

## Inhaltsverzeichnis

Einleitung . . . . .	5
Nanette Rißler-Pipka	
Der Autor als Individuum und das Individuum als Autor . . . . .	21
Leo Spitzer und die Stilistik der ersten Hälfte des 20. Jahrhunderts	
Diego Stefanelli	
“Dice Bernardo a Cristo que un Portero al Papa” . . . . .	35
Indagaciones Multi-Autorales a partir de la reescritura hispánica de un soneto de Burchiello en un manuscrito florentino	
Alan Joaquín Pérez Medrano	
Estilometría y la Edad Media castellana . . . . .	49
José Manuel Fradejas Rueda	
Fernando de Herrera y la autoría de <i>Versos</i> . . . . .	75
Un primer acercamiento al drama textual desde la Estilometría	
Laura Hernández Lorenzo	
Góngora hors norme? . . . . .	91
Étude stylométrique d'un motif gongorin	
Marie-Églantine Lescasse	
In Search of a New Language . . . . .	117
Measuring Style of Góngora and Picasso	
Nanette Rißler-Pipka	
Delta Inside Valle-Inclán . . . . .	151
José Calvo Tello	
Psychologie des Stils . . . . .	165
Autorpsyche, Erzählanalyse und die Poetik Antonin Artauds	
Andrea Gremels und Maren Scheurer	
Interkulturalität und Autorschaft . . . . .	181
Tamara Jadrežić und Alida Bremer	
Marijana Erstić	
Intención autoral y mercado . . . . .	195
Una propuesta	
Shelley Godsland	

## Anhang

Tabellen und Code . . . . .	209
Abbildungsverzeichnis. . . . .	211
Verfasser- und Schlagwortindex . . . . .	215

# Einleitung

## Theorien von Autorschaft und Stil in Bewegung: Stilistik und Stilometrie in der Romania

Nanette Rißler-Pipka (Karlsruhe)

**ZUSAMMENFASSUNG:** Zwischen dem Verdikt vom Tod des Autors (Barthes 1968) und der folgenden Re-definition des Autorbegriffs (Foucault, Genette, Eco, etc.) liegt die Notwendigkeit literaturgeschichtlicher Klassifikation von Autoren, Stilen und Epochen auf der einen Seite und linguistischer Beschreibung von Sprachstilen auf der anderen Seite. Im Literatursystem hat sich die Bedeutung des „Autors“ zurecht in einem Geflecht von Produktionsbedingungen, Text, Leser, Performanz, Diskurs, Medien, Distribution und Rezeption relativiert. Wenn es in diesen theoretischen Vorbemerkungen mehr um den Begriff des Stils geht als um denjenigen des Autors, dann liegt das vor allem daran, dass im Zuge der Autorschaftsdebatte die Frage des Stils und seiner Vermessung oder theoretischen Beschreibung seit Spitzer eher einer pragmatisch orientierten Linguistik überlassen wurde. Die Kluft zwischen einer literaturwissenschaftlich-theoretischen und mathematisch-quantitativen Diskussion zu Stil und Autor soll anhand einiger Beispiele, Begriffe und Methoden hier ein Stück weit überwunden werden.

**ABSTRACT:** In spite of the long discussion about authorship and individual style beginning with Roland Barthes' verdict of the author's death and followed up by the re-definition of the author (Foucault, Genette, Eco, etc.), there is still a literary-historical classification of authors, styles and epochs needed. Linguists need to describe different style of speech and literary criticism wants to know to which epoch an author belongs. For good reasons now the author is seen as dependant figure in a system of production conditions, text, reader, performance, discourse, media, distribution and reception. If the following introduction as preliminary remarks concern the concept of style more than that of the author, then this is mainly due to the fact that in the course of the debate about authorship the question of style and its measurement or theoretical description has been left to pragmatic linguistics. But we should try to overcome the gap between the discussion about authorship and style in literary theory and in the computational, quantitative approach with the help of examples and the explanation of methods.

**SCHLAGWÖRTER:** Autorschaft; Stilistik; Stilometrie

**KEYWORDS:** authorship; style; stylometry

Eines der entscheidenden Desiderate zwischen der jungen Disziplin Digital Humanities und der traditionsreichen Romanistik ist ohne Zweifel der

fehlende Austausch, der zu Vorurteilen auf beiden Seiten führt.<sup>1</sup> Im wissenschaftlichen Alltag nutzen viele Romanist/innen<sup>2</sup> ganz selbstverständlich digitale Ressourcen und Tools, ohne dies als Veränderung ihrer bisherigen Arbeitsweise wahrzunehmen. Umgekehrt scheinen in den DH die Funktionsfähigkeit und die Ergebnisse der neuesten Tools wichtiger zu sein als eine Rückführung der Ergebnisse in die Fachwissenschaft. Fast glaubt man auch, im Stil der Kommunikation untereinander sei eine gemeinsame Debatte aufgrund unterschiedlicher Fachvokabeln und Modelle (allein das Wort ‚Modell‘ bedeutet innerhalb der Geisteswissenschaften und der DH etwas Grundverschiedenes)<sup>3</sup> kaum noch möglich. Daher versteht sich dieses Beiheft der *Romanischen Studien* auch als Versuch, das Ruder noch herumzureißen, hin zu einer fachlichen Kommunikation, die Hürden beseitigt und ergänzende (wenn auch nicht zwingend gemeinsame) Ergebnisse erzielt.

<sup>1</sup> Vgl. dazu: Andreas Fickers, „Digital History: On the Heuristic Potential of Thinkering“, Tagung: *Digitale Geschichte (de)konstruieren* – Lille, 27.–29.11.2017. <https://live3.univ-lille3.fr/collections/dh-nord-2017-deconstruire-lhistoire-numerique>.

<sup>2</sup> Im vorliegenden Artikel wird versucht, weibliche und männliche Formen gleichermaßen anzusprechen. Auch im Hinblick auf die #metoo-Debatte und den mit dem Vorwurf der Vergewaltigung konfrontierten Franco Moretti (allerdings bislang ohne Anklage, Moretti bestreitet die Vorwürfe), der im Folgenden mehrfach zitiert wird, sollte jeder Text als Teil von gesellschaftlichen und kulturellen Diskursen einen kleinen Teil dazu beitragen, Stellung zu beziehen. Die Politik mancher Zeitschriften, Beschuldigte wie Moretti nicht mehr zitieren zu dürfen, halte ich allerdings für naiv: Seine Texte verschwinden nicht, und ein Lektüerverbot macht sie nur interessanter. Wer wollte denn alle Filme, die Weinstein produziert hat, nun aus dem kollektiven Gedächtnis löschen?

<sup>3</sup> Gemeint ist auf der einen Seite eine Abbildung von realen Phänomenen durch rechnerische Modelle, die auf diese Weise Dinge zeigen, die wir durch reine Beobachtung nicht erkennen könnten, z. B. Statistik (wie im Topic Modeling). Auf der anderen Seite handelt es sich in den Geisteswissenschaften um theoretische Modelle, Denkmodelle, die im Grunde die Methode abbilden, wie z. B. der hermeneutische Zirkel (Hermeneutik), die Vorstellung eines impliziten Lesers (Rezeptionsästhetik) oder die Betrachtung des Textes als beschreibbares Modell (Strukturalismus). Wie man an dieser Auflistung bemerkt, nutzen neuere Literaturtheorien weniger den Modellbegriff, weil z. B. die Idee des Diskurses nur noch schlecht in einem eingegrenzten Modell abgebildet werden kann. Doch im Grunde gilt für beide Disziplinen der aus der Statistik stammende Satz von George Box: „All models are wrong, but some are useful“ – Box nutzt hier rhetorisch auch die Synonyme mathematisches Modell und Kunstmodell und führt weiter aus: „he [the scientist] must not be like Pygmalion and fall in love with his model“ (Box, „Science and Statistics“, *Journal of the American Statistical Association* 71, Nr. 356. (Dec. 1976): 791–99, hier 792). Keinem Modell gelingt es, die Wirklichkeit in allen Details abzubilden, aber es kommt auf den Erkenntnisgewinn durch die Modelle an. S. u. für genauere Ausführungen zu den Begriffen ‚Modell‘, ‚Norm‘, ‚Motiv-Thema‘.

Um vorweg einige Missverständnisse auszuräumen, muss der Begriff ‚Stil‘ deutlich von ‚Stilometrie‘ – aber auch von ‚Stilistik‘ abgegrenzt werden. Stilometrie vermisst von Algorithmen erkannte, graduell messbare (in einen statistischen Wert übersetzbare) Unterschiede zwischen Texten. Die zu vergleichenden Texte sollten etwa in der Wortanzahl und Epoche (und im Genre) übereinstimmen, um verwertbare Ergebnisse zu erhalten. Bezeichnet man die so festgestellten Differenzen zwischen Texten als Autorenstil, liegt darin selbstverständlich eine nicht zu unterschätzende Interpretation, die auf einer ganz bestimmten, sicherlich begrenzten Definition von Stil beruht. Je nach vermessenen Faktoren handelt es sich um eine selektiv induktive Beschreibung des Autorenstils. Denn nur die Tatsache, dass die Ergebnisse nachweislich die Autor/innen und ihre Texte korrekt einander zuordnen, lässt den Schluss zu, dass diese vermessenen Elemente für den Autorenstil verantwortlich sind. Dies war in der Stilometrie schon in den 1960er Jahren eine sensationelle Entdeckung,<sup>4</sup> dass von den vielen möglichen Faktoren, die wir als individuellen Stil einer/s Autorin/s bezeichnen möchten, ausgerechnet ein so simples Element wie die Häufigkeit und Verteilung der MFW (most frequent words) ausschlaggebend sein soll. Dabei muss einem grundlegenden Missverständnis vorgebeugt werden, das zu falschen Erwartungen seitens der Literaturwissenschaft führt: Es ging bei dem Verfahren der Stilometrie (das auch noch weit mehr Elemente als die MFW untersuchen kann, je nach ‚Schule‘)<sup>5</sup> meist nicht um literaturwissenschaftliche oder linguistische Untersuchungen des Autorenstils, sondern um Autorschaftsattributions sowie um forensische Linguistik, die abgesehen von einigen Spezialfällen der Literaturgeschichte meist nicht im Zentrum unseres Interesses steht. Dennoch lädt das Verfahren der Stilometrie ganze Forschergruppen von (überwiegend Literatur-)Wissenschaftler/innen aktuell dazu ein, über den Mehrwert der Methode jenseits von Autorschaftsattributions nachzudenken.<sup>6</sup> Nichtsdestotrotz lieferten Fälle von Autorschafts-

<sup>4</sup> Vgl. Frederic Mosteller und David Wallace, *Inference and Disputed Authorship: The Federalist* (Reading, MA: Addison-Wesley, 1964).

<sup>5</sup> Vgl. das Tool JGAAP von Patrick Juola, das verschiedene Elemente der Texte vermisst und daher nicht zur Klassifikation (Vergleich), sondern auch zur Verifikation genutzt werden kann. Vgl. dazu z. B. Juola, „Stylometry and Immigration: A Case Study“, *Journal of Law and Policy* 21, Nr. 2 (2013): 287–98.

<sup>6</sup> Vgl. die im letzten Jahr von Schöch u. a. gegründete europäische COST CA16204 Action *Distant Reading for European Literary History* ([http://www.cost.eu/COST\\_Actions/ca/CA16204](http://www.cost.eu/COST_Actions/ca/CA16204)); das vorwiegend nordamerikanische Forschernetzwerk *NovelTM Text Mining the Novel: A Multi-University Digital Humanities Initiative* (<https://novel-tm.ca/>); die Nachwuchsforschergrup-

attribution zunächst auch im Kreise der Literaturwissenschaft den Beleg für verlässliche Ergebnisse der Methode. Gerade ein Beispiel wie der von Laura Hernández Lorenzo vorgelegte Fall des Dichters Fernando de Herrera zeigt, dass auch mit einer gelungenen Autorschaftsattribution neue Debatten angestoßen werden können. Auch sprachhistorische Untersuchungen und Spuren von Übersetzern können mithilfe von Stilometrie geleistet und nachgewiesen werden (vgl. dazu den Beitrag von Fradejas Rueda). Doch auch ‚Fehler‘ in der Autorschaftsattribution können, literaturwissenschaftlich betrachtet, zu wertvollen Erkenntnissen führen. So führt z. B. ein Korpus, das durchmischte Gattungen enthält, dazu, dass nicht die Autor/innen, sondern die Gattungen zugewiesen werden. Daraus leitet sich das Forschungsgebiet der Gattungsstilistik ab.<sup>7</sup> Denn hier scheint man auch von der polemisch geführten Autorschaftsdebatte befreit. Ebenso wie Autorenzuweisungen sind allerdings auch Gattungszuschreibungen gerade innerhalb der modernen Literatur problematisch und dynamisch an Fachdiskurse gebunden. Da können stilometrische Untersuchungen, die nur den Text bzw. viele Texte als Grundlage haben, etwas mehr Sachlichkeit in die hermeneutische Debatte bringen. Selbst für einen einzigen Autor wie Ramón Valle-Inclán können verschiedene eigene Gattungen und Schreibstile definiert werden (vgl. dazu Calvo Tello). Grundlage der Diskussion bleibt sicher in allen Bereichen, welche Faktoren als ausschlaggebend und charakteristisch für den ‚einzigartigen‘ Autorenstil angesehen werden und damit ist man auch wieder mitten drin in der Debatte um die Figur des Autors, seine Funktionen, Zuschreibungen, etc. (vgl. dazu die Beiträge von Pérez Medrano und Godsland).

## Eine Theorie des Stils und seiner Begrifflichkeiten

Stil kann sehr viel bedeuten – und dementsprechend gibt es auch ebenso viele Vorstellungen von Stilistik. Die Beiträge von Gremels/Scheurer und von Stefanelli greifen in diesem Band die Frage nach der Definition von Autorenstil und Stilistik dezidiert auf. Um literarische Wiedererkennbarkeit zu garantieren und/oder um als Gruppe gemeinsam aufzutreten (wie z. B. die Sur-

---

pe CLIGS (von Christof Schöch geleitet, Universität Würzburg/Trier: <https://cligs.hypotheses.org>) sowie die Entwicklergruppe des Tools *stylo: Computational Stylistics* (Eder, Rybicki, Kestemont, <https://sites.google.com/site/computationalstylistics>).

<sup>7</sup> Vgl. die Gruppe CLIGS (computergestützte Gattungsstilistik), sowie Christof Schöch, „Corneille, Molière et les autres: stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik“, *PhiN* 7, 2014: 130–57, <http://web.fu-berlin.de/phin/beiheft7/b7to8.pdf>. Vgl. im vorliegenden Band den Beitrag von Calvo Tello.

realisten), werden zum Teil bewusst Stilmerkmale eingesetzt, die nur schwer zu vermessen scheinen. Wie sollte man z. B. die Tendenz zur Sinnverweigerung oder Dekonstruktion in den Avantgarden als Stilmerkmal vermessen? Im Gegensatz dazu erscheint es aber umso erstaunlicher, dass auch diese Autor/innen allein aufgrund der statistischen Worthäufigkeit mit stilometrischen Methoden auseinandergehalten werden können. Das bedeutet keineswegs, dass damit der diskursiv nur schwer zu erfassende surrealistische Stil ‚vermessen‘ sei, sondern dass unabhängig von essentialistisch, biographischen Beobachtungen Autor/innen offensichtlich unbewusst (und damit ganz im Sinne des Surrealismus) die kleinen Füll- und Funktionswörter (MFW) in einer unverwechselbaren Weise verwenden. Bringt uns dies jedoch in der Lösung literaturwissenschaftlicher Fragestellungen weiter? Ja und nein, könnte man antworten: Stilometrie ersetzt keine Diskussionen um die Semantik, die Polyvalenz, die Ironie des Textes, aber um beim Beispiel des Surrealismus und der Avantgarden zu bleiben: Stilometrie kann mit sehr schlichten und überprüfbaren Mitteln essentialistische Interpretationen, die sicher seit Leo Spitzer nicht ausgestorben sind, widerlegen. Außerdem können beispielsweise Fragen kollaborativer Autorschaft, die in den Avantgarden als bewusstes Mittel der Textkreation (die gerade nicht mehr auf eine/n Autor/in zurückgeführt werden sollte) eingesetzt wurde, neu betrachtet werden.

Wenn Spitzer über seine stilistischen Untersuchungen zu Diderot anmerkt: „Dies ist der einzige Artikel dieser Sammlung, in dem ich mir den Versuch gestattet habe, zur Seele des Menschen, nicht nur des Autors, vorzustößen“,<sup>8</sup> muss man dies natürlich nicht mehr ernst nehmen, aber wir sollten nicht vergessen, dass es Jacques Derrida zu verdanken ist, diese Vorgehensweise dekonstruiert zu haben. Gerade mit dem Gebrauch des Verbes „vorstoßen“ entlarvt sich Spitzer hier als aggressiver Wahrheitssucher, den Derrida in *Éperons, les styles de Nietzsche* (1978) mithilfe des Stils selbst (der sozusagen zurückschlägt) dekonstruiert.<sup>9</sup> Nichtsdestotrotz haben digitale stilometrische Methoden und die Stilistik im Sinne Spitzer und Co. einige Gemeinsamkeiten. Die postmoderne Theorie kann die Autorfigur und den Individualstil weder gegenstandslos werden lassen, noch die Betrachtung der/s Autor/in als Akteur/in im literarischen System mit einem bestimm-

<sup>8</sup> Leo Spitzer, „Der Stil Diderots“, in ders., *Texterklärungen* (München: Hanser, 1969), 144–75, hier 144.

<sup>9</sup> Vgl. Jacques Derrida, *Éperons, les styles de Nietzsche* (Paris: Flammarion, 1978), z. B. „le style s’avancerait comme l’éperon...“ (29).

ten Wirkungsgrad verweigern<sup>10</sup> (vgl. dazu auch die besondere Relevanz der Interkulturalität im Beitrag von Erstić).

Betrachten wir die Gemeinsamkeiten von Stilistik und Stilometrie, aktualisieren wir gleichzeitig ein Feld der romanistischen Forschung, das weltweit Anerkennung gefunden hat und in der Lage ist, Lösungsansätze für sprachwissenschaftliche und literaturwissenschaftliche Probleme gleichermaßen zu finden. In beiden Teildisziplinen der Romanistik ist wie in allen Gebieten der Wissenschaft eine Notwendigkeit zur Klassifizierung gegeben, auch wenn diese Möglichkeit mit Foucaults *Les mots et les choses* (1966) oder Borges' „La biblioteca de Babel“ (1941) zurecht in Zweifel gezogen wurde. Diese Einteilungen als Mittel des Erkenntnisgewinns werden über Vergleiche vorgenommen: seien es nun Texte, Autor/innen, Epochen oder Gattungen, die untereinander verglichen werden, um zu möglichst genauen Beschreibungen derselben zu gelangen. Auf Textebene ist die Stilistik mit einer Systematik vorgegangen, die sich sehr gut mit digitalen Untersuchungsmethoden nachbilden lässt bzw. mit diesen verwandt ist. Das grundlegende Problem für beide Seiten ergibt sich aus der Frage, was man als Grundlage für einen Vergleich nimmt:

Da aber gerade ein Unterschied, eine Abweichung, die Wahrnehmung des Stils möglich macht, bedienen sich Stilforscher meistens einer Methode des Vergleichs: sie bemühen sich, das System des Stils in bezug auf die Struktur der Sprache zu rekonstruieren. Die Schwierigkeit besteht darin, herauszubekommen, in welcher Form man diese Struktur erfassen soll.<sup>11</sup>

Riffaterre kritisiert die nicht ausreichend formalisierte Herangehensweise der Stilistik, die von einer Autorintention ausgehe, d. h. die/der Autor/in setzt bewusst bestimmte Stilmittel ein, um diesen oder jenen Effekt zu erzielen. Diese Perspektive führe letztlich zu einer psychologisierenden Interpretation, die mehr über die Persönlichkeit der/des Stilforscher/in sage als über diejenige der/des Autor/in oder gar des Textes.<sup>12</sup> Es bleibt allerdings auch für die strukturelle Stilistik ebenso wie für die Digital Humanities ein Trug-

<sup>10</sup> Vgl. dazu die ausführliche Debatte in *Rückkehr des Autors: zur Erneuerung eines umstrittenen Begriffs*, hrsg. von Fotis Jannidis u. a. (Tübingen: Niemeyer, 1999) und auch Matthias Schaffrick und Marcus Willand, *Theorien und Praktiken der Autorschaft* (Berlin: Walter de Gruyter, 2015).

<sup>11</sup> Michael Riffaterre, *Strukturelle Stilistik* (München: List, 1973), 85.

<sup>12</sup> Vgl. Riffaterre, *Strukturelle Stilistik*, 90: „kommt man zu der Überzeugung, daß die Absicht das Definitionsmerkmal des Stils ist“ und später 95: „Unter diesen Bedingungen ist der Stilforscher nur mehr ein Virtuose ...“.



schluss, dass die genaue Definition der vermessenen Elemente den blinden Flecken der/des Forscher/in verhindern könne. Spätestens bei der Interpretation der Ergebnisse kommt es sowohl bei Riffaterre also auch in der aktuellen Stilometrie darauf an, was hervorgehoben wird und was vertuscht wird. Vielleicht macht das Offenlegen der Ergebnisse und der genauen Methode (die im Idealfall intersubjektiv nachvollzogen werden kann) aber auch eine Neu-Interpretation oder Erweiterung einfacher.<sup>13</sup>

Der Vorwurf, den Franco Moretti bei der Präsentation der französischen Übersetzung seiner „Pamphlete“ (*La littérature au laboratoire*, 2016) anbrachte, gilt vermutlich in großen Teilen weiterhin:

Digital Humanities – this has become a practice and not a theory: so people doing things together – I tried to give this research as much a twist towards theory as possible. [...] The ‚Zeitgeist‘ of Digital Humanities leads towards untheorized results<sup>14</sup>

Nicht zuletzt aus diesem Grund fühlte sich der Verband der deutschsprachigen DH offensichtlich zu einem Motto der Jahrestagung verpflichtet, das nun die Theorie in die noch junge Disziplin bringen soll: „Kritik der digitalen Vernunft“ (DHd2018, Köln). In ihrer Keynote betonte Sybille Krämer, dass DH theoretisch die Möglichkeit hätten, blinde Flecken der Geisteswissenschaften aufzudecken und andererseits greife die „Kulturtechnik digitaler Literalität“ auf diejenige „alphabetischer Literalität“ zurück. Es ist also eine gemeinsame Theorie der Digitalität in den Geisteswissenschaften möglich?<sup>15</sup> Eine Frage, die Gefahr läuft an den Hürden der sich immer mehr spezialisierenden Fachdiskurse zu scheitern.

<sup>13</sup> Vgl. dazu den Dialog der Ergebnisse zur Frage des apokryphen *Quijote* und der Identität Avellanedas: Nanette Rißler-Pipka, „Avellaneda y los problemas de la identificación del autor: propuestas para una investigación con nuevas herramientas digitales“, in *El otro Don Quijote: la continuación de Fernández de Avellaneda y sus efectos*, hrsg. von Hanno Ehrlicher (Augsburg, 2016), 27–51, <https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/3704> und die Antwort darauf von Javier Blasco, „Avellaneda Desde La Estilometría“, in *Cervantes: los viajes y los días*, hrsg. von Pedro Ruiz Pérez (Madrid: Prosa Barroca y SIAL Ediciones, 2016), 97–116.

<sup>14</sup> Franco Moretti, „La littérature au laboratoire“ interviewt von Jérôme David, librairie l’Acacia, Paris, 16.10.2016, <https://youtu.be/PNcDCYnUZ8E?t=214>.

<sup>15</sup> Vgl. Sybille Krämer, „Der ‚Stachel des Digitalen‘ – ein Anreiz zur Selbstreflexion in den Geisteswissenschaften? Ein philosophischer Kommentar zu den Digital Humanities“, *Konferenzabstracts DHd2018 Kritik der digitalen Vernunft* 26.02.–02.03.2018, 17, <http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf>. Vgl. außerdem die Symposienreihe „Digitalität in den Geisteswissenschaften“, <http://digitalitaet-geisteswissenschaften.de>.

Löst aber Moretti sein Versprechen von mehr Theorie selbst ein? Seine theoretischen Anleihen stammen fast ausschließlich aus der Stilistik (Spitzer, Auerbach, Curtius), die er als „existing theories“ nun mithilfe der vorhandenen digitalisierten Korpora testen möchte.<sup>16</sup> Allein die Tatsache, dass wir die Masse an vorhandenen Büchern nicht mehr lesen können, bezeichnet Moretti als „pure theory“.<sup>17</sup> Unter Theoriebildung verstehe ich jedoch etwas mehr und auch etwas anderes als das abstrakte Philosophieren über Buchtitel oder über die Anzahl von Büchern, Wörtern, token und types.

Unterschiedliche Methoden zu vergleichen oder auch unterschiedliche Lektüren ein und desselben Textes können zu einer Selbstreflexion und damit zur Theoriebildung führen. Die Frage, was der individuelle Stil einer/s Autor/in ist, kann immer nur im Einzelfall und für die jeweilige Methode definiert werden, aber sicher nicht generalisiert. Die Stilfrage ist wie der Autorbegriff selbst hochgradig historisch dynamisch und kontextabhängig.<sup>18</sup> Vergleicht man vor diesem Hintergrund nun den Ansatz von Moretti mit demjenigen Leo Spitzers, wird deutlich, dass zwar beide Mikro- und Makroanalyse verwenden, aber doch logisch genau gegenteilig vorgehen. Spitzer schließt sowohl mikro- als auch makroanalytisch induktiv vom einzelnen Textabschnitt (z. B. dem Bericht des Théramène in Racines *Phèdre*) auf das Gesamtwerk des Autors und damit auf den Autorenstil (Spitzer kommt damit zum Merkmal der „klassischen Dämpfung“, die er als Erfindung Racines betont).<sup>19</sup> Moretti geht dagegen deduktiv vor, indem er eine vorherige These oder ein bestehendes Konzept mithilfe vieler Einzelexte testen möchte. Im Grunde liegt allerdings beiden Methoden das Sammeln von Belegen zugrunde, welche, philosophisch betrachtet, auf Spuren, Indizien

<sup>16</sup> Franco Moretti, „Literature Measured“, Literary Lab, Pamphlet 12, 6ff., <https://litlab.stanford.edu/LiteraryLabPamphlet12.pdf>. Dabei ist die Idee, traditionsreiche Erkenntnisse der Forschung mit neuen Methoden zu testen, in den DH und auch in anderen Wissenschaften sehr verankert (im Gegensatz zur Literaturwissenschaft). Vgl. dazu Christof Schöch, „Wiederholende Forschung in den digitalen Geisteswissenschaften: zwei literaturwissenschaftliche Fallstudien“, Vortrag: *dhd2017, Digitale Nachhaltigkeit*, Bern, Februar 2017, <https://zenodo.org/record/277113#.WpUcuajOXcs>.

<sup>17</sup> Moretti, „La littérature au laboratoire“.

<sup>18</sup> Vgl. zu einer Stildebatte, die eine Brücke zwischen Stilometrie und Stilistik schlägt: Berenike Herrmann, Karina van Dalen-Oskam und Christof Schöch, „Revisiting Style, a Key Concept in Literary Studies“, *Journal of Literary Theory* 9, Nr. 1 (2015): 25–52.

<sup>19</sup> Vgl. Leo Spitzer, „Der ‚Bericht des Théramène‘ in Racines ‚Phèdre‘“, in *Racine*, hrsg. von Wolfgang Theile (Darmstadt: Wiss. Buchgesellschaft, 1976) 43–98; sowie ders., „L'effet de sourdine dans le style classique: Racine“, *Études de style* (Paris: Gallimard, 1970) 208–335.

und Fahrten zurückgeführt werden können.<sup>20</sup> Auch benutzen beide Methoden Repräsentationen von Realität, die man im algorithmischen Sinne als Modelle bezeichnen könnte (Spitzer reduziert das Funktionieren von Racines Gesamtwerk, seines Stils auf einzelne Passagen und auf das Modell der klassischen Dämpfung, während Moretti meint, durch möglichst viele Texte die Gesamtheit der literarischen Schriften vermessen zu können).

## Eine Kombination quantitativer und qualitativer Methoden?

Zuletzt basiert die Einschätzung von Stil und Autorschaft vielleicht noch mehr als andere Konzepte in der Literatur- und Kulturwissenschaft auf Interpretation. Der entscheidende Unterschied liegt vielleicht eher darin begründet, dass Literaturwissenschaftler/innen gelernt haben, Texte zu interpretieren, aber weniger gewöhnt sind, Graphen und die dahinterliegenden Zahlen (statistische Werte) oder gar die verwendeten mathematischen Modelle zu interpretieren und ihre Passgenauigkeit zu beurteilen. Es ist unbestreitbar, dass kaum jemand auf angemessen hohem Niveau gleichzeitig Literaturwissenschaftler/in (noch schwieriger Romanist/in) sein kann und Informatiker/in, Computerlinguist/in, etc. Zwar stanno sich im Kern der DH immer mehr Forscher/innen neben einer Grundausbildung in den Geisteswissenschaften auch mit Kenntnissen im Bereich der Programmierung, Modellierung etc. aus, aber die eigentliche Lösung des Problems muss in einem für die traditionelle Literaturwissenschaft sehr ungewohnten Prinzip des Teamworks und der kollaborativen Autorschaft von wissenschaftlichen Arbeiten liegen. Das Nebeneinander von quantitativen und qualitativen Methoden der im vorliegenden Beiheft vorgestellten Ansätze zum Thema von Stil und Autorschaft kann demnach nur ein erster Schritt hin zu mehr konkreter Zusammenarbeit sein.<sup>21</sup> Das bedeutet keineswegs, dass sich alle

---

<sup>20</sup> Vgl. dazu die postmoderne Philosophie der Spur und Fahrte, die sich im Strudel der Signifikanten verliert – z. B. Jacques Derrida, *Marges de la philosophie* (Paris: Ed. Minuit, 1972) – oder auch, davon angeregt, die Ausführungen Sibylle Krämers, „The Humanities going digital? Vom alphanumerischen Zeichenraum zum vernetzten Datenraum“ <http://digitalitaet-geisteswissenschaften.de/the-humanities-going-digital>.

<sup>21</sup> Vor allem im sehr umfangreichen Beitrag von Marie-Eglantine Lescasse ist ein ernsthafter Versuch unternommen worden, beide Perspektiven zu verbinden. Dies ist nur möglich, weil das Projekt des OBVL (Observatoire de la vie littéraire)-Paris „Édition digitale et étude de la polémique autour de Góngora“ (mit 36 Forscher/innen) sowie das Forschungszentrum CLEA (Civilisations et Littératures de l'Espagne et d'Amérique du Moyen-Age aux Lumières) der Sorbonne (beides unter der Leitung von Mercedes Blanco) mit entsprechend großer Erfahrung in der Hispanistik und DH dahinter steht.

Probleme in diesem Bereich nur noch kollaborativ lösen lassen, sondern dass die Chance eines digital verzerrten Blicks<sup>22</sup> auf altbekannte, aber auch auf nie beachtete Texte, genutzt werden kann. Dieser Prozess wird ohne Zweifel noch Jahre oder Jahrzehnte dauern, denn auf Seiten der DH fehlen noch die angeblich ‚bahnbrechenden‘ Ergebnisse, die, zurückgetragen in die Fachwissenschaft, dort auf so viel Bewunderung und Veränderung stoßen, dass kein Weg mehr an ihnen vorbeiführt. Auf Seiten der Literaturwissenschaft (um konkret im Kontext der Perspektive dieses Bandes zu bleiben) fehlt die Offenheit für eine experimentelle Arbeitsweise, für neue, gemeinsame Fragestellungen (denn es lassen sich eben nicht notwendigerweise, die gleichen Fragen mit neuen Methoden beantworten). Für Letzteres, für das Finden der gemeinsamen Fragestellung, wird es vermutlich einige Zeit brauchen.

Andrew Piper betont zurecht in seinem Aufruf „Send us your null results“ in *Cultural Analytics*, dass von Seiten der DH nicht so sehr mit den erstaunlichen Ergebnissen geprahlt werden sollte und der übliche Vorwurf, alle Ergebnisse seien vorher auch schon bekannt gewesen (nur jetzt schicker visualisiert), nicht von der Hand zu weisen ist. Doch auch auf die wenig signifikanten Ergebnisse sollte man wissenschaftlich stolz sein, denn in einer den Geisteswissenschaften eher fremden Art und Weise gehört das Iterative des statistischen und algorithmischen Experiments notwendig zum Erkenntnisgewinn dazu. Erst auf diese Weise kann auch geklärt werden, dass digitale Methoden keine Wunderwerke sind, die die kritische Interpretation und das Lesen von Texten unnötig erscheinen lassen, sondern wo ihre Grenzen sind.

If we were to take the set of all experiments ever conducted with a computer on some texts, I would expect that in (at least) 95% of those cases the procedure yielded no insight of interest. In other words, positive results would be very rare. And yet, miraculously, all articles in CA report a positive result (mine included).<sup>23</sup>

Gerade das Scheitern am Material, dem Text, der Widerspenstigkeit einer Sprache, die sich nicht normieren und damit nicht auslesen lässt, erscheint mir als Nachzeichnen einer Grenze zwischen Lektüre und machine learning erkenntnisreich genug (vgl. den Beitrag zu Góngora und Picasso in diesem

---

<sup>22</sup> Hier beziehe ich mich auf den Ansatz von Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism, Topics in the Digital Humanities* (Urbana: University of Illinois Press, 2011). Ramsay vergleicht den digitalen Blick auf den Text mit der avantgardistischen Verfremdung oder des russischen Formalismus.

<sup>23</sup> Andrew Piper, „Send us your null results“, *Cultural Analytics* (5. März 2018), <http://culturalanalytics.org/2018/03/send-us-your-null-results>.

Band). Umgekehrt hat auch die hermeneutische Textanalyse (ebenso wie die systemanalytische, formalistische, struktur- oder poststrukturalistische und empirische) ihre Grenzen und ihre blinden Flecken dort, wo die Materialität des untersuchten Mediums beginnt (Schrift, Papier etc.).<sup>24</sup> Gleichzeitig ist mit einem solchen Nachzeichnen keine Sackgasse erreicht, sondern es wird die Weiterentwicklung der Methode auf der einen Seite angestoßen, und auf der anderen Seite ergeben sich andere, neue und vorher nicht in Betracht gezogene Fragestellungen.

### Von unterschiedlichen Begriffen, Missverständnissen und anderen Fallstricken zwischen Stilistik und Stilometrie

Zu Beginn habe ich auf die disziplinär unterschiedlichen Bedeutungen von Begriffen wie ‚Modell‘ hingewiesen. Dies gilt auch für ‚Autor‘ und ‚Stil‘ – ließe sich doch trefflich darüber diskutieren, inwieweit Autor/innen noch Macht über ihren Stil haben und inwieweit er von Einflüssen anderer Texte, Autor/innen, Medien, Kulturen (vgl. den Beitrag von Erstić), Buchmarktmechanismen (vgl. den Beitrag von Godsland) etc. beeinflusst ist und wie man diese kaum fassbar beschreibbaren Einflüsse vermessen möchte. Die Stilometrie ist m. E. dazu in der Lage, ein wenig Polemik aus der Debatte um Autor und Stil heraus zu nehmen. Denn die Tatsache, dass so einfache Parameter wie die MFW verlässlich Stil und Autor bestimmen können, zeigt sowohl dass die Autorperson sich nicht vollständig hinter postmoderner Theorie verstecken kann und auch, dass Stil als komplexes Phänomen zwar weiterhin existiert, aber als reines Unterscheidungskriterium auch auf schlichter Ebene der Wortverteilungen betrachtet werden kann.

An drei Begriffen, die viele Analysen betreffen, aber weniger ‚vorbelastet‘ sind als ‚Stil‘ und ‚Autor‘, möchte ich abschließend die Problematik der gemeinsamen Kommunikation und Zusammenarbeit an konkreten Beispielen erläutern.

**MODELL** Auch in der sogenannten analogen Welt, die wir sehen und anfassen können, gibt es zahlreiche Arten von Modellen. Als Literatur- und Kultur-

---

<sup>24</sup> Im Bewusstsein, dass das weite Feld der Materialität hier nur angesprochen werden kann, sei auf Roland Barthes, *Variations sur l'écriture* (Mainz: Dieterich'sche Verlagsbuchhandlung, [1973] 2006) verwiesen, sowie auf die Untersuchungen der 1980er Jahre *Materialität der Kommunikation*, hrsg. von Hans Ulrich Gumbrecht und Karl Ludwig Pfeiffer (Frankfurt a.M.: Suhrkamp, 1988) und auf die aktuelle Debatte in *Handbuch Literatur & Materielle Kultur*, hrsg. von Susanne Scholz und Ulrike Vedder (Berlin und Boston: De Gruyter, 2018), <https://doi.org/10.1515/9783110416497>.

wissenschaftlerin nutze ich beispielsweise das Architekturmodell (der Architekten Josep Lluís Sert und Luis Lacasa) des Pavillons der spanischen Republik, in dem 1937 auf der Weltausstellung in Paris Picassos *Guernica* erstmalig ausgestellt wurde.<sup>25</sup> Während der Modellbegriff aus der Architektur sehr konkret verständlich als Miniaturabbildung eines möglichen Gebäudes genutzt wird, sieht es im Bereich der Theorie von Repräsentation, Epistemologie und Wahrnehmung schon schwieriger aus. Denn auch das Gemälde von Diego Velázquez *Las Meninas* (1656) gilt als Modell für ein neues Denken von Repräsentation und Wahrheit, das Foucault prominent in *Les mots et les choses* (1966) diskutiert hat.

Der Modellbegriff in den DH umfasst verschiedene Arten von algorithmischen Modellen, die jeweils eine vereinfachte Darstellung eines realen oder gedanklichen Problems sind und sich für die verschiedenen Fälle mehr oder weniger gut eignen (und dementsprechend angepasst werden müssen). Der Informatiker David Mimno, Spezialist für Topic Modeling, erklärte in seinem Vortrag „From Numbers to Evidence“ (Stanford, 2015), dass Computer-Modelle für verschiedene Objekte unterschiedlich gut geeignet sein können. In seinem Beispiel, einem aus Minecraft-Cubes erstellten Modell der *Sagrada familia*, vermisste er den „emotional effect of the building“. Letzteres ist ein etwa ähnlich unsicher und individuell unterschiedlich beschreibbarer Effekt wie er bei der Lektüre eines literarischen Werks entsteht. Das bedeutet, dass dieser schwer fassbare „emotional effect“ auch schlecht algorithmisch modellierbar ist, womöglich aber mit Worten umschreibbar.

**NORM** Auch hier gibt es sowohl geisteswissenschaftlich betrachtet als auch auf Seiten der Computerlinguistik und Informatik unterschiedliche Bereiche, die mit dem Begriff ‚Norm‘ zusammenhängen. Während spätestens seit den Avantgarden die Frage der Norm im Sinne von Normalität als künstlerische Abgrenzung des Besonderen im Normalen diskutiert wurde<sup>26</sup> und die diskurstheoretische Beschreibung des Normalismus durch Jürgen Link<sup>27</sup> seit den 1970er Jahren die geisteswissenschaftliche Debatte beherrscht, hat

<sup>25</sup> Vgl. die verschiedenen Ansichten unter: <http://catalogo.artium.org/book/export/html/6086> sowie <http://www.steponybarno.es/blog/2009/08/18/el-pabellon-espanol-de-1937-josep-lluís-sert> (beide zuletzt abgerufen am 07.03.2018).

<sup>26</sup> Vgl. zum besonderen Aspekt des Normalen im Gegensatz zum ‚Verrückten‘ auch den psychoanalytischen und psychopathologischen Diskurs bzgl. Artaud im Beitrag von Gremels/Scheurer.

<sup>27</sup> Vgl. Jürgen Link, *Versuch über den Normalismus: wie Normalität produziert wird* (Göttingen: Vandenhoeck & Ruprecht, 2006).

Norm im Sinne von Normalisierung in Bezug auf die Maschinenlesbarkeit von Texten und Datenbanken eine wesentlich praktischere Funktion (das Eliminieren von nicht-lesbaren Zeichen, von Redundanzen, etc.). Im Kontext von Literatur kommen noch normbildende Poetiken hinzu, deren Effekte auf den Stil einer ganzen Gruppe von Autor/innen je nach Art der Vorgaben (Normen) messbar sind (vgl. dazu die genaue Analyse von Lescasse bzgl. der Normabweichungen im Werk Góngoras). Gerade die Frage, ob man ausgiebig diskutierte Normabweichungen bestimmter Autoren, die sich meist als Stilabweichungen genauer beschreiben lassen, nun detaillierter bestimmen kann, wirft erneut das Problem auf, ob durch DH-Methoden neue Erkenntnisse gewonnen werden (vgl. dazu das Beispiel von Lescasse in diesem Band sowie den Vortrag von Schöch<sup>28</sup>).

**MOTIV UND THEMA** Was ein literarisches Motiv ist und was ein Thema ist, wurde im literaturwissenschaftlichen Gebiet der Motiv- und Stoffgeschichte ausgiebig diskutiert und heute im Prinzip ad acta gelegt.<sup>29</sup> Dagegen ist das begrifflich verwandte Topic Modeling in den DH neben Word Embeddings und Stilometrie eine der wichtigsten Methoden der quantitativen Textanalyse. Zwar lassen sich auch stilistisch betrachtet bestimmte wiederkehrende grammatikalische Strukturen und Sequenzen als Motiv bezeichnen (vgl. Lescasse in diesem Band), aber sie bilden ebenso wenig wie die Stilometrie auf Grundlage der MFW die semantische Ebene von Texten ab. Während die menschliche Lektüre Motive und Sinn in Texten durch das Herstellen von logischen Zusammenhängen zwischen Wörtern, Sätzen, Abschnitten, aber auch – und vor allem – über Texte und Medien, Diskurse, Kulturen und Hypertexte hinweg herzustellen vermag, bleibt die algorithmische Lektüre beschränkt auf das verwendete Korpus und auf das zugrunde gelegte Modell. „Topics“ sind daher nicht notwendigerweise identisch mit dem literarischen Thema oder Motiv, sondern können nur durch einen zweiten Schritt in der menschlichen Interpretation als solche angesehen werden.

---

<sup>28</sup> Leider wurde aus dem Vortrag noch kein fertiger Artikel, aber die Folien sind veröffentlicht: <https://christofs.github.io/intertext/#/>

<sup>29</sup> Auch unter neuen Labels bleibt jedoch die literatur-, kultur- und medienwissenschaftliche Forschung in weiten Teilen an Themen und Motiven orientiert, die im Zeitverlauf beobachtet werden: sei Raum, Zeit, Gewalt, Krieg, Terror, Mythen, Traum, Affekte, Grenzen, Migration – um nur einige Themen zu nennen, die auf den letzten Verbandstagen in der Romanistik eine große Rolle spielten. Zwar sind diese Themen wesentlich abstrakter als in der klassischen Motiv- und Stoffgeschichte, aber gerade aus diesem Grund ließen sie sich eher mit der Herangehensweise von Topic Modeling oder Word Embeddings vergleichen.

Zwar haben zahlreiche Studien die Passgenauigkeit und Funktionsfähigkeit der Methode auch für klar strukturierte literarische Korpora belegt,<sup>30</sup> aber es entstehen dabei auch immer die weniger in den Fokus gerückten ‚Non-sense-Topics‘. Der entscheidende Vorteil einer solchen Re-organisation des literarischen Textes ist, dass ohne vorherige Fokussierung auf ein bestimmtes Themenfeld oder Motiv, der Text allein aufgrund von algorithmischen Modellen und Wahrscheinlichkeiten neu nach ‚Topics‘ zusammengestellt wird. Das führt keineswegs zu einer Form von Objektivität der Lektüre, da allein die Auswahl des Korpus und der Parameter das Ergebnis beeinflusst, aber die notwendigen blinden Flecken der Lektüre, die durch das kollektive Imaginäre, den kulturellen und persönlichen Background, etc. bestimmt sind, können durch diesen Schritt weitgehend umgangen werden. Ohne Zweifel kommt bei der Interpretation der algorithmisch gewonnenen Topics wieder ein neuer blinder Fleck hinzu, aber es handelt sich in jedem Fall um eine neue Perspektive.

Das Problem der Begriffe, die in verschiedenen Kontexten und Diskursen unterschiedliche Bedeutungen annehmen und diese im Laufe der Zeit verändern, betrifft demnach nicht nur ‚Stil‘ und ‚Autor‘. Daher finden sich im vorliegenden Band unterschiedliche Konzepte von ‚Stil‘ und ‚Autor‘, die zu meist komplexer sind als diejenigen, die von der Stilometrie tatsächlich vermessen werden können. Dort sind es nur wenige, aber sehr genau benennbare Parameter, während die Stil- und Autorschaftsbegriffe, die auf kulturelle Zusammenhänge, intertextuelle, intermediale und interdisziplinäre Verbindungen verweisen, wesentlich die Interdependenzen und Strategien der einzelnen angesprochenen Diskurse aufdecken. An der Frage wie beide Ansätze sinnvoll zusammen zu bringen sind, muss noch weitergearbeitet werden, aber dass sie sich ergänzen wird hier bereits sichtbar.

<sup>30</sup> Vgl. Christof Schöch, „Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama“, *Digital Humanities Quarterly* 11/2 (2017), 1–53, <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html>; Christof Schöch, Ulrike Henny, José Calvo, Daniel Schlör und Stefanie Popp, „Topic, Genre, Text: Topics im Textverlauf von Untergattungen des spanischen und hispanoamerikanischen Romans (1880-1930)“, *Konferenzabstracts dhd2016*, hrsg. von Elisabeth Burr et al. (Leipzig: nisaba verlag, 2016), 235–38, <http://dhd2016.de/boa.pdf>; Borja Navarro, „A Computational Linguistic Approach to Spanish Golden Age Sonnets: Metrical and Semantic Spects“, in *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, hrsg. von Anna Feldman, Anna Kazantseva, Stan Szpakowicz und Corina Koolen, 2015, 105–13, <http://www.aclweb.org/anthology/W15-0712>.



Denn die Stilometrie und andere DH-Methoden sind keineswegs eine Wiederbelebung der Stilistik und derer überholter essentialistischer und positivistischer Herangehensweise. Im Gegenteil wird die Frage nach dem individuellen Autorenstil ersetzt durch die Frage, wo manifestiert sich Stil im Text – und dabei wird Text als maschinenlesbarer Text (tokens & types) angesehen, in dem die Semantik erstmal keine Rolle spielt. Autorschaft wird reduziert auf ein Signal, das ebenso wie die Gattung als Signal beobachtbare Konsequenzen im Clustering von Texten hat. Das hat nicht notwendigerweise weitreichende Konsequenzen, wie eine erforderliche Neuschreibung der Literaturgeschichte, aber kann im Detailergebnis zu wichtigen theoretischen Fragen führen, wie z. B. warum entscheidet die Wortverteilung überhaupt über das Gattungs- oder Autorsignal oder warum scheint das Gattungssignal meist stärker als das Autorsignal oder wie stark können einzelne Autor/innen ihren Stil auch auf der Mikroebene der MFW so verändern, dass sie die Vermessung des Stils unterlaufen?



# Der Autor als Individuum und das Individuum als Autor

## Leo Spitzer und die Stilistik der ersten Hälfte des 20. Jahrhunderts

Diego Stefanelli (Pavia)

**ABSTRACT:** From a historical point of view the paper examines the role of individuality in the stylistics of the first decades of the 20<sup>th</sup> century. In the context of the crisis of positivistic linguistics and literary studies at the beginning of the 20<sup>th</sup> century the individual played an essential role: At the bottom of modern stylistics is not only the dialectic relationship between the writer and the literary tradition, but also between the speaker and his speech. Leo Spitzer, perhaps the most important style critic of the first half of the 20<sup>th</sup> century, and the authors who contributed mainly to his theory of stylistics, Hugo Schuchardt and Karl Vossler are in the focus of this paper. Particularly Vossler and his important theories of style show the intimate relationship between linguistics and literary studies which is the basis of modern stylistics.

**ZUSAMMENFASSUNG:** Von einem historischen Gesichtspunkt aus untersucht der Beitrag die Rolle der Individualität in der Stilistik der ersten Jahrzehnte des 20. Jahrhunderts. Im Kontext der Krise der positivistischen Sprach- und Literaturwissenschaft am Anfang des 20. Jahrhunderts spielte das Individuum eine wesentliche Rolle: der modernen Stilistik liegen nicht nur die Dialektik zwischen dem Schriftsteller und die literarische Tradition zugrunde, sondern auch die zwischen dem Sprechenden und seiner Sprache. Der Beitrag widmet sich dem vielleicht wichtigsten Stilkritiker der ersten Hälfte des 20. Jahrhunderts: Leo Spitzer – und den Autoren, die die wichtigsten theoretischen Bezugspunkte seiner Stilistik waren: Hugo Schuchardt und Karl Vossler. Vor allem bei Vossler und seinen bedeutungsvollen stilistischen Theorien wird das innige Verhältnis zwischen Literatur- und Sprachwissenschaft als Fundament der modernen Stilistik besonders deutlich.

**SCHLAGWÖRTER:** Spitzer, Leo; Vossler, Karl; Schuchardt, Hugo; Geschichte der Stilistik

An der Grenze zwischen Literatur- und Sprachwissenschaft wurde die moderne Stilistik in den ersten Jahrzehnten des 20. Jahrhunderts geboren.<sup>1</sup> Unter diesem Aspekt ist Leo Spitzers Fall sehr exemplarisch, vor allem in seiner

---

<sup>1</sup> Für eine historische Perspektive auf die Stilistik zwischen dem 19. und dem 20. Jahrhundert in Deutschland und Italien und für eine Vertiefung einiger hier nur skizzierter Aspekte der vielfältigen Geschichte der Disziplin, erlaube ich mir, auf Diego Stefanelli, *Il problema dello stile fra linguistica e critica letteraria: positivismismo e idealismo in Italia e in Germania* (Berlin: Frank & Timme, 2017) zu verweisen.

ersten europäischen Periode. Bei verschiedenen Gelegenheiten betonte der österreichische Romanist, dass die Stilistik eine Brücke zwischen zwei Disziplinen sein sollte:

Es klingt paradox, ist aber doch wahr: die Wortkunst hat bisher in der Sprachwissenschaft relativ wenig Beachtung gefunden. Sie wurde meist der Literatur- oder ästhetischen Forschung überlassen. Am grammatisch-historischen Panzer der Linguisten sind die leichten schwebenden Effekte künstlerischer Wortformung abgeprallt [...]. Die Literaturforscher sind also sprachwissenschaftlich, die Sprachwissenschaftler ästhetisch zu wenig gebildet, als daß die Stilforschung, die an der Grenze beider Disziplinen steht, gedeihen könnte.<sup>2</sup>

Nach vielen Jahren wird ein wichtiger italienischer Sprach- und Literaturwissenschaftler (Benvenuto Terracini), der eine wesentliche Rolle in der italienischen Stilistik zwischen der ersten und der zweiten Hälfte des 20. Jahrhunderts spielte, das tiefe und komplexe Verhältnis zwischen Stilistik und Sprachwissenschaft sehr klar ausdrücken:

Qualsiasi linguista il quale ami riflettere sull'oggetto ultimo del proprio studio, di livello in livello, finirà per risalire fino alla fenomenologia della lingua poetica [...], bene o male, idealisticamente o positivisticamente, finirà quindi per penetrare nel regno della stilistica.<sup>3</sup>

Zwischen dem 19. und dem 20. Jahrhundert liefen viele post-positivistischen Wissenschaften (Literatur- und Sprachwissenschaft, Psychologie, Ästhetik) auf das Thema der Individualität heraus. Am Anfang des zweiten Bandes der *Stilstudien* von Leo Spitzer finden wir ein bedeutungsvolles Zitat aus einem Essay von Miguel de Unamuno: „en las lenguas como en los hombres, persigo la individualidad personal“. Der Satz stammt aus Unamunos Essay „En la calma de Mallorca“, aus dem Buch *Andanzas y visiones españolas*.<sup>4</sup> Es handelt sich um eine Art reiseliterarisches Buch, das „relatos de excursiones [...] por ciudades y campos [...] de España“ (*Prólogo*) enthält. Wie es bei Spitzer oft passiert, wirkt das Zitat auf den ersten Blick ein wenig fehl am Platz. Wenn man es aber im Kontext liest, kann man die Unterschiede zwischen Spitzers sprach- und literaturwissenschaftlichen Beiträgen und Unamunos Essay erkennen:

<sup>2</sup> Leo Spitzer, „Wortkunst und Sprachwissenschaft“, in *Stilstudien*, Bd. II. *Stilsprachen* (München: Hueber, 1928), 498–536, hier 498.

<sup>3</sup> Benvenuto Terracini, *Analisi stilistica: Teoria, storia, problemi* (Milano: Feltrinelli, 1975), 7.

<sup>4</sup> Miguel de Unamuno, „En la calma de Mallorca“, in *Andanzas y visiones españolas* (Madrid: Renacimiento, 1922), 158–71, hier 168.

Ya ve el lector que leo algo más que la media docena de libros que me traje, pero lo demás que leo es en mallorquín. Curiosidad de filólogo. Adondequiera que voy me gusta leer en la lengua de aquel país. En Portugal apenas leo sino portugués y ahora aquí leo mallorquín. Pero cuidando que lo sea y no catalán. No es que haya una gran diferencia entre ellos, pues son hermanos gemelos, pero me gusta apreciar las diferencias más bien que las semejanzas. En las lenguas como en los hombres, persigo la individualidad personal. O si se prefiere, la personalidad individual.<sup>5</sup>

Wie man dem Ende des *Prólogo* entnehmen kann, war die Absicht des Buches von Unamuno nicht wissenschaftlich, sondern sozusagen ästhetisch-impressionistisch: „El que gusta del paisaje literario, va a buscarlo en sí y por sí. Y a esta demanda de la afición estética es a lo que quiere responder la oferta de este libro, lector amigo“.<sup>6</sup> Trotz der Verschiedenheit zwischen Spitzer und Unamuno, muss das Zitat in den Kontext eines generellen Interesses an Individualität in den ersten Jahrzehnten des 20. Jahrhunderts gestellt werden.

Das Individuum spielte eine wichtige Rolle in der Krise der positivistischen Sprachwissenschaft. Die Opposition gegen das positivistische System der Sprachwissenschaft (nämlich die Lautgesetze der Junggrammatiker) stützte sich, sozusagen, auf die Rechte des Individuums. Einer der wichtigsten Protagonisten der methodologischen und theoretischen Kritiken an den Junggrammatiker war der Romanist Hugo Schuchardt, dessen Name sehr wichtig für die Stilistik und vor allem für Spitzer war.<sup>7</sup> Ein Beweis dafür ist Schuchardts Anthologie, die Spitzer 1922 und als zweite Ausgabe 1928 herausgab.<sup>8</sup> In der bedeutungsvollen Vorrede zur 1. Auflage, wurde Schuchardt von Spitzer als „ein Klassiker“ und als „der Ahnherr und Vorahner des sprachwissenschaftlichen Expressionismus“<sup>9</sup> beschrieben. Schuchardt wollte, ebenso wie die neue post-positivistische Sprachwissenschaft, das Detail in Verbindung mit dem Allgemeinen in den Mittelpunkt des Sprachwandels

<sup>5</sup> Unamuno, *Andanzas y visiones españolas*, 167–8.

<sup>6</sup> Unamuno, *Andanzas y visiones españolas*, 8.

<sup>7</sup> Für das persönliche und wissenschaftliche Verhältnis zwischen Schuchardt und Spitzer, ist ihr Briefwechsel ein wesentlicher und unabdingbarer Ausgangspunkt: *Leo Spitzers Briefe an Hugo Schuchardt*, hrsg. und eingeleitet von Bernhard Hurch, unter editorischer Mitarbeit von Niklas Bender und Annemarie Müllner (Berlin und New York: Walter de Gruyter, 2006).

<sup>8</sup> Ich zitiere aus dem „unverändertem reprographischen Nachdruck“ der 2., erweiterten Auflage (Halle: Niemeyer, 1928): *Hugo Schuchardt-Brevier: ein Vademecum der allgemeinen Sprachwissenschaft*, zusammengestellt und eingeleitet von Leo Spitzer (Tübingen: Niemeyer, 1976).

<sup>9</sup> *Hugo Schuchardt-Brevier*, 4.

rücken: „Schuchardts Motto ist ja: ‚Das Kleinste nicht verachten und nach dem Größten trachten‘ [...] – er verbindet Mikroskopie und Makroskopie – er blickt gen Himmel, nachdem er alle Wörterbücher und Traktate mit unvergleichlicher Akribie durchpflügt“.<sup>10</sup>

In Bezug auf die Zentralität der Individualität sind einige Seiten des *Schuchardt-Breviers* besonders bedeutungsvoll. Sie kommen aus zwei wichtigen Werke Schuchardts: das bekannte Pamphlet *Über die Lautgesetze: gegen die Junggrammatiker* (1885),<sup>11</sup> das in Zusammenhang mit der sprachwissenschaftlichen Diskussion über die „Ausnahmslosigkeit der Lautgesetze“ am Ende des 19. Jahrhunderts gelesen werden sollte; und der weniger bekannte Beitrag über *Der Individualismus in der Sprachforschung* (1925).<sup>12</sup> Gegenüber der Ausnahmslosigkeit der Lautgesetze der Junggrammatiker klagte Schuchardt die Zentralität der Individualität der Sprechenden im Sprachwandel ein. Das Individuum ist weniger vorhersehbar als die unpersönlichen und vermutlich mechanischen Lautgesetze. Wenn das Individuum die Sprache „macht“, sind die Regeln dieses individuellen und immer noch dauernden Prozesses mehr psychologisch (z. B. die Analogie) als mechanisch.

So weit direkte Beobachtung an uns selbst oder an Anderen reicht, ist die Aussprache des Individuums von Schwankungen nie frei, worunter ich natürlich keine in strenger Gemäßheit der Zeitfolge auftretende Veränderungen begreife. Mit dieser endlosen Sprachspaltung geht endlose Sprachmischung Hand in Hand. Die Beeinflussung des einen Dialektes durch den anderen, welche den Junggrammatikern zufolge eine Störung der ausnahmslosen Lautgesetze bewirkt, und die Ausgleichung der Individualsprachen, welche denselben Junggrammatikern zufolge ausnahmslose Lautgesetze erst ermöglicht, diese Prozesse von konträrer Wirkung sind im Wesen gleich, sie sind nur verschiedene Mischungsstufen.<sup>13</sup>

Nach vielen Jahren dehnte Schuchardt in *Der Individualismus in der Sprachforschung* diese Idee von Individualität aus:

Alle Kritik, insoweit sie nicht an den äußerlichsten Einzelheiten haftet, ist mehr oder weniger subjektiv; das aber pflegt nicht hinlänglich zum Bewusstsein zu kommen oder geflissentlich verdunkelt zu werden [...]. Dass alles, was wir Sprache nennen, nur eine Abstraktion aus Individualsprachen ist, das hat

<sup>10</sup> Hugo Schuchardt-Brevier, 2.

<sup>11</sup> Hugo Schuchardt, *Ueber die Lautgesetze: gegen die Junggrammatiker* (Berlin: Oppenheim, 1885).

<sup>12</sup> Hugo Schuchardt, *Der Individualismus in der Sprachforschung* (Wien u. Leipzig: Hölder-Pichler-Tempsky, 1925).

<sup>13</sup> Hugo Schuchardt-Brevier, 60.

man längst erkannt; ebenso aber ist alle Wissenschaft nur eine Abstraktion aus Individualwissenschaften.<sup>14</sup>

Nicht nur für Spitzer, sondern auch für andere Sprachwissenschaftler, die am Anfang des 20. Jahrhunderts unzufrieden mit der positivistischen Sprachwissenschaft waren und neue Theorien und Methoden in der Sprachwissenschaft suchten, wurde Schuchardt ein wesentlicher Bezugspunkt. Im Allgemeinen war die post-positivistische Sprachwissenschaft der ersten Jahrzehnte des 20. Jahrhunderts eine methodologisch reiche und kulturell rege Disziplin, die einen Dialog mit anderen Wissenschaften (Literaturwissenschaft, Psychologie, Ästhetik) suchte.

Die Individualität stand im Mittelpunkt dieser vielfältigen und interessanten Periode der Sprachwissenschaft. Sowohl für die neuen sprachwissenschaftlichen Methoden und Theorien (nicht nur Schuchardt aber auch, beispielsweise, die *géographie linguistique* von Jules Gilliéron) als auch für die neoidealistische Ästhetik von Benedetto Croce war das sprachwissenschaftliche Denken Wilhelm von Humboldts wesentlich. Er hatte bereits das Individuum (oder besser gesagt: die Tätigkeit des Individuums) in den Mittelpunkt des Sprachwandels gerückt. Von Humboldt stammte die doppelseitige Definition der Sprache als *ergon* und *energeia*, d. h. als *Tatsache* und *Tathandlung*: die Sprache wurde als individuelle Tätigkeit und Schöpfung betrachtet. Das wurde eine Art Schlagwort vieler sprach- und literaturwissenschaftlichen Bücher und Beiträge dieser Zeit.

In diesem vielfältigen Kontext wurde die moderne Stilistik geboren. Hierzu ist die interessante Einführung von Jean Starobinski zur ersten französischen Übersetzung einiger Schriften von Spitzer<sup>15</sup> besonders bedeutungsvoll:

Derrière Spitzer (comme derrière son ami Karl Vossler) se profile la grande figure de Wilhelm von Humboldt, pour qui l'œuvre de langage, *ergon*, renvoyait à un pouvoir intérieur, *energeia*, à la fois propre au sujet parlant et à sa communauté historique. L'œuvre est donc abordé comme l'expression d'une activité psychique qui l'a conditionnée et façonnée; l'œuvre est la *Tatsache* qui porte la marque d'une *Tathandlung*. Dans une conception qui privilégie à ce point le sujet parlant et l'acte expressif, la stylistique devient la discipline souveraine

<sup>14</sup> Hugo Schuchardt-Brevier, 89.

<sup>15</sup> Leo Spitzer, *Études de style* (Paris: Gallimard, 1970).

parmi les sciences du langage : elle seule est en mesure de percevoir ce que l'acte de parole comporte d'unique et de créateur.<sup>16</sup>

Die Individualität war verbunden mit einem anderen wichtigen Begriff, der die Sprachwissenschaft dieser Zeit charakterisierte: die Affektivität. Auch in diesem Fall handelt es sich um einen innerlich dialektischen Begriff, da man die Affektivität nur in ihrem Verhältnis zur Intellektualität verstehen kann. Das begriffliche Paar Affektivität vs. Intellektualität war die theoretische und praktische Basis einer sehr wichtigen Stilistik der Zeit: die sprachliche *stylistique* von Charles Bally,<sup>17</sup> die sich auf die Interaktion zwischen *affectivité* und *intellectualité* stützte, wie man im *Traité de stylistique française* liest:

Parmi les catégories formelles qui déterminent la comparaison entre deux ou plusieurs faits de langage et permettent de les caractériser, il y en a une qui nous est apparue avec une valeur générale à laquelle aucune autre ne peut prétendre : c'est la présence, en proportion variable, d'éléments intellectuels et d'éléments affectif; les opérations de l'intelligence et les mouvements de la sensibilité concourent à la formation du système expressif, comme ils se partagent toute notre vie intérieure. La distinction entre ces deux ordres de faits domine donc toute notre étude ou, pour mieux dire, elle en est l'objet et la raison d'être.<sup>18</sup>

Interessanterweise finden wir am Ende des 19. Jahrhunderts eine ähnliche, allerdings aus einem verschiedenen Kontext stammende, Opposition in der deutschen Romanistik. In einem methodologischen Beitrag (*Methodik und Aufgabe der sprachwissenschaftlichen Forschung*) im von ihm herausgegebenen *Grundriss der Romanischen Philologie* hatte Gustav Gröber eine solche psychologische Trennung zwischen einer subjektiven (oder affektiven) und einer objektiven (oder intellektuellen) Syntax schon theoretisch beleuchtet. Dabei handelte es sich nicht nur um eine sprachwissenschaftliche Division, insofern Gröber die traditionellen rhetorischen Kategorien mit dieser Theorie interpretieren wollte, sondern sozusagen um eine psychologisch-sprachwissenschaftliche Interpretation der rhetorischen Kategorien.

Die in der empirischen Syntax zu unterscheidende objektive und subjektive Gedankendarstellung treten sowohl in der gewöhnlichen, wie in der künstlerischen Rede verbunden auf [...]. Die Vermischung der beiden Formen der

<sup>16</sup> Jean Starobinski, „Leo Spitzer et la lecture stylistique“, in *La relation critique*, édition revue et augmentée (Paris: Gallimard, 2001), 57–108, hier 73.

<sup>17</sup> Vgl. Anamaria Curea, *Entre expression et expressivité: l'école linguistique de Genève de 1900 à 1940*. Charles Bally, Albert Sechehaye, Henri Frei (Lyon: ENS Éditions, 2015).

<sup>18</sup> Charles Bally, *Traité de stylistique française* (Genève-Paris: Klincksieck, 1909), 155.



Gedankendarstellung in den meisten Arten der Rede kann nicht zur Rechtfertigung derjenigen Grammatiker dienen, die sie nicht trennen [...]. Notwendig muss die wissenschaftliche Betrachtung das sprachlich Verbundene nach seinen verschiedenen Seiten auseinander legen, um es bekannt zu machen, und notwendig muss in der Regel die psychologische Radix, müssen die bei der Gedankengestaltung wirkenden psychische Faktoren angezeit sein.<sup>19</sup>

Gröbers syntaktische Theorie war die methodologische Basis eines der ersten Beispiele der modernen Stilistik: Karl Vosslers Analyse des Stiles der *Vita* von Benvenuto Cellini (1899). Der Untertitel erklärt die Perspektive des Aufsatzes: *Versuch einer psychologischen Stilbetrachtung*. Das ehrgeizige Ziel des Beitrages war, die syntaktische Theorie Gröbers auf ein konkretes literarisches Beispiel anzuwenden. Unter einem psychologischen Gesichtspunkt analysierte Vossler einige traditionelle rhetorische Kategorien (Rectio, Permutatio, Pleonasmen, Ellipsen) als Zeichen der Seele von Cellini. Sprachwissenschaft, Rhetorik und Psychologie gingen Hand in Hand. Natürlich entsprach diese Psychologie nicht der Psychoanalyse von Sigmund Freud, sondern der deutschen Psychologie der letzten Jahrzehnte des 19. Jahrhunderts, nämlich Wilhelm Wundts *Grundriss der Psychologie*. Infolgedessen wurde hier die Individualität als eine psychologische Entität beschrieben, die sich sprachlich zeigt und sprachwissenschaftlich erforscht werden kann:

Die Individualität eines Schriftstellers hat sich in seinem Stil zu offenbaren, und zwar nach verschiedenen Seiten hin, entsprechend den Seelenvermögen, welche die Psychologie unterscheidet. Somit erhalten wir A eine verstandesmäßige Seite des Stils und B eine gefühlsmässige. Das intellektuelle Vermögen wieder lässt sich zerlegen in A I ein analytisch-logisches und A II ein synthetisch-künstlerisches. Dem ersteren entspricht im sprachlichen Ausdruck A I die objektive Rede, dem letzterem A II die subjektive, die bewusst oder unbewusst künstlerische, oder wie sie Gröber schlechtweg nennt: die affektische Rede. Zugleich aber enthält die affektische Rede den Ausdruck für das Gefühlsvermögen, Ethos und Pathos, so dass sie einer doppelten Betrachtung unterliegt: nach der künstlerisch-intellektuellen Seite hin fragen wir (A II): über welche Mittel verfügt der Schriftsteller, um seine Gefühle auszudrücken? und bei Untersuchung seines Gefühlsvermögens fragen wir (B): welcher Art sind die Gefühle und Affekte, die seinem Ausdruck zu Grunde liegen?<sup>20</sup>

<sup>19</sup> Gustav Gröber, „Methodik und Aufgabe der sprachwissenschaftlichen Forschung“, in *Grundriss der romanischen Philologie*, Bd. I (Straßburg: Trübner, 1888), 209–50, hier 213–4.

<sup>20</sup> Karl Vossler, „Benvenuto Cellini's Stil in seiner Vita: Versuch einer psychologischen Stilbetrachtung“, in *Beiträge zur romanischen Philologie: Festgabe für Gustav Gröber* (Halle: Niemeyer, 1899), 414–51, hier 420–1.

Nach dem Erscheinen der Analyse von Cellinis *Vita* diskutierten Vossler und Benedetto Croce – sowohl in Artikeln<sup>21</sup> als auch in ihrem reichen Briefwechsel<sup>22</sup> – über die Nützlichkeit und die ästhetische Legitimität der syntaktischen Theorie von Gröber. Croce schrieb gerade seine *Estetica come scienza dell'espressione e linguistica generale*, die 1902 erschien und wesentlich für Vossler (sowie für viele anderen Literatur- und Sprachwissenschaftler) war. Die Identifikation der Sprachwissenschaft mit der Ästhetik brachte eine theoretische Delegitimierung der positivistischen Sprach- und Literaturwissenschaft mit sich sowie der traditionellen rhetorischen Kategorien. Grammatik, Rhetorik und Stilistik können praktisch und didaktisch nützlich sein, aber sie waren theoretisch nicht gültig. Für viele Forscher hatte Croces Ästhetik eine revolutionäre und sozusagen befreiende Bedeutung: Croce drückte die Unzufriedenheit mit den positivistischen Methoden aus, die viele Literatur- und Sprachwissenschaftler um die Wende des 19. und 20. Jahrhunderts fühlten.

Die Stilistik von Karl Vossler wurde wesentlich von Croces ästhetischen Theorien beeinflusst. Das Verhältnis zwischen Vosslers stilistischer Theorie und Croces *Estetica* ist sehr komplex. Bei diesem wesentlichen Thema der modernen Stilistik kann ich mich daher hier nicht aufhalten. Was mich interessiert ist, dass Vossler in Croces neoidealistischer Ästhetik eine theoretische Legitimation der Individualität sehen konnte. Dabei handelte es sich jedoch nicht um die psychologische Individualität, die Vossler, via Gröber, aus Wundt und der deutschen Psychologie ableitete, sondern um eine philosophisch fundierte Individualität, deren Wurzeln in der idealistischen Philosophie liegen. In der langen und vielfältigen Entwicklung seiner Philosophie vertiefte Croce seine Idee des Individuums, das in seinem philosophischen System eine wichtige Rolle spielte.<sup>23</sup> für Vossler und die Stilistik war vor allem die *Estetica* der ersten Jahre des 20. Jahrhunderts wichtig. Hier finden

<sup>21</sup> Vgl. Benedetto Croce, „Di alcuni principî di sintassi e stilistica psicologiche del Gröber“, *Atti dell'Accademia Pontaniana* 29 (1899), danach in Benedetto Croce, *Problemi di estetica e contributi alla storia dell'estetica italiana* (Napoli: Bibliopolis, 2003), 141–50 und Vosslers Besprechung in *Literaturblatt für germanische und romanische Philologie* 21 (1900): 143–4. Vgl. auch Benedetto Croce, „Le categorie rettoriche e il Prof. Gröber.“ *Flegrea* 2 (1900): 83–8 (danach in *Problemi di estetica*, 151–6).

<sup>22</sup> *Carteggio Croce-Vossler 1899-1949*, hrsg. von Emanuele Cutinelli Rendina (Napoli: Bibliopolis, 1991).

<sup>23</sup> Vgl. Marcello Mustè, „Individuo e individualità nel pensiero di Croce“, in *Croce e Gentile: la cultura italiana e l'Europa* (Roma: Istituto della Enciclopedia Italiana, 2016), 296–302, [http://www.treccani.it/enciclopedia/individuo-e-individualita-nel-pensiero-di-croce\\_%28Croce-e-Gentile%29/](http://www.treccani.it/enciclopedia/individuo-e-individualita-nel-pensiero-di-croce_%28Croce-e-Gentile%29/).

wir die Bestätigung, dass am Anfang des neuen Jahrhunderts die Sprachwissenschaft und die Ästhetik ein gemeinsames Interesse für die Individualität hatten:

La Linguistica ha scoperto anch'essa il principio dell'individualità irriducibile del fatto estetico, allorché ha affermato che la parola è il realmente parlato, e che non vi sono due parole veramente identiche; distruggendo così i sinonimi e gli omonimi, e mostrando l'impossibilità di tradurre davvero una parola in un'altra, dal così detto dialetto alla così detta lingua o dalla così detta lingua materna alla così detta lingua straniera.<sup>24</sup>

Nach der Analyse von Cellinis *Vita* entwickelte Vossler ein von Croce beeinflusstes, aber wesentlich unabhängiges sprach- und literaturwissenschaftliches Denken, das sich vor allem mit der Opposition zwischen Individualität und Gesamtheit in der Sprache beschäftigte. Im 1905 geschriebenen Buch *Sprache als Schöpfung und Entwicklung*, das mit *Positivismus und Idealismus in der Sprachwissenschaft* (1904) eine wichtige Wende in Vosslers wissenschaftlichem Werdegang bedeutete, lesen wir:

Sprache ist zunächst eine rein theoretische, anschauende und individuelle Tätigkeit, also Kunst. Als solche hat sie überhaupt keine Entwicklung, sondern wird von jedem Individuum, das einen geistigen Eindruck zum sprachlichen Ausdruck bringt, immer von neuem und immer wieder anders erzeugt. Demnach müssen die sprachlichen Ausdrucksformen zunächst ohne Bezugnahme aufs Praktische, d. h. als Kunst oder als Schöpfung, d. h. theoretisch-historisch oder ästhetisch betrachtet werden.

Insofern aber die Sprache zum praktischen Verkehr der Individuen untereinander verwendet wird, ist sie nicht mehr individuelle, sondern kollektive Schöpfung, nimmt auf die empirische Wirklichkeit Bezug, fixiert sich und wandelt sich je nach den Kulturbedürfnissen der Sprachgemeinschaft und muss deshalb in zweiter Linie nicht mehr als Schöpfung, sondern als Entwicklung betrachtet werden.<sup>25</sup>

Das war nicht nur ein theoretisches Programm, insofern Vossler diese doppelseitige Perspektive auf die Sprache mit praktischen Beispielen konkretisierte. Schon in *Sprache als Schöpfung* wurde diese Dichotomie in zwei verschiedene Forschungsfälle exemplifiziert: einerseits analysierte Vossler den individuellen Stil eines Autors mittels des sorgfältigen *close reading* eines einzelnen Textes (nämlich die berühmte Fabel von La Fontaine, *Le Corbeau et le*

<sup>24</sup> Ich zitiere aus der dritten Auflage: Benedetto Croce, *Estetica come scienza dell'espressione e linguistica generale: teoria e storia*, terza edizione riveduta (Bari: Laterza, 1908), 166.

<sup>25</sup> Karl Vossler, *Sprache als Schöpfung und Entwicklung: eine theoretische Untersuchung mit praktischen Beispielen* (Heidelberg: Winter, 1905), 19.

Renard); andererseits studierte er die geistesgeschichtliche Entwicklung der französischen Sprache als eine kollektive Entität.<sup>26</sup>

Diese komplexe Beziehung zwischen dem individuellen Stil eines Autors und dem gesamten Stil eines Volks charakterisierte das ganze Werk von Vossler. Nach einigen Jahren erklärte er diese Idee in einem Beitrag, dessen Titel sehr bedeutungsvoll ist: *Der Einzelne und die Sprache*. Er wurde 1919 in Zeitschrift *Logos* publiziert und danach im wichtigen, aber heutzutage ziemlich vergessenen Buch *Gesammelte Aufsätze zur Sprachphilosophie* (1923) erneut publiziert. Hier lesen wir:

An und für sich hat die Sprache weder Regeln noch Ausnahmen, weil sie zwischen beiden hin und hergeht, weil sie weder dem Einzelnen noch der Gemeinschaft ganz gehört, sondern die bewegliche Mittlerin zwischen dieser und jenem spielt. Aber eben kraft ihrer Mittlerrolle kann die Sprache dem Einzelnen ein völlig anderes Gesicht zeigen als der Gesamtheit. Sie kann dem Einzelnen Dinge erlauben, zu denen sie sich niemals der Gesamtheit als solcher gegenüber hergibt.<sup>27</sup>

Vosslers Fall beweist die Vielfältigkeit des Begriffes der Individualität in der Stilistik der ersten Hälfte des 20. Jahrhunderts. Im schon zitierten letzten Beitrag der *Stilstudien* Spitzers erhält man eine eindeutige methodologische Erklärung:

Die geschilderte Forschungsrichtung scheint mir nicht nur von Bedeutung für die Wortkunst, sondern auch für die Sprachwissenschaft, denn bekanntlich ist die Allgemeinsprache nichts als ein Durchschnitt von Individualsprachen, die Grammatikalisierung verschiedener Sprechakte [...], alle Neuerung geht von schöpferischen Einzelnen aus, nihil est in syntaxi quod non fuerit in stylo. Syntax, ja Grammatik sind nichts als gefrorene Stilistik [...]. Zwei lateinische Sätze möchte ich als Mottos über meine Stilforschung setzen: „individuum non est ineffabile“ – die individuelle Stilsprache läßt sich beschreiben, eben durch sprachwissenschaftliche Methoden – und: „oratio vultus animi“ – diese Stilsprache ist die biologisch notwendige Auswirkung der Individualseele (man könnte auch für den zweiten Satz: „Le style c'est l'homme“ sagen,

<sup>26</sup> Die Titel und die Kollokation dieser Studien sind innerhalb des Buches schon bedeutungsvoll: der erste Beitrag, „Ein Beispiel ästhetischer Stilanalyse“, befindet sich im zweiten Kapitel, *Sprache als Schöpfung*, während der zweite, „Beispiele aus der Entwicklungsgeschichte des Französischen“, im dritten Kapitel, *Sprache als Entwicklung*, zu finden ist.

<sup>27</sup> Karl Vossler, „Der Einzelne und die Sprache“, *Logos* 8 (1919–1920): 266–302, danach in Karl Vossler, *Gesammelte Aufsätze zur Sprachphilosophie* (München: Hueber, 1923), 152–209, hier 155–6.

wenn Buffon an der Originalstelle seines Discours sur le style sein *le style est l'homme même* nicht in ganz anderem Sinne gebraucht hätte).<sup>28</sup>

Die zwei Mottos fassen die hauptsächliche Charakteristik der Stilistik der ersten Jahrzehnte des 20. Jahrhunderts zusammen: nicht nur das unlösbare Verhältnis zwischen Sprach- und Literaturwissenschaft, sondern auch die (problematisch) biologisch-psychologische Perspektive auf den Stil.

Im Mittelpunkt dieser verschiedenen Disziplinen stand die Individualität. Am Ende eines anderen Beitrages der *Stilstudien* (*Inszenierende Adverbialbestimmung in der neueren französischen Literatur*) liest man, dass der Protagonist der literatur- und sprachwissenschaftlichen Forschung Spitzers das Individuum, oder, besser zu sagen, das schöpferische Individuum ist:

Das „Wie ich es sehe“ des Künstlers bestimmt sein „Wie ich es sage“. Aus dem Literaturstil geht dann die Ausdrucksweise in die Allgemeinsprache über, wie denn stets das einsam und kraft einer inneren Notwendigkeit von einzelnen Errungene zur bequemen Beute der ewig reproduktiven Gemeinschaften wird. Bei der Erforschung sprachlicher Tatsachen wird es daher in Zukunft notwendig sein, vom Habituellen aufs Okkasionelle und von diesem auf die Neuerungen des schöpferischen Individuums zurückzuschließen.<sup>29</sup>

Wie Starobinski geschrieben hat, stehen der Abstand des Individuums gegen seine eigene Sprache und seine Aussöhnung durch die Vermittlung des Textes im Mittelpunkt der stilistischen Forschungen von Leo Spitzer:

Le style n'est [...] ni le particulier pur, ni l'universel, mais un particulier en instance d'universalisation, et un universel qui se dérobe pour renvoyer à une liberté singulière. Telle est du moins l'acception équilibrée que Spitzer confère à la notion de style. Elle implique la révolte de l'individu, et sa réconciliation par l'intermédiaire de l'œuvre. Reconnaître un écrivain à travers son style, c'est reconnaître à la fois une conscience qui s'affirme dans son irréductible recel, et l'énergie d'une parole qui traverse l'intervalle.<sup>30</sup>

In seinem ziemlich berühmten Essay über Jules Romains, der in der Zeitschrift *Archivum Romanicum* 1924 erschien, wollte Spitzer die expressionistische Poesie des französischen Schriftstellers analysieren. Seine sprach- und kulturwissenschaftliche Forschungsfrage war: wie kann ein französischer Dichter seine Sprache (die rationale und klare Sprache des Klassizismus, die

<sup>28</sup> Spitzer, „Wortkunst und Sprachwissenschaft“, 516–20.

<sup>29</sup> Leo Spitzer, „Inszenierende Adverbialbestimmung in der neueren französischen Literatur“, in *Stilstudien*, 125–65, hier 164.

<sup>30</sup> Starobinski, *Leo Spitzer et la lecture stylistique*, 81

Sprache der *clarté*) in eine expressionistische, sozusagen deutsche, Sprache verwandeln? Diese Frage hatte zwei einleitende Voraussetzungen:

1. einerseits, ein sehr verbreitetes kulturgeschichtliches *cliché*: grammatisch und kulturell war das Französische eine rationale, impressionistische, klare Sprache, während das Deutsche das Gegenteil, die perfekte Sprache des Expressionismus war;
2. andererseits, das Verhältnis zwischen Schriftsteller und Sprache als eine dynamische Dialektik: um seine neuen geistigen Ansprüche auszudrücken, sollte Romains mit seiner Sprache sozusagen einen Kompromiss eingehen.

Romains war kein Deutscher, aber er wollte wie ein deutscher expressionistischer Dichter schreiben. Jeder expressionistische Dichter hat drei Möglichkeiten, um seine Sprache expressionistisch zu benutzen:

Jeder Künstler des Wortes erbt die Worte der gegebenen Sprache [...]. Drei Wege bieten sich dem expressionistischen Wortkünstler, um das herkömmliche Sprachbild zu erschüttern: Auflösung der Syntax und Zusammenballung neuer Worte (parallel der Auflösung der malerischen Komposition), Erweiterung der Wortbedeutung (parallel den Dimensionsverschiebungen der neueren Maler), kurz gesagt sprachlicher Kubismus oder sprachlicher Expressionismus.<sup>31</sup>

Romains' Strategie war die dritte: die „Erweiterung der Wortbedeutung“. Im Unterschied zu den deutschen expressionistischen Dichtern konnte er seine expressionistische „Vergewaltigung des Realen“ nur durch eine Verstärkung seiner Muttersprache (des klaren und immer noch klassizistischen Französisch) erlangen.

Der Sprachforscher hat die Pflicht, das Neue an Ausdrucksmöglichkeiten abzustecken, das der Ausdrucksünstler der französischen Sprache gebracht hat: jene inbrünstige Durchseelung und bewegungszitternde Durchwühlung, die die sprachliche Spiegelung des Unanimismus ist. Und immerhin hat Romains als formbeherrschter Romane es verstanden, die Vergewaltigung des Realen, die im Wesen seiner Dichtergeneration liegt, nicht in einem vergewaltigten, sondern bloß in einem ausdrucks gesteigerten Französisch verdichtet zu haben.<sup>32</sup>

Im Mittelpunkt dieses Beitrages stand also das dialektische Verhältnis zwischen der grammatischen und sprachgeistigen Charakteristik einer Sprache

<sup>31</sup> Leo Spitzer, „Der Unanimismus Jules Romains' im Spiegel seiner Sprache (Eine Vorstudie zur Sprache des französischen Expressionismus)“, in *Stilstudien*, 208–300, hier 287–8.

<sup>32</sup> Spitzer, „Der Unanimismus Jules Romains“, 298–9.

(das Französische) und einem Schriftsteller, der neue poetische sowie kulturelle Ansprüche ausdrücken wollte.

Das ist nur ein Beispiel dafür, dass die Individualität in der Stilistik Spitzers der 20er und 30er Jahre eine wesentliche Rolle spielte. Spitzer spiegelte eine allgemeine Tendenz der Literatur- und Sprachwissenschaft der ersten Jahrzehnte des 20. Jahrhunderts wider, die den Stil und die Individualität in den Mittelpunkt der allgemeinen Krise des Positivismus rückte. Wenn Spitzer vielleicht der interessanteste Stilkritiker der ersten Hälfte des 20. Jahrhunderts ist, sollte man ihn und seine Stilistik in einen historisch und kulturell weiteren Kontext stellen. Wie der Beitrag hier nur andeuten konnte, leistet die historische Perspektive auf die erste Phase der modernen Stilistik einen nützlichen Beitrag, um die faszinierende Komplexität der Stilistik und deren praktische und theoretische Probleme zu begreifen.





# “Dice Bernardo a Cristo que un Portero al Papa”

## Indagaciones Multi-Autorales a partir de la reescritura hispánica de un soneto de Burchiello en un manuscrito florentino

Alan Joaquín Pérez Medrano (Berlin)

**RESUMEN:** El manuscrito Magliabecchiano VII, 353 de la Biblioteca Nacional de Florencia en su variedad antológica contiene un texto peculiar en lo que respecta al *corpus* de Burchiello (1409–1449) quien encarna una de las más importantes figuras de la poesía cómica italiana y del *Quattrocento* toscano. El mencionado texto testimonia una reescritura poética, en lengua española, de uno de los sonetos del poeta florentino (Fucilla, 1960). Si bien la imitación de dicho estilo es prolífica en diferentes momentos de la literatura italiana, será interesante observar y analizar en qué modo los textos “burchiellescos” se relacionan también con el contexto hispánico y en el contexto de la discusión teórica general respecto las funciones autorales (Jannidis, 2003) considerando también las implicaciones inherentes al análisis contemporáneo de textos antiguos.

**ZUSAMMENFASSUNG:** Das Manuskript Magliabecchiano VII, 353 der Nationalbibliothek von Florenz enthält innerhalb seiner vielfältigen Anthologien einen besonderen Text des Korpus von Burchiello (1409–1449) betreffend, der eine der wichtigsten Figuren aus der italienischen Komik und des toskanischen Quattrocento aufleben lässt. Der erwähnte Text zeugt von einer poetischen Neufassung eines Sonetts des florentinischen Dichters in spanischer Sprache (Fucilla, 1960). Zwar ist die Nachahmung dieses Stils durch die italienische Literaturgeschichte hinweg bekannt, aber es lohnt doch zu analysieren wie sich die ‚burchiellesken‘ Texte auf den hispanischen Kontext beziehen und wie in diesem Kontext die Frage der Autorfunktion (Jannidis, 2003) diskutiert werden kann. Dabei sind die zwangsläufigen Einschränkungen zu beachten, die eine aktuelle Analyse von alten Texten mit sich bringen.

**PALABRAS CLAVE:** Burchiello; Poesía cómica italiana; Quattrocento italiano; Magliabecchiano VII; Autorialidad; Jannidis, Fotis; Fucilla, Joseph

**SCHLAGWÖRTER:** Burchiello; Italienische Komik; Quattrocento; Magliabecchiano VII; Autorfunktion; Jannidis, Fotis; Fucilla, Joseph

En 1960 en la revista *Romance Notes* Joseph Fucilla presentaba un estudio donde fue puesta en análisis la breve tradición textual (compuesta de seis versiones) del soneto con el incipit “Dixo un portero al Papa, aquí ha llegado”. Todos y cada uno de los textos de esta pequeña tradición fueron reconocidos como anónimos. En realidad, son marcas externas las que nos proporcionan

información adicional sobre cada una de las versiones, dicha información trata incluso de la suerte de algunos de los manuscritos, pero más concretamente proporciona pistas sobre cierto carácter antológico en la configuración general de los manuscritos que contienen alguna versión del soneto en cuestión. Es interesante la presencia de títulos como por ejemplo: “Bariedad de Sonetos recocidos de diferentes autores por Ignacio de Toledo, año 1627” o también otras marcas de posteriores investigaciones filológicas, por ejemplo, los “273 Sonetos” que habrían sido agrupados por el filólogo francés Foulché-Delbosch. Así pues, la pertenencia de nuestro soneto a un programático *corpus* de Autor no es una de sus características. La única información relativa a la autoría del soneto la brinda el manuscrito perteneciente al fondo Magliabechiano (Manuscritto VII, 353) de la Biblioteca Nacional de Florencia.<sup>1</sup> Mediante una nota al margen, igualmente anónima, se reconoce en el rimador toscano Domenico di Giovanni *detto il Burchiello* la “fuente” original de nuestro soneto, concretamente con el soneto “Dice Bernardo a Cristo: E’ c’è arrivato”. En su ensayo Fucilla presentó interesantes hipótesis respecto a la génesis de la pequeña matriz textual en castellano, de entre ellas resulta significativo mencionar que el soneto del fondo Magliabechiano, junto con aquel de la Biblioteca Matritense, pudieran constituir la base de la tradición.<sup>2</sup> Además Fucilla confrontó la matriz de sonetos en castellano con su correlativa fuente italiana:

CLXXX (FD 180: L1 L2 Mg7 Ag’ Mrc2)

Dice Bernardo a Cristo: “E’ c’è arrivato,	
Signor mie caro, un peccator cotale:	2
arso egli ha chiese e rubato spedale,	
uomo micidiale egli è sempre stato;	4
e tutto il tempo suo t’ha bestemmiato	
sforzò la madre e fatto ha ogni male,	6
uccise un prete la notte di Natale.	
Potrebbesi purgare il suo peccato?”	8

<sup>1</sup> “The poem in question is to be found in the Magliabechiano Manuscript, VII, 353, in the Biblioteca Nazionale in Florence, substantial anthology of a miscellaneous character. An anonymous annotator has, as a matter of fact, detected the Burchiello source and so indicates it in a marginal notation”. Joseph Fucilla, “Variations of a Spanish Adaptation of a Burchiello Sonett”, *Romance Notes* 1, No. 2 (Spring, 1960): 146–50.

<sup>2</sup> Cfr. Fucilla, “Variations”, 150.

A san Bernardo rispondeva Cristo:  
 “Non per viaggi né per digiunare, 10  
 né per orare o piangere o star tristo  
 ma digli che se moglie vuol pigliare 12  
 il porrò allato a Giovanni Battisto,  
 se questa pena in pace vuol portare. 14

Bernardo non pensare,  
 ch’a sufferir la moglie ell’è gran doglia, 16  
 perché ella stessi non sa che si voglia”.<sup>3</sup>

.....

Dixo un Portero al Papa, aquí ha llegado,  
 Señor, un hombre de perversa vida, 2  
 traydor, ladrón, pirata y omicida,  
 que de su hermana y madre mal ha usado. 4

El Santo Sacramento ha profanado  
 y a dos Obispos les quitó la vida, 6  
 a dos Papas veneno en la comida  
 por pocos intereses les a dado. 8

A echado las reliquias en el fuego,  
 haze burla de todo el invisible, 10  
 ¿podráse perdonar tanta insolencia?

Respondió el Papa, sí, con tal que luego 12  
 se case que es la cosa más terrible  
 que se puede dar en penitencia.<sup>4</sup> (fol. 9v) 14

El parentesco entre los dos ejemplares italiano y español resulta interesante en vista de que (no obstante sus adaptaciones culturales y temporales) mantienen cercanía en los contornos tradicionales del retrato de un “personaje ruin” y en la temática de corte misógino. Por otra parte las variaciones en la matriz textual hispánica, como puede observarse en nuestro apéndice, se relacionan internamente de un modo relativamente unitario, no obstante las variaciones naturales de la tradición manuscrita, permitiendo así que cada uno de los textos hispánicos sean considerados parte de una matriz prototípica. Aparecen variaciones que resultarían más significativas en el sentido de

<sup>3</sup> Burchiello, *I sonetti del Burchiello* (Bologna: Commissione per i testi di lingua, 2000), 174.

1: Dixe Mg7; E’ m’è venuto (cfr. v.4) L1 Mrc2 3: chiesa L1; OM. et L1 4: sempremai (=) Mrc2; suto (cfr. v. 1) L1 Mrc2 5: tempo tuo Ag’ 6: et à facto Mg7 10: viaggio Mrc2; o per d. L1 11: non per o. L2 Mrc2; et piangere et star t. FD 12: c’vuol Mrc2 14: e’ vuol Mrc2 16: OM. a L1.

<sup>4</sup> Fucilla, “Variations”, 146.

que podrían evidenciar una deliberada modificación contextual y no sólo un cambio provocado por las condiciones materiales de una tradición manuscrita, por ejemplo, el uso de “datario” en vez de “portero” en el verso primero o también en el verso segundo la aparición de “pirata” dentro de los adjetivos aplicados al “hombre de perversa vida”, de quien es posible observar una gradación en los crímenes cometidos, pasando de la violación (en una versión más relacionada con el correspondiente italiano) hasta el asesinato de miembros de la propia familia. Al respecto Fucilla comenta:

The chief distinction between the two sonnets just cited is that in his remanient the Spanish imitator softens the tone of the original a bit by transferring its locale from heaven to earth, that he multiplies the crimes committed by the sinner, and that he does not make use of a coda. Otherwise, the imitation successfully competes with its Italian exemplar in the verve which is infused into the half-humorous misogynistic diatribe.<sup>5</sup>

Estructuralmente las variaciones entre la versión italiana y la castellana son más que lógicas; es decir la exclusión de la *coda* (un rasgo a característico de la poesía cómica del siglo XV italiano) y también el cambio de la estructura de la rima (de ABBA ABBA CDC DCD DEE que es el esquema típico *burchiellesco* a ABBA ABBA CDE CDE). El hecho de que el glosador anónimo haya reconocido la fuente *burchiellesca* parecería resolver sin mayor problema los misterios autorales sobre la proveniencia de este soneto en su versión castellana, sin embargo, nada hay de simple en tal relación. Antes de comentar la problemática autorral que define al caso de Burchiello me gustaría hacer una hipótesis sobre el modo en que nuestro glosador habría reconocido la fuente *burchiellesca*.

Si bien Fucilla refiere la poco probable influencia de Burchiello en España señala la amplia posibilidad de que ediciones italianas estuvieran al alcance de hispanistas en el Renacimiento y en el Siglo de Oro:

Although there is apparently no published evidence on the influence of Burchiello in Spain, the large number of sixteenth century Italian editions of his poems that were available to Spaniards of the Renaissance and Golden Age suggest the possibility that eventually something of considerable value will turn up.<sup>6</sup>

Retomando la hipótesis sobre nuestro glosador, podemos suponer que el reconocimiento de cierto repertorio temático y genérico (después reconocido en el texto hispánico) le habría posibilitado reconocer cierto “estilo de autor”

<sup>5</sup> Fucilla, “Variations”, 147.

<sup>6</sup> Fucilla, “Variations”, 146.

y por ende la fuente italiana. Tal proceder podría corresponder a las características del concepto *Autorbild* de Fotis Jannidis:

Unter einem Autorbild verstehe ich die Summe alles Wissens eines Lesers über einen realen Autor. Der Begriff Wissen ist hier sehr weit zu nehmen; er umfaßt alle Annahmen und Spekulation. Zu diesem Wissen gehören auch die Erinnerung an die Lektüreerfahrung und die Rückschlüsse auf den Autor aufgrund seines Textes. Die Organisation dieses Autorwissens ist eine spannende Frage. Es ist offensichtlich personal organisiert und hat narrative und typologische Aspekte. Zwei Bezüge scheinen mir besonders relevant zu sein. Zum einem ist die Organisation des Wissens über einen bestimmten Autor immer abhängig von den allgemeineren Annahmen darüber, wie Autoren und Künstler Werke schaffen.<sup>7</sup>

La imagen de Burchiello como “Autor-Fuente” que el glosador habría identificado partiría entonces de su conocimiento como lector basado en su experiencia con las rimas del toscano, generada dentro de los contornos de una edición *burchiellesca* a partir del siglo XVI;<sup>8</sup> habrá que aclarar que las primeras ediciones de las rimas de Burchiello en el contexto italiano se definían por su poca claridad atributiva ya que la firma “Burchiello” muchas veces no plantea fronteras claras con los otros participantes “imitadores” del estilo. La condición de los manuscritos no es menos problemática ya que es verificable un carácter sumamente misceláneo y definido por la intercambiabilidad de los textos.

En tal estado de la cuestión nuestro glosador habría reconocido de algún modo el soneto como parte del *corpus* de Burchiello, aun cuando en una lectura global de los sonetos de Burchiello podría percibirse que los rasgos del texto en cuestión se encuentran alejados de la configuración más típica que encarnaría la esencia más definitoria del fenómeno de la poesía de Burchiello en el siglo XV. Dicho de otra manera, es posible que en nuestro soneto sea reconocida la fuente *burchiellesca*, paradójicamente sin que sean evidentes los rasgos mayoritariamente característicos del estilo de Burchiello. En ese sentido debemos observar que el conocimiento del lector-glosador sobre un “autor real” no sólo se define necesariamente por la imagen ligada a un personaje histórico del siglo XV, como es el caso de Domenico di Giovanni (quien tradicionalmente encarna la figura autoral de Burchiello), sino que se expande proyectando diversos “espectros autorales” en los contornos de una

<sup>7</sup> Fotis Jannidis, “Autor, Autorbild, Autorintention”, *Editio* 16 (2003): 27.

<sup>8</sup> Quizás también pudo suceder con algún testimonio manuscrito, esta es una situación que, sin embargo, parece menos probable.

tradición *comico realistica*. Antes de continuar habrá que hacer dos precisiones respecto al nombre de Domenico di Giovanni y el apodo Burchiello. Los datos históricos sobre Domenico di Giovanni, son más bien parcos y si bien circunscriben su existencia a la Florencia de la primera mitad del siglo xv, aluden más precisamente a su profesión de barbero que a su actividad como poeta. Existen muy pocas coordenadas historiográficas sobre aquel sujeto marginal y bizarro sobre el que el imaginario literario italiano ha erigido la imagen de Burchiello. Sobre el origen del apodo surgen varias hipótesis, una de las cuales resulta relevante en vista de que emerge de las características estilísticas del *corpus* a él circunscrito. El término *burchiello* hace referencia a una pequeña embarcación mercantil del tiempo, cuya característica principal sería estar cargada de mercancías de naturaleza variada y no precisamente ordenada. El término *alla burchia*, ya en el contexto de la creación literaria, referiría alegóricamente a este proceder. El estilo de Burchiello implica primordialmente la composición poética con variados e incluso contrapuestos recursos semánticos dispuestos en una articulación discursiva poco convencional, cuya cohesión sintáctica produce un efecto cómico y de extrañamiento. Se habla continuamente de la poética de la paradoja significativa, del *non-sense burchiellesco*, de una poética material de la fragmentariedad y lo surreal. En todo caso un estilo poético peculiar en la tradición poética italiana en general y en la tradición cómico burlesca en particular. A manera de ejemplo de tal estilo presentamos el siguiente soneto:

El dispota di Quinto e 'l gran Soldano	
e trentasette schiere di pollastri	2
fanno coniar molti fiorin novastri	
come scrive el psalmista nel Prisciano;	4
e dicesi nel borgo a san Friano	
ch'egli è venuto al porto de' Pilastrì	6
una galea carica d'impiastrì	
per guarir del catarro Monte Albano.	8
Mille franciosi assai bene incaciati,	
andando a Valembrosa pe'cappegli	10
furon tenuti tutti isvemorati;	
Toian gli vide e disse: "Végli, végli!	12
E' non son dessi, el bagno g'ha scambiati,	
o e'gli à barattati in alberegli".	14

Allora e fegategli  
 gridaron tutti quanti “Cera, cera”,  
 e l'aringhe s'armoron di panziera.<sup>9</sup> 16

Nos limitaremos a comentar el primer cuarteto. “El despota di Quinto” (que podría ser relacionado con Cosimo dei Medici) es un epíteto en sí paradójico dadas las reducidas dimensiones de la región toscana de Quinto, sin embargo, es relacionable a un “gran Sultán” no siendo el caso de las “treinta y siete tropas de polluelos” (vv. 1, 2), el curioso grupo de animales y monarcas acuñaría (*fanno coniar*) moneda nueva (*fiorin novastri*). El hecho sería relatado por *el psalmista* (el Rey David, antonomasia de la tradición bíblica) en el *Prisciano* (una de las más importantes gramáticas medievales). El efecto cómico de esta última adjudicación podría corresponder análogamente a la frase: “Como dice Homero en el Corán”. La secuencia sujeto plural + verbo + complemento es lógica en un plano sintáctico, pero traiciona las expectativas convencionales mediante la acumulación de referentes alejados semánticamente creando paradojas e incongruencias. En el cuarto verso de la estrofa podemos también observar los referentes de Burchiello que adoptan como campo instrumental alusiones bíblicas (a autores, personajes y libros de esta tradición) y de igual

<sup>9</sup> “1 Citato insieme al Sultano di Babilonia, è forse Cosimo de’ Medici il Vecchio (la villa medicea di Castello, oggi sede dell’Accademia della Crusca, è appunto a Quinto alto, a nord di Firenze) 2 numerale iperbolico [...] 4 psalmista. David, ma il termine è attestato nel traslato furbesco ‘saccente’ [...]; il *Prisciano* è l’autore delle *Institutiones grammaticae*, testo fondamentale per tutto il Medioevo 5 borgo san Frediano, di là d’Arno 6 via de’ Pilastri, presso il mercato di Sant’Ambrogio 7 le “galee” facevano la spola fra la Toscana e le Fiandre, cariche di merci [...]; *impiastri*: cataplasmi, spesso smerciati da impostori e ciarlatani [...] 8 Monte Albano: il monte che domina Empoli, ma anche la sede mitica dei carolingi. 9 per *francioso* è attestato un traslato osceno “membro” [...]; sull’associazione con il cacio [...] 10 Vallombrosa: sede del famoso monastero, sarà qui nome parlante di tipo boccaccesco [...] 11 “furono presi tutti per matti” (il termine *smemorato* o *svemorato*, qui in forma prostetica, vale propriamente “fuori dalla memoria, impazzito” [...] 12 *Toian*: personificazione della rocca di Toiano, conquistata dai Fiorentini, o forse per metonimia, la campana della rocca, trasportata a Firenze [...], *Végli, véglil*: guardali, guardali (ma anche “attenzione” [...]; “voce propria dei “bambini” che l’usano quando alcuno di loro pericola di cascare”) 13–14 *E’ non son dessi*: il tema dello sdoppiamento o perdita dell’identità era assai popolare nel primo Quattrocento, quando vennero prodotti il diffuso cantare *Geta e Birria* e la *Novella del Grasso legnaiuolo* [...]; sottinteso il soggetto *mille franciosi*, che il lavaggio, ma propriamente il bagno termale, ha ora fatto diventare lucidi (*albergli* erano detti dei vasi allungati di terracotta invetriata in uso fra gli speciali) [...] 15 i *fegategli* occorrono con funzioni allusive anche molto diverse [...] quello di “cinedi” sembra qui preferibile [...] 17 la *panziera* è un panciotto di maglia metallica usato in battaglia [...]” Burchiello, *I sonetti del Burchiello* (Torino: Einaudi, 2004), 3–4.

modo medievales; en otros sonetos.<sup>10</sup> Los efectos cómicos y la fragmentación a varios niveles son un rasgo también característico del estilo de Burchiello presente incluso a nivel de estrofas dentro de un mismo soneto. El poema “El despota di Quinto e ‘l gran Soldano” es un ejemplar ideal del fragmentario esquema *alla burchia* que forma parte esencial del estilo de Burchiello,<sup>11</sup> sin embargo, la cuestión autoral se podría complejizar aún más en vista de que el estilo *alla burchia* habría sido ya inaugurado antes de la existencia de Domenico di Giovanni por el también toscano Mariotto di Nardo di Cione *detto* Orcagna, cuya producción también forma parte del fenómeno global de Burchiello en el siglo xv pero que también aparece en otros códices extraños al *corpus*. Resulta importante considerar las problemáticas autorales en relación también a las condiciones materiales de los documentos que conforman el *corpus* burchiellesco, es decir, los códices y primeras ediciones cuya característica principal es la intercambiabilidad y la volatilidad en esta época de transición del texto manuscrito al texto de imprenta. Esta situación ha impuesto un reto ecdótico que en tiempos recientes ha visto inaugurada su sistematización con la primera edición crítica del Prof. Michelangelo Zaccarello en el año 2000, quien mediante el rigor filológico y la consciencia de la peculiaridad del caso literario, optó por determinar el caso de Burchiello como una tradición *vulgata* más que una tradición de autor. Respecto al soneto “Bernardo a Cristo: E’ c’è arrivato”, resulta interesante observar que éste forma parte de un grupo complicado en lo que se refiere a la radiografía de la *vulgata* elaborada por Zaccarello; en este segmento del *corpus* es posible formar hipótesis sobre la no “genuinidad” del grupo de sonetos, y en nuestro caso concreto la atribución de “Dice Bernardo a Cristo: – E’ c’è arrivato” a Niccolò Cieco en algunos manuscritos:

A questa sezione dobbiamo poi aggiungere il gruppo CLXXX; CLXXXV [...] che (secondo l’ipotesi basata sullo studio delle sequenze) rappresenta un fascicolo originariamente collocato fra CXXXI e CXXXII e successivamente spostato e trascritto a fine silloge in ordine diverso (la seriazione ‘regolare’ che passa in FD). È possibile che alla base dello spostamento ci fosse anche la consapevolezza della non genuinità di molti di quei testi: CLXXX *Dice Bernardo a Cristo: – E’ c’è arrivato* è attribuito da vari codici a Niccolò Cieco.<sup>12</sup>

<sup>10</sup> Cfr. Alan Pérez Medrano, “Burchiello: Travesía poetica” (Tesis, Universidad Nacional Autónoma de México, 2015), 9–15.

<sup>11</sup> Cfr. Michelangelo Zaccarello, “Introduzione” in *I sonetti del Burchiello* (2004), XIII.

<sup>12</sup> Michelangelo Zaccarello, “Radiografía della Vulgata Fiorentina FD” in *I sonetti del Burchiello* (2000), CXXIV.



En este grupo del *corpus* también encontramos otro caso interesante, análogamente con nuestro tema en el caso hispánico, éste involucra a Burchiello en la contraparte, ósea esta vez como “adaptador-imitador” de un soneto del poeta cómico Cecco Angiolieri contemporáneo de Dante. En esencia el proceder es muy similar al que involucra el texto hispánico y el italiano:

### Cecco Angiolieri

I' son sì magro che quasi traluco, de la persona no, ma de l'avere;	2
ed abbo tanto più a dar che avere, che m'è rimaso vie men d'un fistuco.	4
Ed èmmi sì turato ogni mi' buco, ch'i' ho po' che dar e vie men che tenere:	6
ben m'è ancora rimas'un podere che frutta l'anno il valer d'un sambuco!	8
Ma non ci ha forza, ch'i' so' 'nnamorato; ché s'i' avesse più òr che non è sale,	10
per me saria 'n poco temp'assommato.	
Or mi paresse almeno pur far male!	12
Ma com' più struggo, più son avviato di voler far di nuovo capitale. <sup>13</sup>	14

### Burchiello

I' son sí magro, che quasi traluco della persona e cosí dell'avere:	2
che s'io vo per la via, son per cadere, sí poco è l'esca, di ch'i' mi conduco.	4
Così ho io turato ogni mie buco, ch'i' non ho piú che dar, né che tenere,	6
ma ben m'è certo rimaso un podere, che frutta l'anno un bel fior di sambuco.	8
Ma non mi curo, sí sono avviato, che s'io avessi in mano il Sangredale,	10
in picciol'ora si saria fondato.	

<sup>13</sup> Cecco Angiolieri in *Poeti Giocosi del tempo di Dante*, ed. Mario Marti (Milano: Rizzoli, 1956), 211.

E d'ogni mio principio arrivo male,	12
di collo ad ogni amico i' son cascato,	
nimico mi diventa ogni uom mortale.	14
Gli ucce' che batton l'ale,	
e gli animali che son sopra la terra,	16
le bestiê fiere, ognuna mi fa guerra. <sup>14</sup>	

En este pequeño recorrido hemos observado la aparición de varias presencias autorales en nuestro fenómeno en cuestión, tanto en la tradición italiana cuanto en la hispánica, cuyas variaciones significativas se presentan a veces anónimamente y otras con la posibilidad de ser atribuidas a un nombre (Orcagna, Nicolò Ciecco, Cecco Angiolieri, etc.). En nuestra perspectiva, todas y cada una de estas presencias participan de diferentes modos en la configuración que da volumen un “Autor poliédrico”, como lo es Burchiello, a quien podríamos caracterizar como el generador de una suerte de *Ars poetica* de la intercambiabilidad, en sintonía con el carácter decididamente experimental de la primera mitad del siglo xv italiano. El concepto de Jannidis de *Autor-konfiguration* nos resulta oportuno para observar en Burchiello a dicho “Autor poliédrico” y complejo que, como hemos observado, es reconocido como una esencia autónoma y al mismo tiempo un autor relacional dentro de la tradición cómica y burlesca:

Wir sprechen von *Autorfunktion*, um eine kognitive Operation zu bezeichnen, die dem ‘Autor’ zugeschrieben wird. z.B. die Auswahl des sprachlichen oder thematischen Materials aus der Menge aller zeitgenössisch verfügbaren Wörter und Themen für ein Werk. Im Anschluß an diese Zuschreibung kann man nun als Editor wiederum weiterführende Operationen vornehmen, z.B. die Herkunft des Materials verfolgen. Zumeist werden solche Autorfunktionen nicht individualisiert zugeschrieben, sondern in typisierten Komplexen, die ihre Begründung in den historischen Künstlerbildern finden. Ich schlage als Bezeichnung solcher Komplexe den Begriff *Autorkonfiguration* vor.<sup>15</sup>

En nuestro caso, las “elecciones de los materiales lingüísticos y temáticos” realizadas por un sujeto en determinado contexto histórico (que identificamos como *Autorfunktionen*) son condensadas en Burchiello como una *Autorkonfiguration* global y porosa, para la cual, la obra poética de Domenico di Giovanni debió ser un factor determinante mas no necesariamente exclusiva. Las variaciones hispánicas presentadas en esta ocasión pueden considerarse ejemplos excepcionales, pero significativos, de transmigraciones y re-

<sup>14</sup> Burchiello, *I Sonetti* (2004), 261–2.

<sup>15</sup> Jannidis, “Autor”, 28.

elaboraciones de estilos, géneros y tópicos, cuya base es un bagaje literario-tradicional global y que fueron adaptadas a la *praxis* de una lengua y contexto literario diferentes. Dichas dinámicas podrían ser mejor entendidas en una dimensión europea al evidenciar lazos estilísticos en lo que respecta al género cómico-burlesco por un lado, pero también resultan significativas en sus especificidades contextuales en cada cultura. Los análisis a partir de las varias modalidades de presencia del “Autor”, en ese sentido, pueden también proporcionar perspectivas funcionales para atravesar y analizar fenómenos literarios de naturaleza transcultural.

**Apéndice: Comparación de versiones con base en el ensayo de Fucilla<sup>16</sup> se toma como texto prototípico la correspondiente al (Mg1) Magliabechiano VII, 353.**

Bnm1	Bnm2	Bacaa	Bnm3	Bm1	Bm2
portero	portero	a llegado	portero... a llegado	portero... a llegado	Dijo un Datario... a llegado.
malvada vida	malvada vida	...de muy mala vida,	... hombre de malvada...,	malvada vida	
ladrón, traydor, perjuro y...	traydor, ladron, herético, ...	ladrón, herético, traidor i omicida,	traidor, ladrón, erético, omisida,	ladrón, traydor, erético, omicida	... erético, homicida
...su madre y ermanas ha forçado.	que a sus padres y hermanos muerte ha dado.	...a su madre i hermanos muerte ha dado.	que a su madre y hermano muerte a dado.	que a su padre y hermanos muerte ha dado.	que a su Padre y Hermanos muerte ha dado.
El santo...,	Quatro o cinco sagrarios ha robado,	...santo sacramento a...		El sancto sacramen-to a...	El sacra-mento santo a...

Continúa en la siguiente pagina.

<sup>16</sup> Fucilla, “Variations of a Spanish Adaptation of a Burchiello Sonett”, 146–150.

<b>Bnm1</b>	<b>Bnm2</b>	<b>Baca</b>	<b>Bnm3</b>	<b>Bm1</b>	<b>Bm2</b>
a un obispo dio muerte conocida,	a su patria en son de paz dexa vendida,	y dió muerte a un obispo no entendida,	...un santo obispo le...,	y a un cardenal dió muerte no entendida,	y de su madre a sido matrizada,
...Papas ponçoña en la bebida	a tres Papas ponçoña en la...	Y a tres Papas veneno en...	y a tres Papas ponzoña en la ...,	y a... ponçoña en la...	y a...
por muy poco interese les...	por vil y baxo precio él ha echado	de su cudi-cia i interés llevado	... muy [poco] dineroles ha echado.	por muy pequeño precio les a echado	por muy poco dinero...
Ha echado...	Arrojó unas reliquias dentro un...,		A puesto las rreli-quias por el suelo,		Ha puesto mil reliquias...
tiene por burla todo lo...	hizo... lo invisible.	hace,...lo visible,	ase burlado de todo lo...	...lo invisible...	y en nada tiene todo lo...
¿Podráse...	¿Puédese... tal insolencia?	¿puedese...	¿Puédese...esta insolencia?	¿y puede-se...	¿Podráse permitir tal...
Responde el Papa: sí,...	El Papa dixo:...	El Papa dixo, si,...	El Papa dixo, si,...	El Papa dixo, si...	El Papa dijo, si,...
tome mujer y sufra si es posible	tome mujer y sufra si es posible	tome mujer i sufra si es posible	tome mu-xer i sufra si es posible	tome mujer y sufra si es sufrible	este se... y pase si es posible
tan agradable mal en...	tan insufrible mal en...	tan insufrible carga en...	tan insufrible mal en...	tan insufrible mal en...	tan insufrible mal en...

Tab. 1: Comparación de versiones con base en el ensayo de Fucilla

**Bnm1** (Biblioteca Nacional Matritense 198, fol 72), **Bnm2** (Biblioteca Nacional Matritense 305, fol \*\*), **Baca** (Biblioteca Antequerana de la Caja de Ahorros de Antequera, *Variedad de Sonetos recogidos de diferentes autores por Ignacio de Toledo, año 1627*, Vol. 1 fol. 221), **Bnm3** (Biblioteca Nacional Matritense Ms 19387, fol. 79), **Bm1** (British Museum, Ms. 18, 706, fol. 138v. – letra de los años 1598 a 1612–), **Bm2** (British Museum, Ms. 10, 257 –its Handwriting is of the 18<sup>th</sup> Century).



## Estilometría y la Edad Media castellana

José Manuel Fradejas Rueda (Valladolid)

**RESUMEN:** En este trabajo se muestra que para los análisis estilométricos con la librería *stylo* se pueden utilizar transcripciones semipaleográficas de textos españoles medievales. Se ha puesto en prueba con las obras completas de don Juan Manuel (regularizadas y semipaleográfica), con las obras que conforman el *Scriptorium* alfonsí (semipaleográficas). Este corpus ha arrojado un resultado curioso, pero aún inexplicable, y es que considerando tan solo las 100 palabras más frecuentes, el sistema ha podido establecer que las obras del *Scriptorium alfonsí* se agrupan en dos grandes bloques: 1) obras traducidas del árabe y 2) obras de fuente latina. Alentados por los resultados, y tras estudiar los trabajos de *lexómica*, se ha analizado el *Libro de la caza de las aves* de Pero López de Ayala y el *Libro de la montería*. Los resultados, aunque conocidos, han sido muy interesantes ya que al dividir los textos en pasajes de 800 o 1000 palabras, el sistema ha sido capaz de detectar la existencia de las distintas fuentes de las que han bebido tanto Ayala como el *Libro de la montería* y agruparlas entre sí.

**ZUSAMMENFASSUNG:** Der Beitrag zeigt, dass für stilometrische Untersuchungen mit dem *Stylo*-Paket für R auch semi-paläografische Transkriptionen mittelalterlicher Texte genutzt werden können. Verglichen werden dazu die gesammelten Werke Juan Manuels (regularisiert und semi-paläographisch) mit den Werken aus dem *Scriptorium* von Alfonso X (semi-paläographisch). Mit diesem Korpus wurden interessante aber kaum erklärable Ergebnisse erzielt. Beispielsweise kann die Methode schon bei 100 MFW (most frequent words) die Werke des *Scriptorium* von Alfonso X in zwei Blöcke gruppieren: 1. Werke, die aus dem Arabischen übersetzt sind, 2. Werke, die aus dem Lateinischen stammen. Angespornt durch diese Ergebnisse und im Zusammenhang lexometrischer Studien am Wheaton College, wurde die gleiche Methode auf einen Vergleich des *Libro de la caza de aves* von Pero Lopez de Ayala (in einer modernisierten Fassung) und des *Libro de la montería* (in einer semi-paläographischen Fassung) angewendet. Die bereits bekannten Ergebnisse wurden noch spannender, weil durch eine Aufteilung der Texte in Abschnitte von 800 und 1000 Wörtern mit der stilometrischen Methode die unterschiedlichen Quellen in den Texten nachgewiesen werden können.

**PALABRAS CLAVE:** estilometría; *lexómica*; autoría; fuentes; Edad Media; español medieval

**SCHLAGWÖRTER:** Stilometrie; *Lexomik*; Autorschaft; Quellen; Mittelalter; Altspanisch

<sup>o</sup> Este artículo forma parte de las investigaciones del proyecto *Edición crítica (digital) de las Siete Partidas*, financiado por el MINECO y el FEDER (ref. FFI2016-75014-P AEI-FEDER, EU) que se lleva a cabo en la Universidad de Valladolid. Para más información acerca de este proyecto véase el sitio web <https://7partidas.hypotheses.org/> y el repositorio <https://github.com/7PartidasDigital>.

## Introducción

En un trabajo reciente pude comprobar que la estilometría automatizada, ya sea aplicando el *script* de aprendizaje de programación en R publicado por Jockers<sup>1</sup>, ya sea con el paquete *stylo*, funciona correctamente a la hora de identificar un texto moderno o contemporáneo escrito en español<sup>2</sup>. También se han realizado algunos ensayos para ver si esas mismas técnicas, y en especial el paquete *stylo*<sup>3</sup>, son aplicables a textos de etapas anteriores cuya autoría es altamente disputada, como quiénes son (o no son) los autores del llamado *Quijote de Avellaneda* o el del *Lazarillo de Tormes*<sup>4</sup>. Lo cierto que esos análisis parecen contradecir todo lo dicho hasta el momento sobre la autoría de estas dos obras. Ante el éxito de estos ensayos, y alguno más del que tengo noticias, me planteé la posibilidad de aplicar estas técnicas a textos de un pasado más lejano, tan lejano que a veces el gran problema subyace en la defectuosa transmisión textual que han sufrido. Hemos estado haciendo pruebas, no se las puede considerar de ninguna otra manera, con textos de una serie de autores de los que tenemos un relativo buen conocimiento, en algunos muy profundo, desde el punto de vista de la filología más tradicional, y los resultados son los que exponemos a continuación.

## Juan Manuel

Es un tema clásico de la historia de la literatura medieval española afirmar que las obras de don Juan Manuel (1282–1348) se pueden dividir en dos grandes bloques: uno de influencia alfonsí y otro de un estilo más personal. En el primer bloque se incluyen la *Crónica abreviada*, el *Libro de la caza* y el *Libro del*

---

<sup>1</sup> Matthew L. Jockers, *Text Analysis with R for Students of Literature* (Cham: Springer, 2014), doi:10.1007/978-3-319-03164-4.

<sup>2</sup> José Manuel Fradejas Rueda, “El análisis estilométrico aplicado a la literatura española: las novelas policíacas e históricas”, *Caracteres: estudios culturales y críticos de la esfera digital* 5, n.º 2 (2016): 196–245.

<sup>3</sup> Maciej Eder, Jan Rybicki, y Mike Kestemont, “Stylo: a Package for Stylometric Analyses”, 2013.

<sup>4</sup> Nanette Rissler-Pipka, “Avellaneda y los problemas de la identificación del autor: propuestas para una investigación con nuevas herramientas digitales”, en *El otro Don Quijote: la continuación de Fernández de Avellaneda y sus efectos*, ed. Hanno Ehrlicher (Ausburgo: Institut für Spanien, Portugal- und Lateinamerikastudien, 2016), 27–51; Nanette Rissler-Pipka, “Der falsche Quijote? Autorschafts attribution für spanische Prosa der frühen Neuzeit”, en *DHd 2016 Modellierung, Vernetzung, Visualisierung*, 2016, 212–17; Nanette Rissler-Pipka, “Digital Humanities und die romanische Literaturwissenschaft. Der Autorschaftsstreit um den Lazarillo de Tormes”, *Romanische Forschungen* 128, n.º 3 (2016): 316–42.



*caballero e del escudero*; y en el segundo, el resto de sus obras. La consideración alfonsina del *Libro del caballero e del escudero* se debe a que se inspiró en el título 21 de la segunda partida; la del *Libro de la caza* se debía a que siempre se ha dicho que lo escribió siguiendo un desaparecido, e ignoto, libro de la caza escrito por el rey Alfonso X<sup>5</sup>, sin embargo, se ha podido demostrar que Juan Manuel sigue el modelo del *De arte venandi cum avibus* de Federico II<sup>6</sup>. Resta una tercera obra de corte alfonsí: la *Crónica abreviada*, que es un resumen de la *Estoria de España* de Alfonso X.

El análisis del texto con el paquete *stylo* se ha realizado sobre la edición normalizada de Carlos Alvar<sup>7</sup> y nos muestra, como se observa en la figura 1, que la *Crónica abreviada* es una obra que nada tiene que ver con el estilo de don Juan Manuel, que es un elemento haplofílico dentro del dendrograma, o por usar el término estadístico, un dato atípico.

---

<sup>5</sup> Esta afirmación se basa en los que dice don Juan Manuel en el prólogo al *Libro de la caza*: “Et el dicho rey don Alfonso deseando el saber, como dicho es, et pagándose de todas las cosas nobles et apuestas et sabrosas et aprehochosas, entendiendo que en la caça ha estas quatro cosas muy complidamente a los que quieren usar della como deven, et non dexar por ella otros fechos mayores, ca los que en otra manera caçassen, aunque guardassen el sabor et la apostura de la caça, non guardarían la nobleza nin el aprovechamiento, por ende mandó fazer muchos libros buenos en que puso muy complidamente toda la arte de la caça, tan bien del caçar, como del benar, como del pescar. Et puso muy complidamente la teórica et la práctica como conviene a esta arte; et tan complidamente lo fizo, que bien cuidan que non podría otro emendar nin eniader ninguna cosa más de lo que él fizo, nin aun fazer tanto nin tan bien como él”, Juan Manuel, *Obras completas*, ed. Carlos Alvar y Sara Finci (Madrid: Fundación José Antonio de Castro, 2007).

<sup>6</sup> José Manuel Fradejas Rueda, “La influencia del *De arte venandi cum avibus* de Federico II en el *Libro de la caza* de Juan Manuel”, en *Los libros de caza* (Tordesillas y Valladolid: Seminario de Filología Medieval, Universidad de Valladolid, 2005), 41–54; José Manuel Fradejas Rueda, “Juan Manuel y Federico II de Hohenstaufen”, en “*Siempre soy quien ser solía*”: estudios de literatura española medieval en homenaje a Carmen Parrilla, vol. 1 (La Coruña: Universidade da Coruña, Servicio de Publicaciones, 2009), 137–48.

<sup>7</sup> El mismo editor (Juan Manuel, *Obras completas*) nos facilitó una copia en formato Word de los textos tal y como los entregó a la editorial. Ese fichero se convirtió a texto plano y cada una de las obras se guardó en un fichero independiente que se han identificado como: ABREVIADA.txt (99767 palabras-token; 6943 palabras-tipo), ARMAS.txt (7576 palabras-token; 1395 palabras-tipo), ASUNCION.txt (2862 palabras-token y 638 palabras-tipo), CAB\_ESC.txt (26508 palabras-token; 2695 palabras-tipo), CAZA.txt (28647 palabras-token; 2916 palabras-tipo), ENFENIDO.txt (15748 palabras-token; 1856 palabras-tipo), ESTADOS.txt (103243 palabras-token; 5922 palabras-tipo) y LUCANOR.txt (77764 palabras-token; 5717 palabras-tipo) lo que ha supuesto un corpus de 362115 palabras-token y 15821 palabras-tipo. Una vez obtenidos estos materiales, se han procesado con *stylo*.

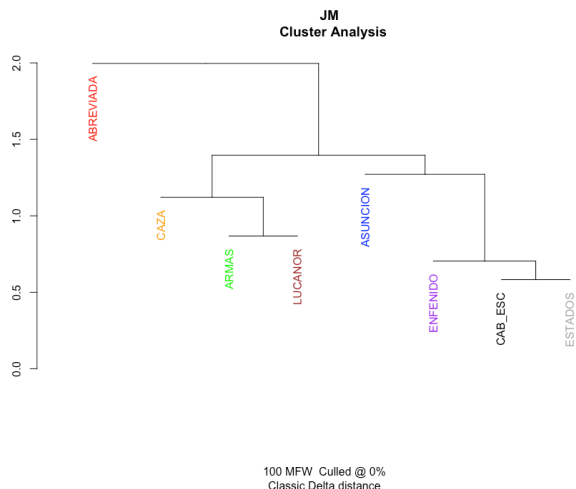


Fig. 1: Obras de don Juan Manuel, según la edición normalizada de Alvar (100 MFV)

Al aumentar el número de palabras que se tienen en cuenta a 1000, la geometría del dendrograma se mantiene, y la *Crónica abreviada* sigue siendo un dato atípico (fig. 2).

El objetivo es comprobar si el hecho de tratarse de versiones normalizadas de un texto medieval y no fieles a un manuscrito, ha podido influir en este resultado puesto que en la edición de Alvar se han regularizado usos gráficos como la alternancia de u/v y i/j/y, que pueden tener valores vocálicos o consonánticos, y se ha acentuado según las reglas ortográficas actuales. Para esta parte del análisis se han manejado las transcripciones semipaleográficas de Ayerbe-Chaux<sup>8</sup> preparadas para el Dictionary of Old Spanish Language (DOSL n adelante) del Hispanic Seminary of Medieval Studies (HSMS en adelante). Ahora bien, se ha tomado la decisión de convertir la nota tironiana, que representa la conjunción copulativa y, que en estas transcripciones se transcriben con &, se han sustituido por e en la *Crónica Abreviada* ya que es la forma mayoritaria cuando no se usa la nota tironiana, y por et en el resto de las obras ya que es la de mayor uso. Asimismo, se han efectuado otros dos cambios: la conversión de la secuencia <n-> en <ñ> y de <c'> en <ç>, dado que estas dos letras no existen en el juego de caracteres ASCII con el que se

<sup>8</sup> Reinaldo Ayerbe-Chaux, *Textos y concordancias de la obra completa de Juan Manuel* (Madison: Hispanic Seminary of Medieval Studies, 1986).

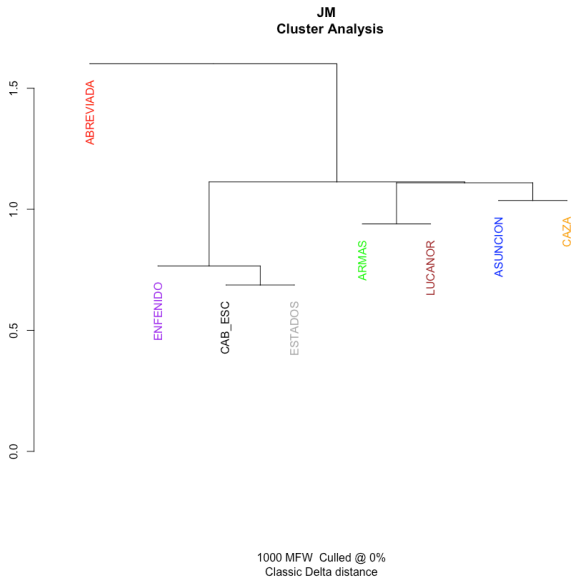


Fig. 2: Obras de don Juan Manuel, según la edición normalizada de Alvar (1000 MFW)

transcribieron originalmente los textos del HSMS<sup>9</sup>. A pesar de estos cambios, no parece que la fidelidad a un testimonio medieval sea de gran valor, como lo demuestran las figuras 3 y 4 en donde se puede ver que la Crónica abreviada sigue siendo un rasgo atípico dentro de la producción textual de don Juan Manuel, tanto si solo se consideran las 100 palabras más frecuentes (fig. 3) como si se tienen en cuenta las 1000 más frecuentes (fig. 4).

## Alfonso X

Bajo el nombre de Alfonso X (1221–1284), la historia de la literatura española agrupa un rico corpus textual que comprende obras legales, históricas, científicas y de recreación. Uno de los temas más debatidos es el referente a la intervención del rey en la redacción de las obras y el equipo con el que contaba en el llamado *Scriptorium alfonsí*. En este punto, se examinarán las obras alfonsíes desde un punto de vista estilométrico para observar cómo se agrupaban los distintos textos que constituyen este rico corpus textual.

<sup>9</sup> Kenneth Buelow y David Mackenzie, *A manual of manuscript transcription for the Dictionary of the Old Spanish language* (Madison Wis.: Hispanic Seminary of Medieval Studies, 1977), 11–2.

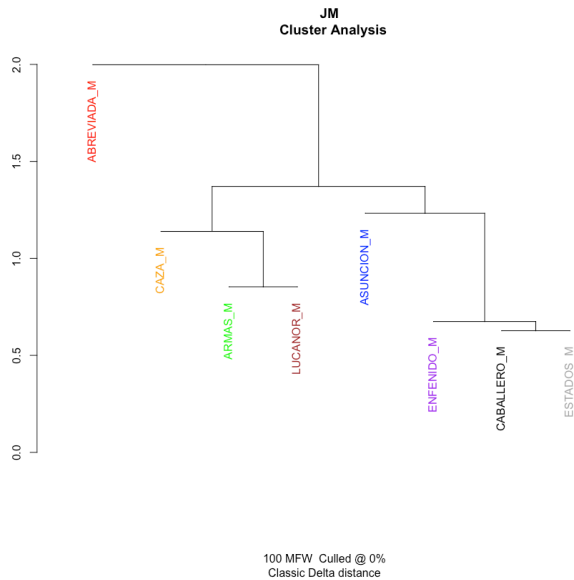


Fig. 3: Obras de don Juan Manuel, según la edición semipaleográfica de Ayerbe-Chaux (100 MFW)

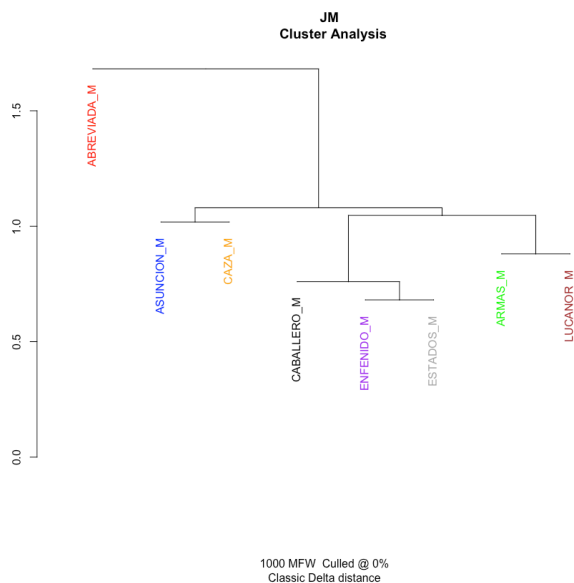


Fig. 4: Obras de don Juan Manuel, según la edición semipaleográfica de Ayerbe-Chaux (1000 MFW)

Para esto se ha contado, de nuevo, con todos los textos del *Scriptorium alfonsí* transcritos para el *Dictionary of Old Spanish Language* del Hispanic Seminary of Medieval Studies, transcripción llevada a cabo por Kasten *et al.*<sup>10</sup> y algunos textos más recopilados por O'Neill<sup>11</sup> y reunidos por Gago-Jover<sup>12</sup>. Como en el caso de las obras de don Juan Manuel, se ha procedido a sustituir la nota tironiana & bien por *e* bien por *et*, de acuerdo con el uso mayoritario en cada testimonio de una u otra forma (véase Tabla 1), de manera que no se pierda información.

Un sencillo análisis de grupos con el paquete *stylo*, en el que solo se han tenido en cuenta las 100 palabras más frecuentes, ofrece un interesante y curioso dendrograma (fig. 5).

Las obras del rey Sabio aparecen distribuidas en dos grandes bloques: a la derecha, traducciones del árabe y a la izquierda, obras históricas y jurídicas con fuente latina. En casi todos los casos la longitud de los subclados es muy corta, lo que indica una gran homogeneidad entre los textos, especialmente en la serie *Judizios, Ymagenes, Ajedez, Moamin, Lapidario* y *Picatrix*.

---

<sup>10</sup> Lloyd August Kasten *et al.*, *The electronic texts and concordances of the prose works of Alfonso X, El Sabio* (Madison: Hispanic Seminary of Medieval Studies, 1997).

<sup>11</sup> John O'Neill y Hispanic Seminary of Medieval Studies, *Electronic texts and concordances of the Madison Corpus of early Spanish manuscripts and printings* (New York: Hispanic Seminary of Medieval Studies, 1999).

<sup>12</sup> Las obras que se han tenido en cuenta son: *Canones de Albateni, Libros de ajedrez, dados y tablas, Libros del saber de astronomía, Libro de las cruces, Estoria de España I, Estoria de España II, General estoria I, General estoria II, General estoria IV, General estoria V, General estoria VI, General estoria VI, Judizos de las estrellas, Lapidario de Alfonso X, Libro de las leyes, Moamyn-Libro de las animalias, Picatrix de Alfonso X, Libro del cuadrante señero, Libro de las formas y de las imagenes, Tablas de Zarquiel, Setenario*. El *Libro de las animalias que cazan*, erróneamente considerado como el *Libro de Moamín*, es transcripción de Cárdenas – *The text and concordance of Biblioteca Nacional Manuscript RES. 270–217 Libro que es fecho de las animalias que caçan, the Book of Moamin* (Madison [Wis.]: Hispanic Seminary of Medieval Studies, 1987) –, mientras que la del *Setenario* lo es de Bonch-Bruevich (*Electronic text and concordances of the Setenario by Alfonso el Sabio: Biblioteca Capitular de Toledo, MS 43–20* (Madison: Hispanic Seminary of Medieval Studies, 1999)). Todo este corpus textual supone un total de 3.234.321 palabras y se halla reunido en la Biblioteca Digital de Textos del Español Antiguo (Francisco Gago Jover, “Obra en prosa de Alfonso X el Sabio: biblioteca digital de textos del español antiguo”, 2011, <http://www.hispanicseminary.org/t&c/ac/index-es.htm>). El recuento es el ofrecido por el HSMS para cada uno de los textos (véase Tabla 2).

	<i>e</i>	<i>et</i>	<i>&amp;</i>
AJEDREZ	474	157	2014
ALBATENI	7	1403	1714
ASTRONOMIA	130	4942	15826
CRUZES	30	1550	5790
CUADRANTE	20	417	341
EE_1	6282	731	13317
EE_2	429	10181	17457
GE_1	626	4477	38388
GE_2	166	6382	13922
GE_4	270	563	37880
GE_5	1	3691	9457
GE_5r	6	3982	11329
GE_6	2	313	1189
JUDIZIOS	2822	3	19984
LAPIDARIO	8	2621	5420
LEY_LBL	2330	43	4325
MOAMIN	10499	10	96
PICATRIX	17	582	2584
SETENARIO	8	1420	3022
YMAGENES	1	58	1164
ZARQUIEL	0	0	12

Tab. 1: Formas y ocurrencia de la conjunción copulativa en el corpus alfonsí

	palabras-token	palabras-tipo
AJEDREZ	53638	1899
ALBATENI	44386	2115
ASTRONOMIA	275364	6238
CRUZES	130770	2815
CUADRANTE	10210	633
EE_1	258521	14620
EE_2	352894	16385
GE_1	535532	20663
GE_2	243965	1447
GE_4	456311	20520
GE_5	146468	10343
GE_5r	193144	12549
GE_6	17338	2839
JUDIZIOS	225456	7413
LAPIDARIO	103121	5655
LEY_LBL	153664	6833
MOAMIN	107190	6345
PICATRIX	34584	4421
SETENARIO	78729	638
YMAGENES	17203	2104
ZARQUIEL	849	127

Tab. 2: Número de palabras-token y palabras-tipo de cada uno de los textos tenidos en cuenta

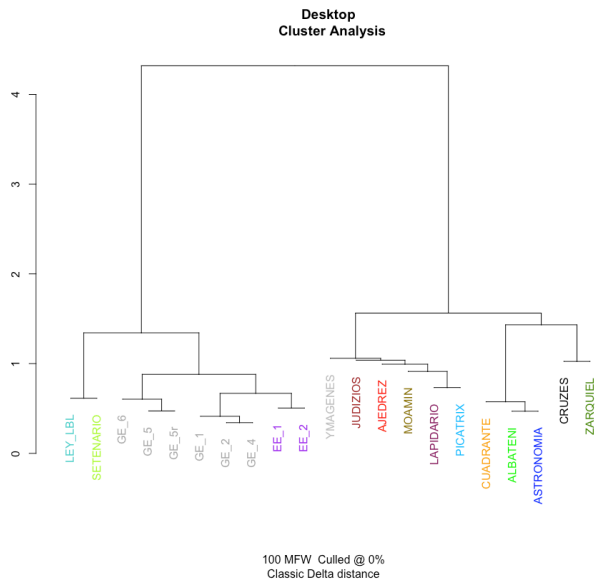


Fig. 5: Obras del Scriptorium alfonsí (100 MFW)

El que el *Libro de los animales que cazan* (o *Moamín*)<sup>13</sup>, obra traducida del árabe cuando aún Alfonso no había accedido al trono, se asiente cómodamente dentro del compacto y homogéneo grupo de obras de origen árabe traducidas a lo largo de 30 años puede indicarnos que la traducción es alfonsí realmente,

<sup>13</sup> Designar a esta obra como *Moamín* se debe a un error que se remonta a finales de los años 1940, cuando Håkan Tjerneld (“Una Fuente Desconocida Del Libro de La Montaña Del Rey Alfonso El Sabio”, *Studia Neophilologica* 22, n.º 2 (21 de enero de 1949): 171–93, doi:10.1080/00393274908587044) estableció la relación entre la versión castellana del *Kitab al-yawarīh*, hasta entonces conocido como *Libro de cetrería* (El Escorial, ms. V.II.19), y un texto editado por el mismo Tjerneld (*Moamīn et Ghatrif: traités de fauconnerie et des chien de chasse* (Stockholm: Fritze, 1945)) debido a la pluma de un cetrero árabe conocido simplemente como *Moamīn* y que François Viré (“Sur l’identité de Moamyn le fauconnier”, *Comptes Rendus de Séances* 111, n.º 2 (1967): 172–5) trató de identificar con Hunayn ibn Ishaq (808–873). Al acceder a principios de los años 1980 a un nuevo testimonio de la versión castellana (olím Phillipps 11719, BNE RES/270) se pudo establecer que el autor era Muhammad ibn Abd Allah ibn Umar al-Bayzar (Muḥammad ibn Abdallah Al-Bāzyār, *Libro de los animales que cazan* = *Kitāb al-Ŷawārīh*, ed. José Manuel Fradejas Rueda (Madrid: Editorial Casariego, 1987)), autor del que años más tarde se publicarían partes del texto árabe bajo el título de *Kitab al-Mutawakkil* (véase Muḥammad ibn Abdallah Al-Bāzyār, *Das Falken- und Hundebuch des Kalifen al-Mutawakkil: ein arabischer Traktat aus dem 9. Jahrhundert*, ed. Anna Akasoy y Stefan Georges (Berlin: Akademie Verlag, 2005)).



aunque se realizara en un momento previo a su coronación como rey, por lo que no se sostiene la apreciación de Fernández-Ordóñez de que el *Libro de los animales que cazan* puede pertenecer al entorno de Fernando III, aunque este no aparece “como responsable o instigador, no digamos autor”<sup>14</sup>.

Incluso si aumentamos a 3000 palabras, entrando en el resbaladizo ámbito de las palabras semánticas, se observa en la figura 6 que seguimos teniendo dos grandes grupos: uno constituido por las obras traducidas del árabe y el otro conformado por las obras históricas y legales con un fortísimo componente de traducción, o fuentes, desde el latín. Pero la situación de estos dos grupos es inversa a la mostrada anteriormente: el grupo árabe está ahora a la izquierda, mientras que el latino se sitúa a la derecha, lo que es poco o nada significativo, puesto que un dendrograma solo representa los grupos, ya que para una estructura de grupos cualquiera existen  $2^{n-1}$  árboles posibles<sup>15</sup>.

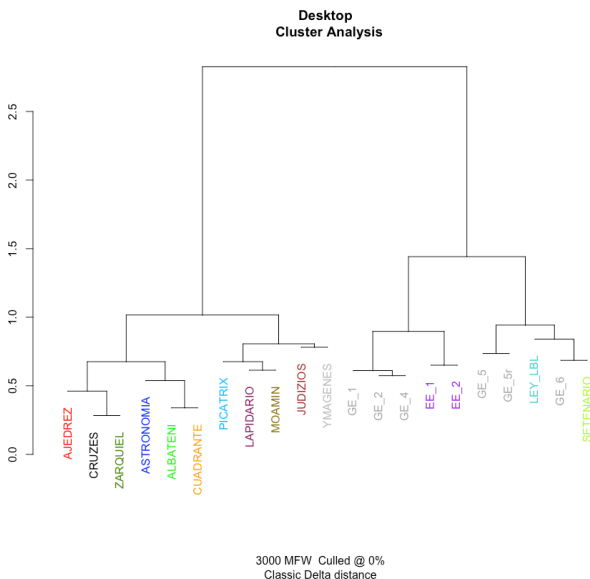


Fig. 6: Obras del *Scriptorium alfonsí* (3000 MFW)

<sup>14</sup> Inés Fernández-Ordóñez, “Alfonso X el Sabio en la historia de español”, en *Historia de la lengua española*, ed. Rafael Cano Aguilar (Barcelona: Ariel, 2004), 386–7.

<sup>15</sup> Hermann Moisl, *Cluster Analysis for Corpus Linguistics* (Berlin: De Gruyter Mouton, 2015), 214.

Otro de los asuntos candentes dentro de los estudios alfonsíes es el de la redacción y finalización de las *Siete Partidas*. En el esquema anterior (fig. 5 y 6), se ha tenido en cuenta el llamado *Libro de las leyes* (British Library, ms. Additional 20787), que corresponde al testimonio más antiguo de la primera partida<sup>16</sup>, pero no existe ningún testimonio de época alfonsí de todas las *Partidas*. Es más, ninguno de ellos es anterior al reinado de Sancho IV y, por lo que parece, la gran mayoría de los conservados son copias realizadas después de 1348, cuando Alfonso XI declara en el Ordenamiento de Alcalá 1348 las *Siete Partidas* como derecho supletorio (título 28, ley 1) y ordena hacer dos ejemplares sellados con su sello de oro y plomo<sup>17</sup>. A finales del siglo XV, en 1491, se fija un texto que se convertirá en el texto normativo durante los 65 años siguientes: la edición de Alonso Díaz de Montalvo. De esta tenemos una versión electrónica cuidada con los mismos criterios de transcripción que el resto del corpus alfonsí<sup>18</sup>. Al incorporarla, se ve que se sitúa junto al *Libro de las leyes*, aunque el *Setenario* se desplaza y se une a GE6, algo que ya habíamos visto en la figura 5. Por otra parte, las dos versiones de las *Siete Partidas* se unen en un grupo absolutamente compacto con una distancia cero, como se puede apreciar en la figura 7, a pesar de que la distancia cronológica entre ambas versiones es de 200 años.

Este tipo de análisis con textos tanto contemporáneos como medievales, como los que acabamos de mostrar, permiten establecer con alta confianza la atribución a un autor de una obra con independencia de si los textos están escritos en lengua actual o en la de tiempos pasados, o si están escritos en forma modernizada o con grafías antiguas. Pero... ¿puede el análisis de grupos, la estilometría, decirnos algo más acerca de una obra? Acabamos de ver que hay algo curioso en los textos alfonsíes: un sistema automatizado separa las obras que son traducción de un texto árabe de las que tienen fuentes latinas, aunque no se puede aventurar que sean traducción directa de un texto latino, ya que cada uno de ellos tiene múltiples fuentes. Uno de los datos relativos de la influencia del árabe en la escritura castellana es la abundantísima coor-

---

<sup>16</sup> Juan Arias Bonet, *Primera partida según el manuscrito Add. 20.787 del British Museum* (Valladolid: Universidad de Valladolid Secretariado de Publicaciones, 1975).

<sup>17</sup> Alfonso Otero Varela, "Las Partidas y el Ordenamiento de Alcalá en el cambio del ordenamiento medieval", *Anuario de historia del derecho español* 1, n.º 63 (1993): 451-548.

<sup>18</sup> Francisco Gago-Jover, "'Siete partidas': textos legales españoles. Biblioteca digital de textos del español antiguo", 2013, <http://www.hispanicseminary.org/t&c/lex/index-es.htm>. La edición de 1491 incorpora unas adiciones de mano de Montalvo, pero se han eliminado para este estudio, ya que podrían introducir ruido.

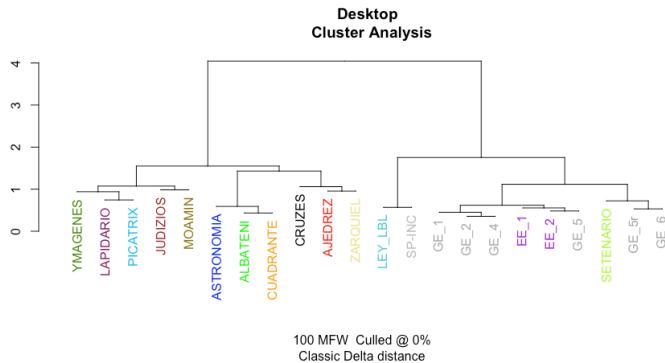


Fig. 7: El corpus alfonsí con las *Siete Partidas* de 1491 (SP-INC) (100 MFW)

dinación copulativa<sup>19</sup>. ¿Qué pasaría si se retirara este marcador discursivo de altísima frecuencia en el corpus alfonsí (209009 casos de *et* y 63886 casos de *e*)? El resultado es que al borrarse de los textos la conjunción copulativa, no varía un ápice la geometría básica del dendrograma, esto es, sigue habiendo un grupo de traducciones del árabe y otro, de fuentes latinas como se aprecia en la figura 8.

## Lexómica

En el Wheaton College han desarrollado un método estilométrico al que designan *lexómica*. Este nuevo método adopta técnicas de la bioinformática, de la estilometría y del análisis textual tradicional, es decir, de los métodos filológicos de siempre. Lo que hacen es utilizar técnicas de la estadística informatizada para identificar patrones que después interpretan usando los métodos filológicos tradicionales. Al inicio de sus investigaciones, pudieron comprobar que “the computational methods told us where in a text to look”<sup>20</sup>.

Según exponen, los métodos lexómicos “differ slightly from pioneering stylometrics analyses in two major ways [...] First [...] we include every word in our analyses. Second [...] we divide our texts into segments and analyze the

<sup>19</sup> Rafael Lapesa, *Historia de la lengua española* (Madrid: Gredos, 1981), 150 y Lola Pons Rodríguez, *La lengua de ayer. Manual práctico de historia del español* (Madrid: Arco Libros, 2010), 358, aunque esta autora no le da validez a este rasgo.

<sup>20</sup> Phoebe Boyd et al., “Lexomic Analysis of Anglosaxon Prose: Establishing Controls with the Old English Penitential and the Old English Translation of Orosius”, *Selim: Journal of the Spanish Society for Medieval English Language and Literature* = *Revista de la Sociedad Española de Lengua y Literatura Inglesa Medieval*, n.º 19 (2012): 7–58, aquí 9.

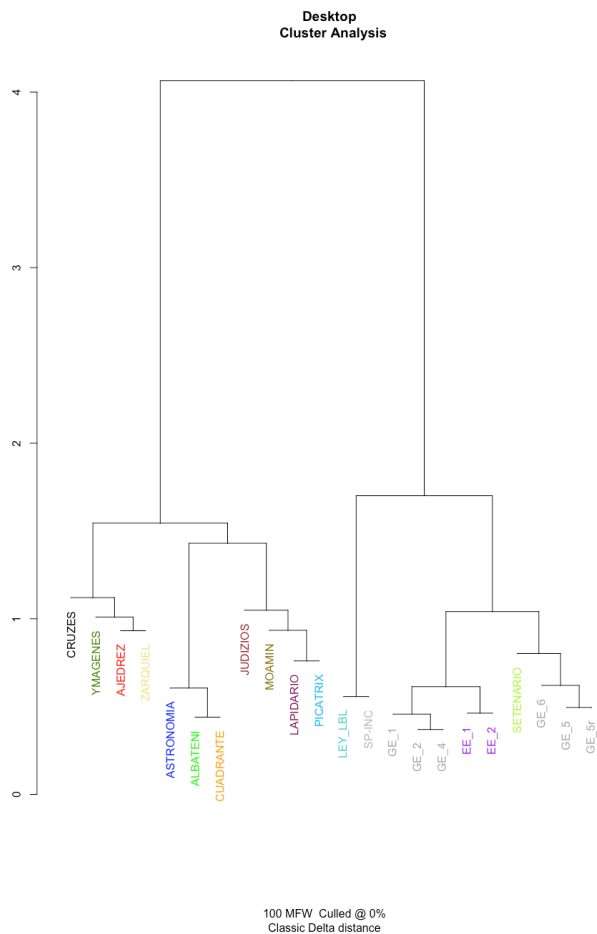


Fig. 8: El corpus alfonsí con las *Siete Partidas* de 1491 (SP-INC) y eliminadas las conjunciones copulativas *e* y *et* (100 MFW)

relationship of these to each other”<sup>21</sup>. Y, aunque reconocen que estos análisis tienen cierta propensión al establecimiento de la autoría, sus objetivos no están “focused primarily on author identification but instead on text’s sources or affinities”<sup>22</sup>. En nota a pie de página reconocen que las fuentes y las afinidades pueden tener que ver con los problemas de autoría, y de hecho lo han utilizado para mostrar que *Guthlac B*, uno de los poemas anglosajones que narran las leyendas de san Guthlac (673–714), lo escribió Cynewulf<sup>23</sup>.

También reconocen que un análisis lexómico puede no decir nada nuevo sobre un texto. Sin embargo, admiten que gracias a la correlación de la geometría de un dendrograma con el conocimiento previo que podemos tener sobre un texto dado, “we can have reasonable confidence in the relationships suggested by our methods when these are applied to texts whose interrelationship is unknown or controversial”<sup>24</sup>, es decir, examinar con los métodos lexómicos textos que sabemos cómo *funcionan* gracias a los métodos filológicos puede servirnos como *control* para aquellos que no conocemos aún<sup>25</sup>.

Downey et al.<sup>26</sup> insisten repetidamente en que la lexómica muestra una relativa homogeneidad en el vocabulario de los autores examinados. Sin embargo, esto es una falsa apreciación ya que no tienen claro qué se ha de entender por el léxico de un texto, o bien no tienen nada claro qué palabras utilizan los programas para realizar el análisis de grupos, aunque, cuando tratan de explicar que el fragmento 17 del *Planctu naturae* de Alain de Lille (a. 1128–1202) es un dato atípico, una rama *haplofilica* o *simplifolius*, reparan en que las palabras más abundantes son *alii*, *alienus* e *isti*, poco informativas semánticamente<sup>27</sup>. Puesto que la lexómica, así como la estilometría, se basan en el análisis de grupos, con independencia de la fórmula matemática empleada (Euclidiana, Delta, Manhattan, Camberra, Eder...), sin hacer uso de las palabras semánticas sino de las gramaticales, que suelen ser las más frecuentes, la homogeneidad de estilo a la que se refieren parece que es independiente del contenido temático del texto.

<sup>21</sup> Boyd et al., “Lexomic Analysis”, 9.

<sup>22</sup> Boyd et al., “Lexomic Analysis”.

<sup>23</sup> Michael D C Drout et al., “Of Dendrogrammatology: Lexomic Methods for Analyzing Relationships among Old English Poems”, *Journal of English and Germanic Philology* 110, n.º 3 (2011): 301–36.

<sup>24</sup> Drout et al., “Of Dendrogrammatology”, 38, y Sarah Downey et al., “Lexomic Analysis of Medieval Latin Texts”, *The Journal of Medieval Latin*, n.º 24 (2014): 225–75, aquí 232.

<sup>25</sup> Boyd et al., “Lexomic Analysis”, 14, y Downey et al., “Lexomic Analysis”, 232.

<sup>26</sup> Downey et al., “Lexomic Analysis”.

<sup>27</sup> Boyd et al., “Lexomic Analysis”, 240.

También afirman que el “lexomic analysis is based on critical editions”<sup>28</sup>, debido a que “previous lexomic research suggests cluster analyses based on critical editions are more reliable than those based on diplomatic texts”<sup>29</sup>. Pero, como acabamos de ver, en el caso de don Juan Manuel, los dendrogramas de las obras de este autor, ya se hayan basado en el análisis de la edición normalizada de Alvar<sup>30</sup>, ya sea en la edición semipaleográfica de Ayerbe-Chaux<sup>31</sup> no varían. Lo mismo cabe decir de las transcripciones de las obras de *Scriptorium alfonsí*<sup>32</sup>, que no parecen afectar en nada los resultados, por lo que no se puede afirmar que el “lexomic analysis goes somewhat against the grain of manuscript-approaches”<sup>33</sup>. Es más, ellos mismos reconocen que “the differences between diplomatic and critical editions are relatively invisible to lexomic methods [...] because the most common words in Anglo-Saxon are those least likely to be emended”<sup>34</sup>.

Partiendo de las ideas y sugerencias del grupo formado por Downey, Boyd y Drout et al.<sup>35</sup> y conociendo perfectamente el comportamiento del *Libro de la caza de las aves* de Pero López de Ayala (1332–1407), y del *Libro de la montería* (siglo XIV, c. 1350), hemos realizado un análisis lexómico de ambos textos, pero con el paquete *stylo*, puesto que los procesos que ejecuta el software diseñado por este grupo, Lexos<sup>36</sup>, también los puede realizar *stylo* sin mayores problemas, ya que las fases del trabajo son las mismas. La única diferencia de importancia es que Lexos se puede encargar de fragmentar los textos, bien por número de palabras, bien por medio de hitos (*milestones*)<sup>37</sup>, mientras que en

<sup>28</sup> Boyd et al., “Lexomic Analysis”, 18.

<sup>29</sup> Michael D. C. Drout et al., *Beowulf Unlocked: New Evidence from Lexomic Analysis* (Palgrave Macmillan, 2016), 19.

<sup>30</sup> Juan Manuel, *Obras completas*.

<sup>31</sup> Ayerbe-Chaux, *Textos y concordancias de la obra completa de Juan Manuel*.

<sup>32</sup> Gago Jover, “Obra en prosa de Alfonso X el Sabio.”

<sup>33</sup> Boyd et al., “Lexomic Analysis”, 19.

<sup>34</sup> Boyd et al., “Lexomic Analysis”, 24–5.

<sup>35</sup> Sarah Downey, Michael D C Drout, y Michael J Kahn, “Books Tell Us’: Lexomic and Traditional Evidence for the Sources of Guthlac A California University of Pennsylvania”, *Modern Philology* 110, n.º 2 (noviembre de 2012): 69–86, doi:10.1086/668252; Downey et al., “Lexomic Analysis”; Boyd et al., “Lexomic Analysis”; Drout et al., “Of Dendrogrammatology”; Michael D. C. Drout, *Tradition and Influence in Anglo-Saxon Literature: An Evolutionary, Cognitivist Approach* (New York: Palgrave Macmillan, 2013); Drout et al., *Beowulf Unlocked*.

<sup>36</sup> Se puede trabajar en línea (<http://lexos.wheatoncollege.edu/upload>) o se puede instalar el software en máquinas Windows 7 o superior), Mac OSX o Linux (<https://wheatoncollege.edu/academics/special-projects-initiatives/lexomics/lexos-installers>).

<sup>37</sup> Estos hitos pueden ser capítulos, folios, etc.

*stylo* es una tarea previa, aunque poco complicada con unos pequeños *scripts* en R o cualquier otro lenguaje de programación. Las demás tareas son idénticas y lo único que pueden variar son las fórmulas para calcular las distancias. Tras cargar el texto en el programa, este se encarga de convertirlo todo a minúsculas (para evitar que, por ejemplo, la palabra *Rey* y *rey* cuenten como dos palabras diferentes, cuando, en realidad, son una), y eliminar la numeración arábiga y los signos de puntuación<sup>38</sup>. A todo este proceso los lexicónomas han dado el nombre de *scrubbing*<sup>39</sup>. Posteriormente, el programa divide el texto en palabras (*tokens*) y calcula las frecuencias absolutas y relativas para cada uno de los textos (o fragmentos de texto) y, por último, aplica la fórmula elegida para el cálculo de las distancias y dibuja el dendrograma<sup>40</sup>. *Lexos* puede producir otros elementos de fantasía como nubes de palabras o diagramas de burbujas, elementos innecesarios para extraer conclusiones y tablas estadísticas con informaciones variadas. Por su parte, *stylo* también ofrece tablas estadísticas.

### El Libro de la caza de las aves

El *Libro de la caza de las aves*, escrito por Pero López de Ayala<sup>41</sup> mientras estuvo preso en Óvidos tras la derrota castellana de Aljubarrota (14/8/1385) es un texto que tiene dos fuentes principales: una versión al castellano del *Livro de cetraria* de Pero Menino<sup>42</sup> y lo escrito por el propio Ayala. Está bastante claro cuáles son los capítulos que son traducción de la obra de Pero Menino, los capítulos 11–37 y 39, mientras que se consideran originales de Ayala los capítulos 1–10, 36 y 40–47. La pregunta que nos hacemos es la siguiente: ¿puede el análisis de grupos detectar qué capítulos son de uno y otro autor?

---

<sup>38</sup> También se pueden eliminar espacios en blanco (*whitespace*) e incluso pueden limpiar etiquetas XML, HTML, etc., por lo que no habría problemas para procesar textos codificados en TEI. E incluso se pueden introducir una serie de palabras, bien manualmente bien por medio de un fichero, que no se quiera procesar.

<sup>39</sup> Michael D. C. Drout, *Tradition and Influence*, 52 y Drout et al., *Beowulf Unlocked*, 17.

<sup>40</sup> Durante todo el proceso, el sitio web permite descargar los ficheros intermedios que se van generando: el fichero *limpiado*; una vez segmentado; la tabla con todos los datos tras *tokenizar* (palabras y frecuencias –filas– en cada fragmento de texto –columnas–) que se puede descargar en formato CSV o TSV; y se pueden generar otras estadísticas (número de palabras-tipo; hápax).

<sup>41</sup> Pero López de Ayala, *Libro de la caza de las aves*, ed. José Fradejas Lebrero y José Manuel Fradejas Rueda (Barcelona: Castalia, 2016).

<sup>42</sup> Pero Menino, *Livro de falcoaria*, ed. Manuel Rodrigues Lapa (Coimbra: Imprensa da Universidade, 1931).

Para llevar a cabo este análisis, hemos partido de la edición normativizada del *Libro de la caza de las aves* que se ofrece en el Archivo Iberoamericano de Cetrería<sup>43</sup>. Al tratarse de una página web, se ha de limpiar de los códigos del HTML y de todo el material propio de una página web que no tiene interés alguno para nuestro análisis. Una vez limpio, se ha dividido por capítulos. Esta primera división nos ha proporcionado unas primeras estadísticas básicas, y curiosas (tabla 3)<sup>44</sup>. Así el número de palabras-token va desde el máximo de 6967 (cap. 9) al mínimo de 110 (cap. 45), con una media de 806.

Tab. 3: Estadísticas básicas del *Libro de la caza de las aves*

Capítulo	Tokens	Tipos	Hápx	TTR	Tokens/Obra	Tipos/Obra	Hápx/Obra
AIC-01	1750	602	399	34.4	4.52231439129648	14.2148760330579	18.936877076412
AIC-02	1125	399	253	35.4666666666667	2.9072021086906	9.42148760330578	12.0075937351685
AIC-03	2093	579	346	27.663640707119	5.40868801199059	13.6717827626919	16.421452301851
AIC-04	706	250	158	35.4107648725212	1.82443083443161	5.90318772136954	7.49881347887992
AIC-05	854	347	238	40.632318501171	2.20688942295268	8.19362455726092	11.2956810631229
AIC-06	517	220	144	42.5531914893617	1.33602088017159	5.19480519480519	6.8343616516374
AIC-07	761	313	199	41.1300919842313	1.96656071530093	7.39079102715466	9.44470811580446
AIC-08	398	198	143	49.748743718593	1.02850350156343	4.67532467532468	6.78690080683436
AIC-09	6967	1397	801	20.0516721687958	18.00399796366643	32.987012987013	38.016136687233
AIC-10	446	202	137	45.2914798206278	1.15254412486756	4.76977567886659	6.50213573801614
AIC-11	384	190	134	49.4791666666667	0.992324986433057	4.48642266824085	6.35975320360702
AIC-12	1676	506	294	30.1903907875895	4.33108509703595	11.9480519480519	13.953488372093
AIC-13	533	257	181	48.2176360225141	1.3773677546063	6.06847697756789	8.59041290934979
AIC-14	424	188	119	44.3396226415094	1.09569217251983	4.43919716646989	5.64784053156146
AIC-15	537	243	168	45.2513966480447	1.38770447321498	5.73789846517119	7.9734219269103
AIC-16	531	219	144	41.2429378531073	1.37219939530196	5.17119244391972	6.8343616516374
AIC-17	478	213	141	44.5606694560669	1.23523787373698	5.02951593860685	6.69197911722829
AIC-18	244	117	79	47.9508196721311	0.630539835129338	2.76269185360094	3.74940673943996
AIC-19	703	280	184	39.8293029871977	1.8166782954751	6.61157024793388	8.7327954437589
AIC-20	360	175	115	48.6111111111111	0.930304674780991	4.13223140495868	5.45799715234931
AIC-21	581	248	159	42.6850258175559	1.50140837791043	5.85596221959858	7.54627432368296

<sup>43</sup> <http://www.aic.uva.es/clasicos/ayala/ayala-texto.html>.

<sup>44</sup> El texto del *Libro de la caza de las aves* según esta versión consta de 38697 palabras-token y 4235 palabras-tipo, con lo que obtenemos una TTR de 10.944. Hay 2107 casos de hápx, lo que supone el 5.446 % del texto, y tan solo cinco de ellas –*e*, *que*, *de*, *en* y *el*– son las palabras que ocurren más de 1000 veces suponen el 22,201 %. La primera palabra semántica aparece en la décima posición con una frecuencia absoluta de 478 (*falcón*) y si tuviéramos en cuenta la forma plural (de haberlo lematizado; *falcones* con 203 casos) ocuparía la séptima posición y hubiera supuesto el 1.760 % del texto.



Capítulo	Tokens	Tipos	Hápax	TTR	Tokens/Obra	Tipos/Obra	Hápax/Obra
AIC-22	437	205	142	46.9107551487414	1.12928650799804	4.84061393152302	6.73943996203132
AIC-23	460	191	118	41.5217391304348	1.18872263999793	4.51003541912633	5.60037968675842
AIC-24	274	145	98	52.9197080291971	0.708065224694421	3.42384887839433	4.65116279069767
AIC-25	493	193	128	39.1480730223124	1.27400056851952	4.55726092089728	6.0749881347888
AIC-26	317	155	103	48.8958990536278	0.819184949737706	3.65997638724911	4.88846701471286
AIC-27	924	326	207	35.2813852813853	2.38778199860454	7.69775678866588	9.82439487422876
AIC-28	478	210	139	43.9330543933054	1.23523787373698	4.95867768595041	6.59705742762221
AIC-29	1020	382	245	37.4509803921569	2.63586324521281	9.02007083825266	11.6279069767442
AIC-30	733	308	211	42.0190995907231	1.89420368504018	7.27272727272727	10.0142382534409
AIC-31	462	190	124	41.1255411255411	1.19389099930227	4.48642266824085	5.88514475557665
AIC-32	942	342	218	36.3057324840764	2.43429723234359	8.07556080283353	10.3464641670622
AIC-33	241	116	76	48.1327800829876	0.62278729617283	2.73907910271547	3.60702420503085
AIC-34	721	298	199	41.3314840499307	1.86319352921415	7.03659976387249	9.44470811580446
AIC-35	558	250	171	44.8028673835125	1.44197224591054	5.90318772136954	8.11580446131941
AIC-36	613	259	165	42.2512234910277	1.58410212677985	6.11570247933884	7.8310393250119
AIC-37	440	184	117	41.8181818181818	1.13703904695454	4.34474616292798	5.55291884195539
AIC-38	411	163	96	39.6593673965937	1.06209783704163	3.84887839433294	4.5562411010916
AIC-39	292	163	119	55.8219178082192	0.75458045843347	3.84887839433294	5.64784053156146
AIC-40	686	263	168	38.3381924198251	1.77274724138822	6.21015348288076	7.9734219269103
AIC-41	273	152	109	55.6776556776557	0.705481045042251	3.58913813459268	5.17323208353109
AIC-42	918	337	231	36.7102396514161	2.37227692069153	7.95749704840614	10.9634551495017
AIC-43	775	324	235	41.8064516129032	2.0027392304313	7.65053128689492	11.1532985287138
AIC-44	185	107	82	57.8378378378378	0.478073235651343	2.52656434474616	3.89178927384907
AIC-45	110	80	62	72.7272727272727	0.284259761738636	1.88902007083825	2.94257273778832
AIC-46	1234	463	311	37.5202593192869	3.18887769077706	10.9327036599764	14.7603227337447
AIC-47	1623	506	318	31.1768330252619	4.19412357547097	11.9480519480519	15.0925486473659
AIC-48	1009	427	301	42.3191278493558	2.60743726903894	10.0826446280992	14.2857142857143

Posteriormente, se ha analizado el texto dividido en capítulos con *stylo*, y se le ha pedido que establezca el análisis de grupos de acuerdo con la fórmula Classic Delta con un máximo de 100 palabras. El resultado es el dendrograma de la figura 9 en el que se observa que hay dos grandes grupos de tamaño distinto: el de la izquierda, agrupa los capítulos 1–8 y 42–44, 46 y 48, es decir, la mayoría de los capítulos atribuidos normalmente a Ayala. Esto confirma, *grosso modo*, lo que sabíamos: que hay dos estilos diferentes y que lo podemos atribuir, gracias al conocimiento filológico previo, a una cuestión de fuentes. Sin embargo, hay otros problemas *menores*: el capítulo VIII (AIC-09) se ha asentado dentro de la rama de Menino.

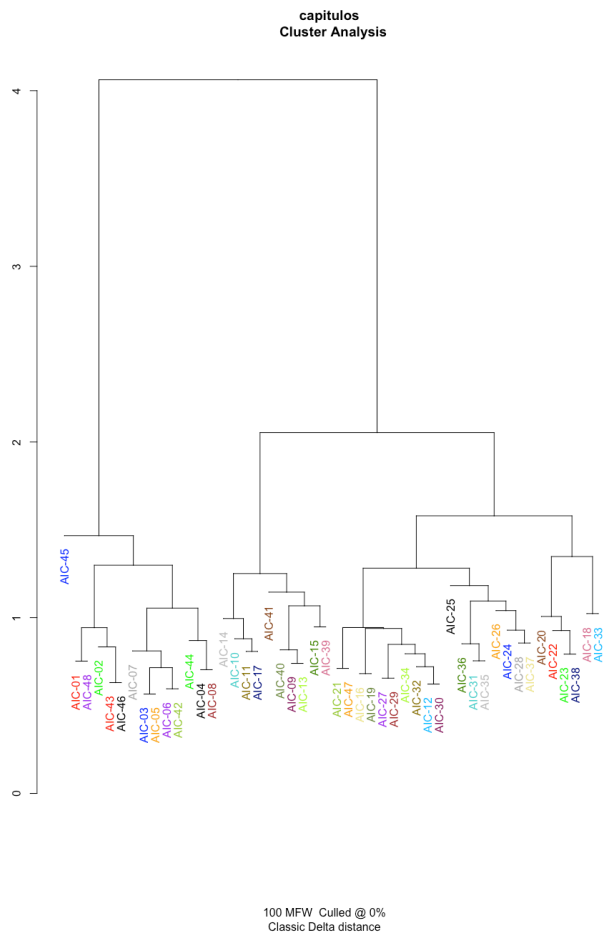


Fig. 9: Los capítulos del *Libro de la caza de las aves* (100 MFW)

Esto nos ha llevado a comprobar que ocurre si aumentamos el número de palabras que se han de tener en cuenta. Al aumentar a las 500 palabras más frecuentes, el resultado ha cambiado y el capítulo VIII (AIC-09) se ha situado cómodamente entre los que creemos de Ayala (fig. 10). Un nuevo aumento a las 1000 más frecuentes vuelve a ofrecer una geometría semejante como se puede observar en la figura 11.

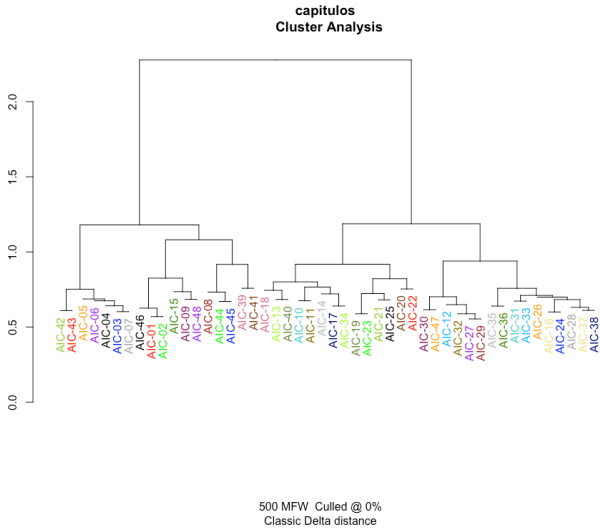


Fig. 10: Los capítulos del *Libro de la caza de las aves* (500 MFW)

Quizá el mayor problema es que es, sin lugar a dudas, el capítulo más extenso (él solo constituye el 18 % del texto) por lo que se procedió a analizar el texto dividido en segmentos de igual longitud. En un primer momento se dividió en pasajes de 1000 palabras para que *stylo* que considerara todas las palabras en cada fragmento. El resultado es una geometría semejante (fig. 12)), con variaciones en el orden de las hojas más bajas y con el capítulo VIII a caballo entre los dos grandes grupos (el capítulo VIII se halla entre los fragmentos 8 y 14), pero delimitado en un subgrupo de este segundo bloque.

Posteriormente, el tamaño de los fragmentos se redujo a 800 palabras, ya que esa era la media de palabras-token<sup>45</sup> para ejecutar la misma prueba con *stylo*. De nuevo, la geometría es semejante (fig. 13). El capítulo VIII, que se

<sup>45</sup> El último fragmento (AIC\_800-48) es un poco más grande que los demás con 1090 palabras.

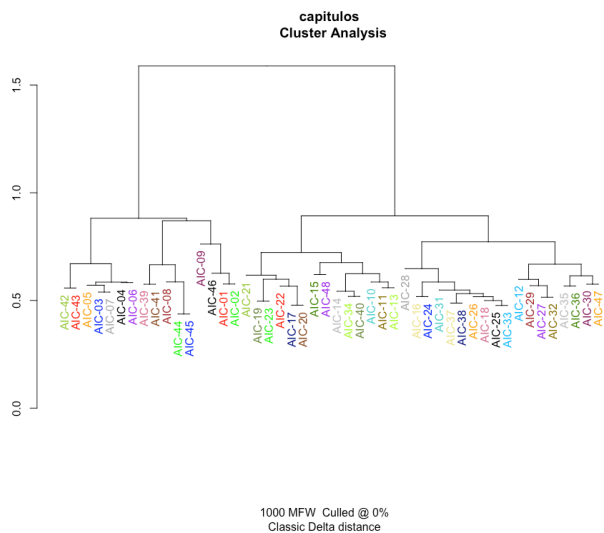


Fig. 11: Los capítulos del *Libro de la caza de las aves* (1000 MFW)

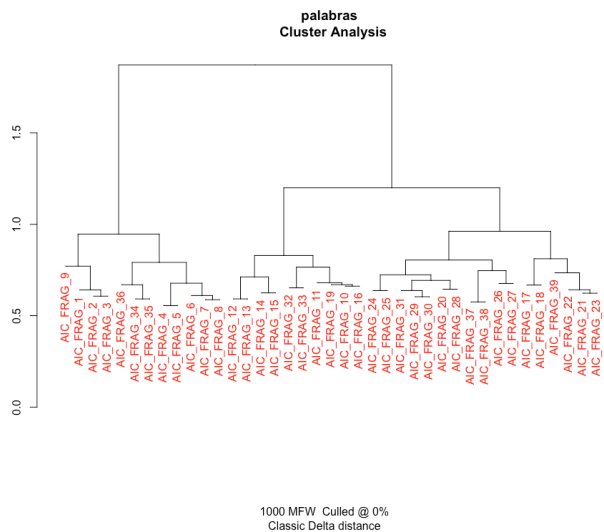


Fig. 12: El *Libro de la caza de las aves* dividido en fragmentos de 1000 palabras

encuentra repartido entre los fragmentos 11 y 20, se encuentra a caballo entre los dos grandes grupos, pero los pasajes del capítulo VIII se han agrupado dentro de un subgrupo homogéneo.

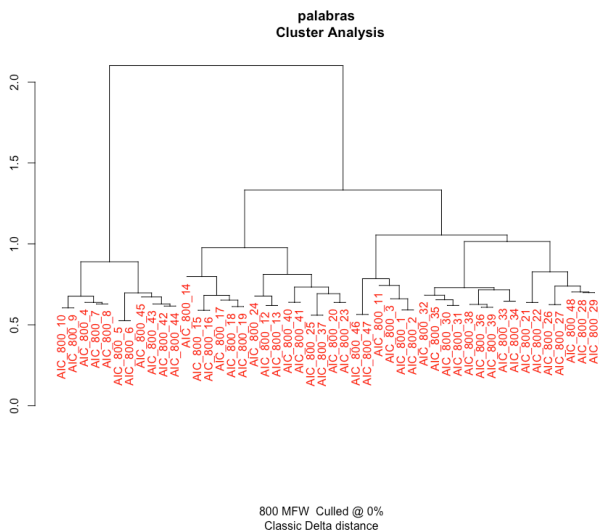


Fig. 13: El Libro de la caza de las aves dividido en fragmentos de 800 palabras

Por todo ello, es necesario revisar con mayor detenimiento este capítulo VIII y su atribución a López de Ayala, aunque no existe ninguna prueba textual de que este capítulo esté basado en la obra de Menino, ya que no tiene ningún lusismo ni falso amigo propio de la traducción ni, por otra parte, ningún testimonio de la obra de Menino habla, ni remotamente, de que el halconero portugués escribiera algo semejante a lo que se expone en este capítulo VIII.

### El Libro de la montería

Otro texto medieval de interés es el *Libro de la montería*. Este tiene una complicada historia de atribuciones desde los que defienden a ultranza que el autor fue Alfonso X a los que creen que fue su biznieto Alfonso XI<sup>46</sup>. Es un problema que no podrá resolverse porque no existe ningún escrito de Alfonso XI y

<sup>46</sup> José Manuel Fradejas Rueda, “El autor del Libro de la Montería: historia y comentario de seis siglos de controversia”, en *Actas II Congreso Internacional de la Asociación Hispánica de Literatura Medieval: Segovia, del 5 al 9 de octubre de 1987* (Alcalá de Henares: Universidad de Alcalá, 1992), 285–312.

de los presuntos libros de caza del rey Sabio, no se sabe nada, salvo que fue el promotor de la traducción al castellano del *Libro de los animales que cazan*<sup>47</sup>, texto que, como se ha visto cómo anteriormente, se asienta con toda naturalidad entre las obras de origen árabe del *Scriptorium* alfonsí.

Esta obra es un amplio tratado de caza mayor, *chasse courrant* o *par force*, dividido en tres libros: el primero, con 46 capítulos, trata de la teoría de caza; el segundo, que a su vez se subdivide en dos partes, se ocupa de las heridas y enfermedades de los perros de caza y, el tercero, con 30 capítulos, describe los cazaderos de los dominios del rey de Castilla y León.

Aunque hay muchas ediciones en el mercado, unas mejores que otras, y para mantenernos dentro de la línea de las características editoriales utilizadas en los casos de don Juan Manuel y el corpus alfonsí, para este análisis se parte de la transcripción semipaleográfica de acuerdo con el sistema del Hispanic Seminary of Medieval Studies que realizó Dennis Seniff en los años 1980 para su tesis doctoral<sup>48</sup> y que corresponde al ms. escorialense Y.II.19, que se estima como el mejor testimonio de esta obra, aunque tiene varios problemas. Uno de ellos es que debió de caer en manos de un encuadernador torpe que descolocó algunos folios, lo cual añade un aliciente a la exploración del texto, puesto que no se pudo segmentar por capítulos, al margen de que habrían sido muy desiguales de tamaño, desde 27 palabras (2.2.13) a los varios cientos de palabras de cualquier capítulo del libro tercero, por lo que se decidió segmentarlo en fragmentos de 1000 palabras, de manera que se consiguieron 117 ficheros de los que los 116 primeros tienen 1000 palabras, mientras que el último tiene 918 (una auténtica suerte)<sup>49</sup>.

Tras analizarlo con *stylo* y aplicando la distancia Classic Delta a las 1000 palabras más frecuentes, se obtuvo el dendrograma de la figura 14, ciertamente ilegible. Pero en él se observan dos grandes grupos muy separados entre sí. ¿A qué responde esta nítida división? Sencillamente, a que el grupo de la parte superior (fig. 15), el más pequeño, corresponde a una fuente específica del *Libro de la montería*: todos esos segmentos, aunque pueda haber solapamientos debido al corte cada 1000 palabras, se remontan a los libros cuarto y quinto del *Libro de los animales que cazan*, ese texto traducido del árabe que se asentaba con comodidad en medio del grupo árabe del *Scriptorium* alfonsí.

<sup>47</sup> Al-Bāzyār, *Libro de los animales que cazan* = *Kitāb al-Īawāriḥ*.

<sup>48</sup> Alfonso, *Libro de la montería: based on Escorial MS Y.II.19*, ed. Dennis P. Seniff (Madison: Hispanic Seminary of Medieval Studies, 1983).

<sup>49</sup> El texto tiene 115912 palabras-token y 7470 palabras-tipo, de las que 3650 son hápax.

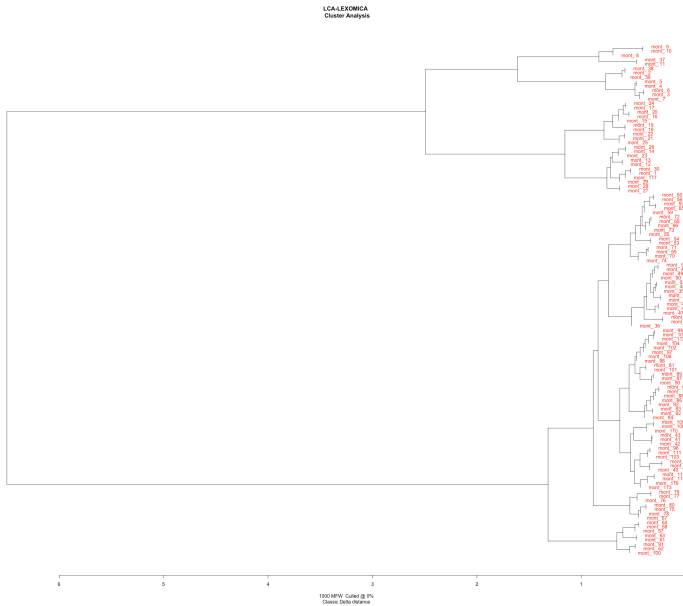


Fig. 14: Dendrograma del *Libro de la montería* dividido en fragmentos de 1000 palabras

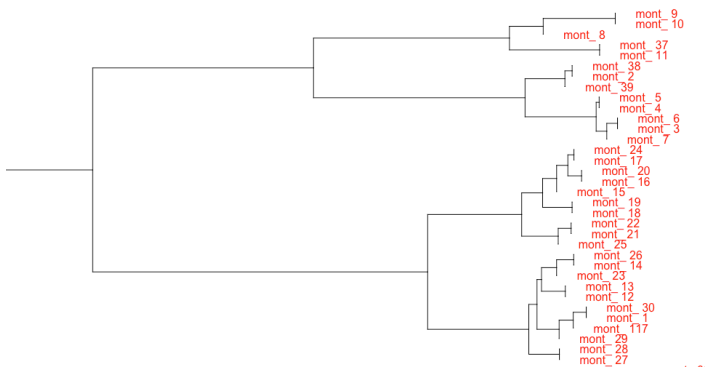


Fig. 15: Detalle de la figura 14

## Conclusión

Convencidos de que los métodos estilométricos son válidos y de que pueden ayudarnos a decidir si un texto es de un autor o no (o, en el peor de los casos, para descartar atribuciones tradicionales), el gran problema es si es necesario normalizar, estandarizar, regularizar los textos que se pretenden analizar cuando se trata de textos de la Edad Media. Se ha visto que los investigadores del Wheaton College abogan por el uso de ediciones críticas, pero estos experimentos con los textos castellanos, tanto si están normalizados, como es el caso de las obras de don Juan Manuel o el *Libro de la caza de las aves* de Ayala, como si se trata de una edición semipaleográfica (Juan Manuel –Ayerbe-Chaux–, *Scriptorium* alfonsí, *Libro de la montería*, parece que no cambian sustancialmente los resultados. Somos conscientes de que estas son unas pequeñas pruebas sobre textos que sabía cómo debían de funcionar: el *Libro de la montería*, el *Libro de la caza de las aves*, el *Scriptorium* alfonsí; ahora bien, queda una pregunta final que aún no sé responder: ¿Qué es lo que hace que un sistema automatizado, que solo ha tenido en cuenta 100 palabras, haya detectado, sin error alguno, que un grupo de obras son traducción del árabe y el otro gran grupo son obras con fuentes, no diré traducción, latinas? Cuestión aparte es el errático comportamiento del capítulo VIII del *Libro de la caza de las aves*.

Decididamente, los métodos estilométricos (y lexómicos) puede que no nos digan nada nuevo de un texto, aparte de confirmar lo que ya sabíamos (que no es poco), pero su importancia radica en las posibilidades que puede ofrecer para futuras investigaciones.



# Fernando de Herrera y la autoría de *Versos*

## Un primer acercamiento al drama textual desde la Estilometría

Laura Hernández Lorenzo (Sevilla)

**RESUMEN:** Además de tratarse de uno de los problemas textuales y de autoría más fascinantes de la poesía del Siglo de Oro, la cuestión en torno a la autenticidad de parte de la obra lírica de Fernando de Herrera ha dividido a la crítica en dos grandes grupos, dando lugar a un *impasse* que continúa hasta hoy. En este artículo se pretende arrojar nueva luz aplicando metodologías provenientes de las Humanidades Digitales y la Estilometría, empleando una técnica desarrollada recientemente por Maciej Eder, Rolling Stylometry, para detectar si hay varias manos en las obras poéticas dubitadas de Herrera, además de confirmar o rechazar la existencia de la mano de Francisco de Pacheco, principal argumento de los estudiosos que niegan la autoría herreriana.

**ZUSAMMENFASSUNG:** Es handelt sich nicht nur um eines der faszinierendsten philologischen Probleme der Lyrik des Siglo de Oro, sondern die Frage nach der Autorschaft und Authentizität des poetischen Werks von Fernando de Herrera trennt die Fachwelt in zwei Gruppen, die bis heute unversöhnlich erscheinen. Durch die Anwendung neuer Methoden aus dem Bereich der Digital Humanities, hier der Stilometrie und genauer der Rolling Stylometry (jüngst von Maciej Eder entwickelt) soll die Frage unter neuem Licht betrachtet werden. Dazu wird die These untersucht, ob verschiedene poetische Werke von Herrera möglicherweise von unterschiedlichen, mehreren Personen verfasst wurden. Der in der bisherigen Forschung als wahrscheinlichste Kandidat für ein solches „Ghostwriting“ ist der Maler und Herausgeber Francisco de Pacheco.

**PALABRAS CLAVE:** estilometría; Herrera, Fernando de; poesía del Siglo de Oro; humanidades digitales; estilística computacional

**SCHLAGWÖRTER:** Stilometrie; Herrera, Fernando de; Siglo de Oro; Digital Humanities

### 1. El drama textual herreriano

El problema textual y de autoría que afecta a la obra poética de Fernando de Herrera (1534–1597) es uno de los más apasionantes de la poesía del Siglo de Oro. A las dificultades de transmisión textual de su poesía, particularmente

---

<sup>o</sup> Este trabajo forma parte de un proyecto de tesis doctoral financiado por una ayuda de Formación de Profesorado Universitario (FPU) del Ministerio de Educación y Ciencia español (MECD) y dirigido por el Dr. Juan Montero, a quien agradezco sus orientaciones y consejos.

oscuras dentro de la ya complicada circulación y transmisión que caracteriza a los textos poéticos áureos, tenemos que sumar el intenso y prolongado debate crítico sobre la autenticidad de parte de su obra y la importancia de la poesía herreriana dentro del panorama lírico español del XVI, en el que, si seguimos a Dámaso Alonso y Antonio Vilanova, esta sería fundamental para entender la evolución del lenguaje poético desde Garcilaso a Góngora.<sup>1</sup> Los textos del poeta áureo más sobresaliente de la llamada escuela poética sevillana y conocido con el sobrenombre de “El Divino” se nos han transmitido a través de dos ediciones impresas y varios manuscritos. Realizamos a continuación un somero estado de la cuestión de los textos poéticos herrerianos y de la polémica conocida como el “drama textual” herreriano.

La primera edición, *Algunas obras*, conocida como *H*, fue publicada en 1582 y ha sido considerada tradicionalmente como la más autorizada. Con 91 poemas, constituye una selección de la poesía herreriana preparada por el propio Herrera. Tenemos constancia, además, de que el propio poeta participó en el proceso de impresión, revisando pruebas de imprenta. Por todo ello, *H* es considerada prácticamente como un autógrafo.

La polémica y el debate crítico se centran en la segunda edición, *Versos de Fernando de Herrera*, publicada de forma póstuma en 1619 y preparada por el pintor Francisco Pacheco (1564–1644), conocida como *P*. Esta contiene 365 poemas supuestamente herrerianos, de los cuales 88 ya aparecían en *H* y 70 en otros manuscritos de poesía herreriana. De estos la inmensa mayoría han sufrido correcciones y modificaciones que en algunos casos se limitan a palabras concretas y en otros se trata de una auténtica reescritura de los poemas. Esta edición incluye también 205 poemas nuevos de Herrera, no conocidos hasta la fecha ni transmitidos a través de otras fuentes.

Además de estas dos ediciones, nos ha llegado un corpus de poemas sueltos herrerianos, de los cuales destacan los incluidos en el manuscrito conservado en la Biblioteca Nacional de España, *Cisnes del Betis*, publicado por José Manuel Blecua y conocido como *B*. Este contiene 130 poemas de Herrera, 84 de los cuales son comunes a *H* y/o *P*.

Dentro de esta compleja transmisión de las obras poéticas herrerianas, la cuestión textual se ha centrado fundamentalmente en la discusión sobre la autenticidad de *Versos* por su carácter póstumo y por las diferencias que pre-

---

<sup>1</sup> Antonio Vilanova, “Fernando de Herrera”, in *Historia general de las literaturas hispánicas*, II, publ. bajo la dir. de Guillermo Díaz-Plaja (Barcelona: Barna, 1968), 689–751.

senta con *Algunas obras*, así como las oscuras circunstancias que rodean a la edición.

La única información que tenemos sobre su proceso de preparación se encuentra en los textos preliminares, caracterizados por la ambigüedad. En primer lugar, en el prólogo de Francisco Medina se nos habla del olvido que padecían las obras de Herrera, elogiándose la labor de Pacheco en recoger sus obras perdidas, pero no se nos dan más detalles sobre las causas de esta pérdida y olvido. El licenciado Enrique Duarte da algunos detalles más en su prólogo. Declara que Herrera tenía preparado un volumen de sus obras poéticas para darlo a la imprenta cuando falleció y que estas se han podido recuperar del “naufragio” gracias a la diligencia de Pacheco:

I es cierto que su memoria uviera quedado sepultada en perpetuo olvido, si Francisco Pacheco Celebre Pintor de nuestra Ciudad, i afectuoso imitador de sus escritos, no uviera recogido con particular diligencia i cuidado algunos cuadernos i borradores que escaparon d’el naufragio, en que pocos días des- / pues de su muerte perecieron todas sus obras Poeticas; que el tenia corregidas de ultima mano, i encuadradas para darlas a la Empreñta. Dexo en silencio la culpa d’esta perdida, porque soi enemigo de sacar en publico agenas culpas, i juzgo por merecedor de gran premio, al que con tantas veras à procurado restaurarla, hurtando muchas oras de su mas forçosa y precisa ocupacion; porque no solo copiò una i dos vezes de su mano lo que aora nos ofrece, pero cumplió lo que faltaba de otros papeles sueltos, que avian venido a manos de diferentes personas, de quien los uvo; i aunque todo ello sea d’el mesmo Autor es cosa cierta, que lo que el tenia escogido, i perficionado para sacar a luz seria de mayor, i de mas acabada perfeccion.<sup>2</sup>

Aunque estos testimonios han hecho desconfiar ya a algunos críticos, sin duda el que ha causado más interrogantes ha sido el poema del propio Pacheco que abre el libro, en el que dice: “Goza, ô Nacion osada, el don fecundo | que t’ofresco, en la forma verdadera | qu’imaginè, d’el culto i gran Herrera; | i el fruto de su ingenio, alto i profundo.”<sup>3</sup> ¿Cómo tenemos que interpretar este “en la forma verdadera | qu’imaginè”?

Quevedo fue el primero en desconfiar y atacar la edición preparada por Pacheco en el prólogo a sus lectores de su edición de las poesías de Francisco de la Torre (Madrid, 1631). Sin embargo, su objetivo es por un lado demostrar que Herrera imitó en su poesía la de Francisco de la Torre y, por otro,

<sup>2</sup> Fernando de Herrera, *Obra poética*, ed. José Manuel Blecua (Madrid: Boletín de la Real Academia Española, 1975), t. II: 25–6.

<sup>3</sup> Herrera, *Obra poética*, t. II, 31.

de acuerdo con Peter M. Komanecky, criticar la edición de Pacheco por su mayor cercanía al estilo poético de Góngora:

the real point Quevedo was trying to make was not that the poems as they appear in Pacheco's edition were apocryphal [...] but rather that Herrera should not have developed a more 'baroque' style in his later years, and if he did, then he was worthy of the strongest possible censure.<sup>4</sup>

En cualquier caso, la desconfianza hacia *P* aumentó cuando en el siglo xx Adolphe Coster realizó la primera edición crítica de la poesía herreriana en 1908, en cuyo prólogo defiende que los poemas publicados en *P* no tienen carácter auténtico y que *P* recogería poemas deteriorados por copistas y correctores, o rechazados por el propio Herrera, si bien unos años más tarde, en 1914, decide publicar *P* y reconoce el interés de los poemas publicados por Pacheco.<sup>5</sup>

Con la publicación del manuscrito *B* en 1948 por José Manuel Blecua se descubre que muchos poemas de *P* que no aparecían en *H* sí estaban en este manuscrito y se confirma la tendencia de Herrera a corregir sus escritos a la que hacen referencia los amigos:

I esta fue la causa, de que Fernando de Herrera pareciesse tan difícil, i tardo en aprovar las obras, que via, no porque no admirasse las suyas, que de ninguna cosa estaba mas lexos; porque, como a ombre a quien el uso i exercicio de aquellas cosas avia dado una mui entera noticia de los preceitos mas ocultos de l'arte, le satisfazian pocas, i sus oidos como capaces de otras mayores desseavan siempre alguna de consumada perfeccion; de que pueden dar testimonio los borradores de sus Versos, que despues de limados muchas vezes, i en espacio de años enteros, apenas le contentavan; i assi desechó muchos, que pudieran ser estimados de los mas entendidos en esta profession.<sup>6</sup>

A partir de entonces, las diferencias que presenta *P* con respecto a *H* han dividido a la crítica en dos grandes grupos:

Por un lado, el primero se iniciaría con la postura de Adolphe Coster, cuyo principal defensor ha sido el reconocido estudioso José Manuel Blecua, y a la que se han adherido otros críticos como Cristóbal Cuevas, Vicente García de Diego o A. David Kossoff, que niegan la autoría herreriana de *P* y sospechan

<sup>4</sup> Peter Komanecky, "Quevedo's Notes on Herrera: The Involvement of Francisco de La Torre in the Controversy Over Góngora", *Bulletin of Hispanic Studies* 52, no. 2 (1975): 123–33, cit. 125.

<sup>5</sup> Fernando de Herrera, *Algunas obras de Fernando de Herrera*, ed. Adolphe Coster (Paris: Champion, 1908); Fernando de Herrera, *Versos de Fernando de Herrera*, ed. Adolphe Coster (Estrasburgo: Bibliotheca Romanica, 1914).

<sup>6</sup> Herrera, *Obra poética*, t. II, 23.

de una segunda mano, presumiblemente la de Pacheco, el editor, que además de pintor era poeta y gran admirador de Herrera:<sup>7</sup>

[...] solo admitiendo la posibilidad de que Pacheco emendase en la “forma verdadera” que imaginó la obra poética de Herrera tienen respuesta las muchas interrogaciones planteadas a lo largo de estas páginas [...] Su único defecto consistió en que se le fué la mano y restauró a veces con demasiados tintes viejos y nuevos las versiones que le ofrecían los cuadernos, borradores y papeles sueltos remitidos con tanta diligencia.<sup>8</sup>

Por otro, los hispanistas italianos Salvatore Battaglia, Oreste Macrí e Inoria Pepe Sarno defienden la autoría herreriana de *P*. Para ellos, las variantes serían de Herrera y mostrarían una evolución de signo prebarroco.<sup>9</sup>

En un primer momento, la hipótesis defendida por los que negaban la autoría de *P* fue que Pacheco había utilizado borradores de Herrera que recogerían versiones anteriores a los poemas publicados en 1582 y, por tanto, descartadas por El Divino. Sin embargo, Salvatore Battaglia a través del análisis de las variantes de los poemas comunes a *B*, *H* y *P* consiguió convencer a casi la totalidad de la crítica de que el orden de los poemas es  $B \rightarrow H \rightarrow P$ ,<sup>10</sup> y a día de hoy solo el hispanista Ricardo Senabre sigue defendiendo que *P* recoge borradores primitivos.<sup>11</sup> Consecuentemente, el debate se ha centrado desde esta fecha definitivamente en la autenticidad de *P*. En los años 50, José Manuel Blecua y Oreste Macrí publicaron diversos artículos intentando cada uno convencer al otro de su postura sin éxito. Los editores modernos de la

<sup>7</sup> David Kossoff, “Another Herrera Autograph: Two Variant Sonnets”, *Hispanic Review* 33, no. 3 (1965): 318–25; David Kossoff, “Herrera, editor de un poema”, in *Homenaje a Rodríguez-Moñino* (Madrid: Castalia, 1966), 283–90; David Kossoff, “Algunas variantes de versos de Herrera”, *Nueva revista de filología hispánica* XI (1957): 57–63; David Kossoff, “Algo más sobre largo y luengo en Herrera”, *Revista de filología española* XLI (1957): 401–10; David Kossoff, *Vocabulario de la obra poética de Herrera* (Madrid: RAE, 1966); José Manuel Blecua, “De nuevo sobre los textos poéticos de Herrera”, *Boletín de La Real Academia Española* XXXVIII (1958): 377–408; José Manuel Blecua, “Los textos poéticos de Fernando de Herrera”, *Archivum* IV (1954): 247–63; Cristóbal Cuevas, “La cuestión textual”, in *Poesía castellana original completa* (Madrid: Cátedra, 1985), 87–99.

<sup>8</sup> José Manuel Blecua, “De nuevo sobre los textos poéticos de Herrera”, *Boletín de La Real Academia Española* XXXVIII (1958): 406–7.

<sup>9</sup> El profesor Oreste Macrí ha sido quien más ha estudiado esta supuesta evolución de la poesía herreriana con diversos artículos que fueron en aumento y se convirtieron finalmente en la monografía Oreste Macrí, *Fernando de Herrera* (Madrid: Gredos, 1972).

<sup>10</sup> Salvatore Battaglia, “Per il testo di Fernando de Herrera”, *Filologia Romanza*, no. I (1954): 51–88.

<sup>11</sup> Ricardo Senabre, “Los textos emendados de Herrera”, *Edad de Oro* IV (1985): 179–93.

poesía de Herrera han seguido las pautas de Blecua, tomando como modelo su edición de la poesía herreriana.<sup>12</sup>

En los últimos años, algunos estudiosos entre los que destacan Inoria Pepe Sarno y Juan Montero han entrevisto la posibilidad de realizar avances en el asunto aplicando la metodología utilizada por Battaglia, esto es, el análisis crítico de las variantes entre los diferentes testimonios a textos sueltos y, en ocasiones, desconocidos de El Divino, que han supuesto interesantes aportaciones.<sup>13</sup>

Sin embargo, parece que no se puede llegar a conclusiones definitivas con la metodología adoptada hasta ahora, por lo que se hace necesario un nuevo enfoque que abarque todo el corpus y que aproveche las posibilidades que ofrecen hoy las Humanidades Digitales y, más concretamente, la Estilometría.<sup>14</sup>

## 2. La Estilometría<sup>15</sup>

La Estilometría, también llamada Estilística Computacional, es una de las ramas más destacadas a día de hoy dentro de las Humanidades Digitales. Sus

<sup>12</sup> Herrera, *Obra poética*.

<sup>13</sup> Juan Montero, “Una versión inédita (con algunas variantes) de la canción ‘Al sueño’ de F. Herrera”, *Cuadernos de investigación filológica XII–XIII* (1987): 117–32; Juan Montero, “Dos textos poéticos de Fernando de Herrera con variantes y un posible soneto desconocido (más una lira antiherreriana)”, *Bulletin hispanique* 107, no. 2 (2005): 605–16; Inoria Pepe Sarno, “La ‘luz’ de la ‘Aurora’: variantes en dos sonetos de Fernando de Herrera”, in *AIH. VIII Congreso*, 1983; Inoria Pepe Sarno, “Marte-Don Juan in una canzone di Fernando de Herrera”, *Archivo hispalense*, 1986; Inoria Pepe Sarno, “Bianco il ghiaccio, non il velo: ritocchi e metamorfosi in un sonetto di herrera”, *Strumenti Critici*, no. XLVI (1981): 458–71; Inoria Pepe Sarno, “Il senso della storia in alcuni sonetti di Fernando de Herrera”, in *Estudios en homenaje a D. Claudio Sánchez Albornoz en sus 90 años* (Ávila-Buenos Aires: Fundación “Claudio Sánchez Albornoz/Instituto de Historia de España”, 1990), 307–19.

<sup>14</sup> Existen ya un par de estudios que aplican enfoques de corpus o computacionales a la obra de Herrera, aunque todavía no a la cuestión del drama textual: E. Espinilla Buisán et al., *Garcilaso y Herrera: vocabulario, frecuencias y concordancias de las anotaciones a la poesía de Garcilaso de Fernando de Herrera* (Barcelona: Edicions de la Universitat de Barcelona, 2003); Laura Hernández Lorenzo, “The Poetic Word of Fernando de Herrera: An Approach through Corpus and Computational Linguistics”, in *CILC 2016. 8<sup>th</sup> International Conference on Corpus Linguistics: EPiC Series in Language and Linguistics*, ed. Antonio Moreno Ortiz and Chantal Pérez-Hernández, vol. 1, 2016, 170–80.

<sup>15</sup> Un estado de la cuestión en español bastante completo en cuanto a los orígenes de la Estilometría lo encontramos en José Manuel Fradejas Rueda, “El análisis estilométrico aplicado a la literatura española: las novelas policíacas históricas”, *Caracteres: estudios culturales y críticos de la esfera digital* 5, no. 2 (2016): 196–246. En este hemos basado el nuestro y a él remitimos para más información.

comienzos se remontan a la llamada *Shakespeare Authorship Question* o la cuestión de si William Shakespeare escribió todo lo que se le atribuye. Los investigadores Augustus de Morgan (1806–1871) y Thomas C. Mendenhall (1841–1924) compararon la longitud de las palabras de las obras atribuidas a Shakespeare con las de otros candidatos como Vere, Bacon y Marlowe. Se trataba, sin embargo, de un procedimiento muy mecánico e insuficiente para llegar a conclusiones sólidas sobre la autoría de un texto. El primer gran éxito fue el caso de los *Federalist Papers* en el que los investigadores Mostaller y Wallace utilizaron por primera vez la frecuencia de las palabras gramaticales o palabras función (*function words*) y consiguieron establecer la autoría de doce de los artículos, reclamada tanto por Hamilton como por Madison. Desde entonces, se han realizado numerosos estudios de autoría utilizando estudios estilométricos, siendo uno de los más sonados el caso de *The Cuckoo's Calling*, firmada por un tal Robert Galbraith, que resultó ser un pseudónimo de J. K. Rowling, tal y como Patrick Juola había obtenido en todos sus análisis estilométricos.<sup>16</sup>

En el caso de la literatura española, sin embargo, los estudios estilométricos son aún bastante escasos, sobre todo si los comparamos con los aplicados a textos en inglés.<sup>17</sup> Entre ellos destaca la aplicación de la Estilometría a las dos obras anónimas más famosas de la literatura española, el *Lazarillo de Tormes* y el *Quijote* de Avellaneda. Aunque no se ha podido determinar la autoría de estas obras, en el caso del *Lazarillo* sí se han podido descartar candidatos<sup>18</sup> y en el del *Quijote* de Avellaneda ha quedado resaltada una, por otra parte, no del todo sorprendente cercanía de esta obra con las del propio Cervantes.<sup>19</sup>

<sup>16</sup> Patrick Juola, “The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions”, *Digital Scholarship in the Humanities* 30, no. 1 (2015): 100–13.

<sup>17</sup> Esta situación ha llevado a la aparición de artículos divulgativos en español sobre los métodos estilométricos como el ya mencionado Fradejas Rueda, “El análisis estilométrico aplicado a la literatura española: las novelas policíacas históricas”; y otros como José Calvo Tello, “Entendiendo delta desde las humanidades”, *Caracteres: estudios culturales y críticos de la esfera digital* 5, no. 1 (2016): 140–76.

<sup>18</sup> Nanette Rißler-Pipka, “Digital Humanities und die romanische Literaturwissenschaft: der Autorschaftsstreit um den Lazarillo de Tormes”, *Romanische Forschungen* 128, no. 3 (2016): 316–42; Javier de la Rosa and Juan Luis Suárez, “The Life of Lazarillo de Tormes and of His Machine Learning Adversities: Non-Traditional Authorship Attribution Techniques in the Context of the Lazarillo”, *Lemir* 20 (2016): 373–438.

<sup>19</sup> Javier Blasco, “Avellaneda desde la estilometría”, in *Cervantes: los viajes y los días*, ed. Pedro Ruiz Pérez (Madrid: Prosa Barroca y SIAL Ediciones, 2016), 97–116; Nanette Rißler-Pipka, “Avellaneda y los problemas de la identificación del autor: propuestas para una investigación con nuevas herramientas digitales”, in *El otro Don Quijote: la continuación de Fernández de Avellaneda*

Por otra parte, en varios congresos y reuniones científicas he podido conocer otros trabajos estilométricos aplicados a la literatura española que se están llevando a cabo sobre textos medievales castellanos, el teatro barroco o la novela de la Edad de Plata. Por lo general, la poesía lírica suele ser el género literario menos abordado a través de métodos computacionales debido a la mayor presencia de subjetividad y este es también el caso en la aplicación de la Estilometría a la literatura española, aunque destaca el análisis que está acometiendo la profesora Nanette Rißler-Pipka con la poesía de Góngora y Picasso y el publicado por Antonio Rojas Castro sobre la fábula mitológica del XVII.<sup>20</sup>

### 3. Preparación de los textos

Para aplicar los métodos computacionales y estilométricos al problema textual de autoría de la poesía herreriana resultaba necesario en primer lugar contar con los textos de Herrera en formato digital. Puesto que la obra poética solo se había digitalizado de forma muy parcial en la web, el primer paso necesario fue acometer la digitalización de esta en su totalidad, la cual se llevó a cabo aplicando el software de OCR (Optical Character Recognition) Adobe Acrobat Pro sobre la edición más autorizada, la de José Manuel Blecua, que contiene además la totalidad de los textos poéticos herrerianos.<sup>21</sup> A continuación, se revisó el corpus y se corrigieron los errores del OCR. Debido a que las particularidades ortográficas de Herrera hubieran entorpecido el análisis estilométrico, aunque estas se mantuvieron por su interés como otro indicador de cambio entre los textos herrerianos, se creó un corpus paralelo en el que esta fue modernizada de acuerdo con las reglas actuales de la Real Academia de la Lengua.

Por otra parte, Francisco Pacheco, el editor de *P*, es el principal sospechoso de los críticos que niegan la autoría herreriana de esta obra, que piensan que pudo haber retocado los poemas para adaptarlos al gusto más barroco del momento de publicación:

Sin embargo, los textos de Pacheco presentan la mixtura más informe y extraña que conoce la poesía española del siglo XVI, y sólo un editor que en 1619 (en pleno furor culterano) quiere dar un tinte nuevo y viejo a la vez a la poesía,

---

y sus efectos, ed. Hanno Ehrlicher (Augsburgo, 2016), 27–51.

<sup>20</sup> Antonio Rojas Castro, “Luis de Góngora y la fábula mitológica del Siglo de Oro: clasificación de textos y análisis léxico con métodos informáticos”, *Studia Aurea* 11 (2017): 111–42.

<sup>21</sup> Herrera, *Obra poética*.



“en la forma verdadera que imaginó”, puede atreverse a cambiar sistemáticamente el pronombre os en vos [...].<sup>22</sup>

Tomando esto en consideración, hemos recopilado su poesía de diferentes fuentes, principalmente del *Arte de la pintura* y del *Libro de los retratos*, junto con la *Biblioteca de autores españoles*, y los poemas recogidos por Asensio.<sup>23</sup> Este corpus fue también modernizado siguiendo los mismos criterios que con la poesía herreriana.<sup>24</sup>

#### 4. Análisis estilométrico

Como hemos visto, el problema de autoría de *P* tiene una gran complejidad, ya que no se trata solo de confirmar si la obra es de Herrera o no lo es, sino más bien de detectar si hay otra mano aparte de la de El Divino, que indudablemente está en las partes coincidentes con *B* y *H* que no han sido modificadas.

En consecuencia, este primer acercamiento estilométrico a los problemas de autoría en la poesía de Herrera se basa en uno de los métodos desarrollados para analizar trabajos colaborativos u obras en las que intervienen dos o más manos, Rolling Stylometry, presentado por Maciej Eder en un artículo publicado en 2016 en *Digital Scholarship in the Humanities (DSIH)*, la revista de la Alliance of Digital Humanities Organisations (ADHO).<sup>25</sup> En este trabajo, Eder

<sup>22</sup> Blecua, “De nuevo sobre los textos poéticos de Herrera”, 396.

<sup>23</sup> José María Asensio, *Francisco Pacheco: sus obras artísticas y literarias. introducción e historia del libro de descripción de verdaderos retratos de ilustres y memorables varones que dejó inédito* (Sevilla: Rasco, 1886), esta obra se ha consultado a través de la Biblioteca Virtual de Andalucía en el siguiente enlace: [www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1017777](http://www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1017777), consulta: 14/02/2018; Adolfo de Castro, ed., *Biblioteca de autores españoles, desde la formación del lenguaje hasta nuestros días: poetas líricos de los siglos XVI y XVII* (Madrid: Rivadeneyra, 1854), esta obra se ha consultado a través del repositorio archive.org en el siguiente enlace: <https://archive.org/details/bibliotecadeauto32madruoft>, consulta: 14/02/2018; Francisco Pacheco, *Libro de descripción de verdaderos retratos de ilustres y memorables varones* (Sevilla: Rafael Tarasco, 1881), esta obra se ha consultado a través de la Biblioteca Virtual de Andalucía en el siguiente enlace: [www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1040827](http://www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1040827), consulta: 14/02/2018; Francisco Pacheco, *Arte de la pintura ...* (Madrid: Galiano, 1866), esta obra se ha consultado a través de Google Books en el siguiente enlace: [https://books.google.de/books/about/Arte\\_de\\_la\\_pintura.html?id=litPAAAYAAJ&redir\\_esc=y](https://books.google.de/books/about/Arte_de_la_pintura.html?id=litPAAAYAAJ&redir_esc=y), consulta: 14/02/2018.

<sup>24</sup> A la hora de preparar esta recopilación, me han resultado de enorme utilidad los materiales que muy generosamente me ha cedido el Dr. Juan Montero, a quien se lo agradezco mucho.

<sup>25</sup> Maciej Eder, “Rolling Stylometry”, *Digital Scholarship in the Humanities* 31, no. 3 (2016): 457–69.

aplica tres técnicas de Machine Learning (Machine Learning classifiers) cuya eficacia ha sido probada a través del paquete *stylo* de R:<sup>26</sup> support vector machines (svm), nearest shrunken centroids (NSC) y la Delta de Burrows.<sup>27</sup> Con estas aborda tres casos de obras con varias manos: el *Roman de la Rose*, poema del medievo francés escrito por Guillaume de Lorris y Jean de Meun; la Biblia de la Reina Sofía, obra con un complicado problema de transmisión en la que han participado varios traductores; y los trabajos colaborativos de Joseph Conrad y Ford Madox Ford. En todos los casos, las señales autoriales se detectan con buenos resultados.

Una versión anterior de Rolling Stylometry, Rolling Delta, que aplica solo la medida de Burrows, había sido aplicada anteriormente a los trabajos colaborativos de Conrad y Ford por los investigadores Jan Rybicki, David Hoover y Mike Kestemont<sup>28</sup> y a los trabajos colaborativos de Charles Dickens por el profesor Tomoji Tabata.<sup>29</sup> Por otra parte, tras la publicación del artículo de Rolling Stylometry, han aparecido más trabajos que aplican esta técnica a otras obras en inglés.<sup>30</sup>

Previamente al análisis con los textos herrerianos, se decidió realizar primero una prueba con un texto de la literatura en español con dos manos y en el que supiéramos antes de realizar el análisis qué parte fue escrita por cada autor. El objetivo era confirmar la eficacia de este procedimiento con textos literarios en español. A tal efecto se aplicó la técnica sobre la obra barroca *Amor es más laberinto*, co-escrita por Sor Juana Inés de la Cruz y Fray Juan de Guevara.<sup>31</sup> Las marcas ‘a’ y ‘b’ marcan las separaciones en jornadas de la obra.

<sup>26</sup> Maciej Eder, Jan Rybicki, and Mike Kestemont, “Stylometry with R: A Package for Computational Text Analysis”, *The R Journal* XX (2008): 1–15.

<sup>27</sup> John Burrows, “Delta: a Measure of Stylistic Difference and a Guide to Likely Authorship”, *Literary and Linguistic Computing* 17, no. 3 (2002): 267–87.

<sup>28</sup> Jan Rybicki, David Hoover, and Mike Kestemont, “Collaborative Authorship: Conrad, Ford and Rolling Delta”, *Literary and Linguistic Computing* 29, no. 3 (2014): 422–31.

<sup>29</sup> Tomoji Tabata, “Stylometry of Collaborations: Dickens, Collins and Their Collaborative Writings”; in *Digital Humanities 2014: Conference Abstracts* (Lausanne: EPFL-UNIL, 2014), 378–80.

<sup>30</sup> Hartmut Ilsemann, “More News on Sir Thomas More”, *Digital Scholarship in the Humanities*, 2017, 1–13; Alexander A.G. Gladwin, Matthew J. Lavin, and Daniel M. Look, “Stylometry and Collaborative Authorship: Eddy, Lovecraft, and ‘The Loved Dead’”, *Digital Scholarship in the Humanities* 32, no. 1 (2017): 123–40.

<sup>31</sup> La colaboración en la escritura de esta obra dramática consta desde la edición princeps de 1692, leyéndose en la portada de la comedia lo siguiente: “AMOR ES MAS LABERYNTO, | COMEDIA, | DE LA QVAL LAS JORNADAS PRIMERA, | y tercera son de la Madre Juana, y la segunda del /Lic. D. Juan de Guevara, Ingenio Conocido | de la Ciudad de Mexico” [folio 389]. La edición puede consultarse a través de la Biblioteca Virtual Miguel de Cervantes en el siguiente enla-



Fig. 1: *Amor es más laberinto*. SVM

Como puede observarse en la figura 1, los resultados reflejan lo que ya sabíamos, que el primer y el tercer acto fueron escritos por Sor Juana y el segundo por Fray Juan de Guevara. La probabilidad de autoría es incluso del 100 % en algunas partes del gráfico.

En este análisis estilométrico de la poesía herreriana, se aplicarán las tres técnicas incluidas en Rolling Stylometry a través del paquete *stylo* en R y utilizando como corpus de referencia por un lado, las obras poéticas que sabemos indudablemente que pertenecen a Herrera, es decir, los poemas sueltos y *H*; y, por otro, los textos poéticos que hemos recopilado de Pacheco. El ‘test corpus’ será tanto *P* como un corpus que hemos preparado con los 205 poemas nuevos de esta edición, al que hemos llamado *P2*. Nuestro objetivo es medir de forma más científica y objetiva la probabilidad de la autoría herreriana de *P*, así como rechazar o confirmar en la medida de lo posible la existencia de una segunda mano y más concretamente la de Francisco Pacheco.

Comenzamos utilizando Rolling SVM y en el primer experimento utilizaremos las mismas variables que utiliza Maciej Eder en su artículo, esto es, muestras de 5000 palabras con solapamiento de 4500. Además, en la visualización de los resultados la separación entre libros de *P* se muestra a través de las marcas ‘a’ y ‘b’.

Con Rolling SVM y las 100 palabras más frecuentes sobre *P* y *P2*, obtenemos las figuras 2 y 3, en las que, como puede verse, la probabilidad de mano de Herrera es casi del 100 %.

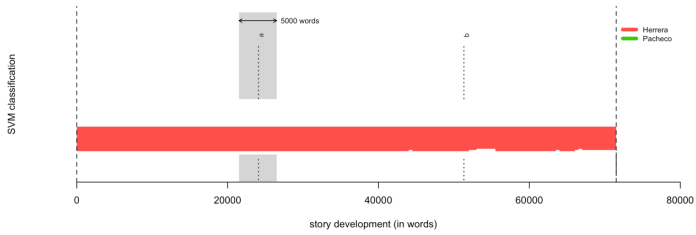


Fig. 2: P. SVM. Samples of 5000 words. 100 MFW

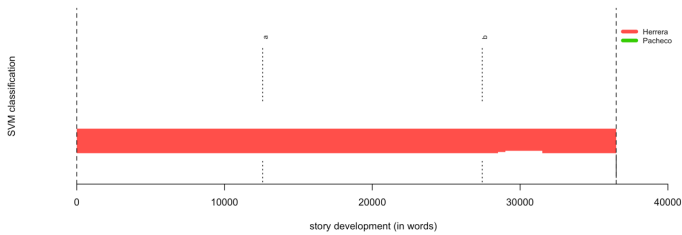


Fig. 3: P2. SVM. Samples of 5000 words. 100 MFW

En la segunda prueba que hemos realizado con SVM hemos utilizado muestras de 3000 palabras con un solapamiento de 2500. Obtenemos las figuras 4 y 5, en las que aparece algo de probabilidad (si bien ciertamente absolutamente mínima) de presencia de mano de Pacheco al comienzo del tercer libro de P.

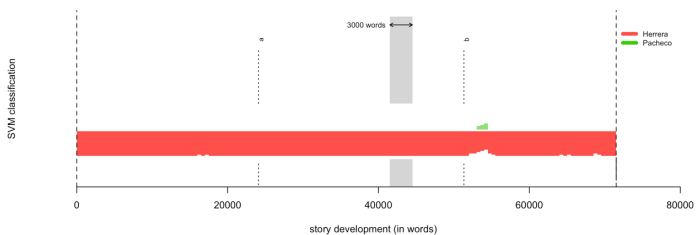


Fig. 4: P. SVM. Samples of 3000 words. 100 MFW

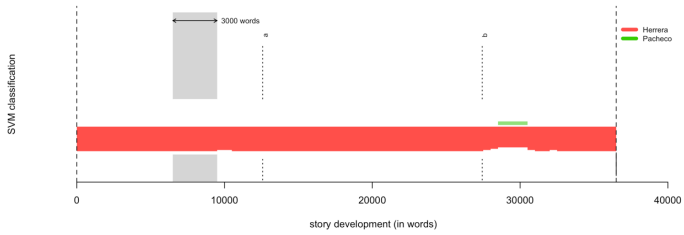


Fig. 5: P2. SVM. Samples of 3000 words. 100 MFW

A continuación, aplicamos NSC con las 50 palabras más frecuentes sobre *P* y *P2* la elección es Herrera para todo el texto en ambos casos, aunque aparece una probabilidad mínima de Pacheco en algunas partes del texto, como puede verse en las figuras 6 y 7.

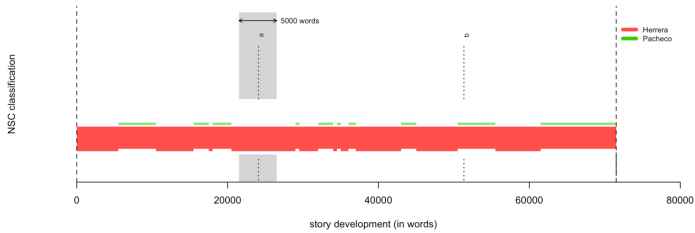


Fig. 6: P. NSC. Samples of 5000 words

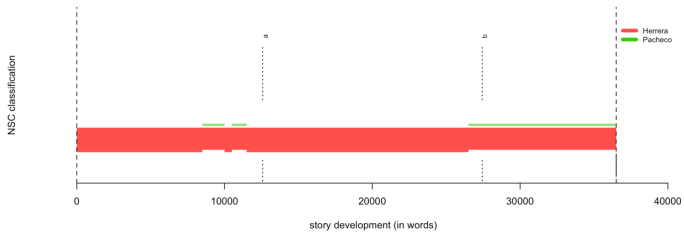


Fig. 7: P2. NSC. Samples of 5000 words

En tercer lugar, aplicamos Rolling Delta con las 500 palabras más frecuentes sobre  $P$  y  $P_2$ . Como puede verse en las figuras 8 y 9, la primera elección es Herrera y la segunda Pacheco.

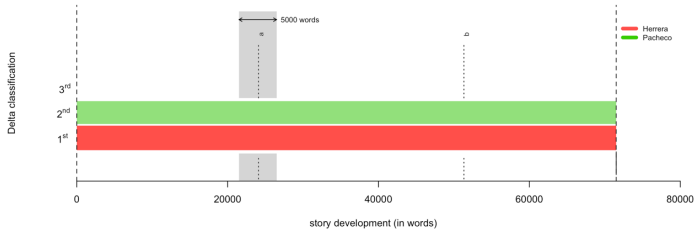


Fig. 8: P.Delta. Samples of 5000 words

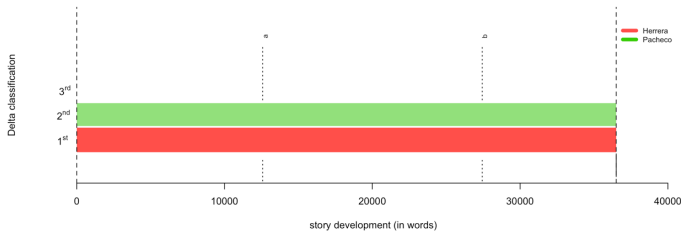


Fig. 9: P2. Samples of 5000 words

Hasta ahora mediante todas las técnicas empleadas el resultado para todo  $P$  y  $P_2$  es Herrera y la probabilidad de presencia de mano de Pacheco parece ser casi inexistente. Para confirmar este resultado y asegurarnos de que las diferencias de tamaño no estuviesen influyendo en el mismo –puesto que el corpus de Herrera es mucho mayor que el de Pacheco, 77044 frente a 5085 palabras–, decidimos realizar un nuevo análisis reduciendo previamente el corpus de referencia de Herrera y  $P$  a 8000 palabras aproximadamente cada uno.

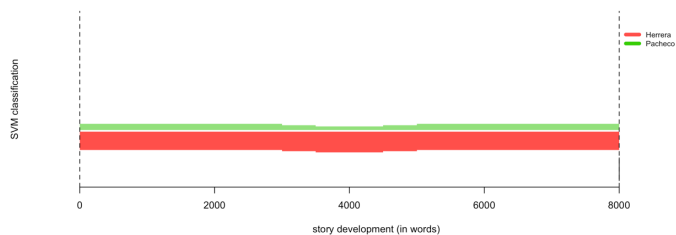


Fig. 10: SVM. Samples of 5000 words. 100 MFW. Aleatory reduced version P

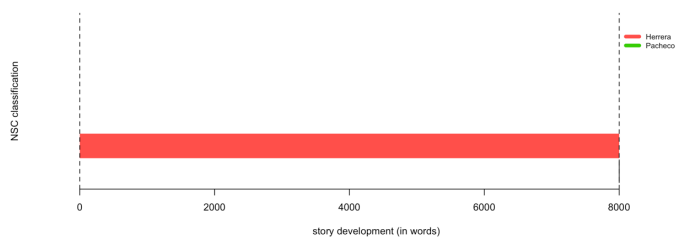


Fig. 11: NSC. Samples of 5000 words. Aleatory reduced version P

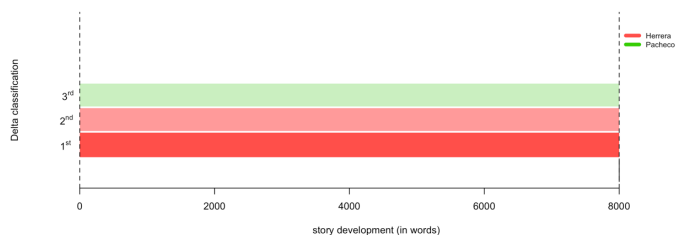


Fig. 12: Delta. Samples of 5000 words. Aleatory reduced version P

Como puede verse en las figuras 10, 11 y 12, aunque la probabilidad de presencia de Pacheco ha aumentado en SVM, en todas las opciones la autoría elegida a lo largo de todo el texto es la de Herrera, no eligiéndose la de Pacheco en ningún caso o fragmento.

## 5. Conclusiones

Con este trabajo se pretende mostrar en un primer nivel la utilidad de las herramientas digitales y, especialmente, de las estilométricas para avanzar en la investigación del drama textual herreriano; y, en un segundo nivel, el interés por supuesto de aplicar la Estilometría a textos en español, como han hecho algunos trabajos pioneros, y de utilizar técnicas como Rolling Stylometry para abordar obras literarias en español con autoría mixta o que pueden tener varias manos, un campo que está todavía casi sin explorar.

Centrándonos en los resultados de nuestro estudio, aparentemente en este primer acercamiento estilométrico parece que la probabilidad de que hubiera mano de Pacheco en *P* sería muy baja, aunque habría que tener en cuenta por un lado, que en el corpus de poesía de Pacheco encontramos temas no presentes en *P* (como la pintura), y, por otro lado, que los resultados podrían variar si obtuviéramos más poemas del autor del *Libro de retratos* para añadir al corpus de referencia. En cualquier caso, con los datos que tenemos a día de hoy el argumento que esgrimen los detractores de la autoría herreriana de que con bastante seguridad la mano del pintor y editor literario sería la responsable de los cambios en *P* en su intento de acercar el estilo herreriano al gusto más barroco de la época pierde fuerza.

Para concluir, nos gustaría señalar que este estudio forma parte de un proyecto mayor en desarrollo y que será necesaria la realización de nuevos análisis y la aplicación de nuevas técnicas estilométricas con un corpus mayor de autores para poder llegar a conclusiones definitivas sobre la autoría herreriana de *P*.



# Góngora hors norme?

## Étude stylométrique d'un motif gongorin

Marie-Églantine Lescasse (Paris)

**RÉSUMÉ :** Cet article a pour objectif de quantifier les spécificités syntaxiques gongorines par rapport à l'usage poétique de son temps, par extraction de séquences grammaticales récurrentes (SGR, ou n-grams de POS-tags) et comparaison avec un corpus-témoin. Après une clarification terminologique autour de la notion de motif, nous avons entamé une caractérisation stylistique de l'œuvre gongorine, grâce à l'observation de la séquence la plus surreprésentée chez notre auteur. L'analyse qualitative nous a conduit à identifier certains effets de sens récurrents qui feraient de cette séquence un *motif* à proprement parler, celui du rapprochement expressif entre entités hétérogènes, et, dans les cas où la séquence intègre un complément antéposé, celui de la recréation poétique du processus d'intellection du monde, allant du particulier au général.

**ZUSAMMENFASSUNG :** Die Besonderheiten der gongorinischen Syntax werden quantitativ im Vergleich zum poetischen Sprachgebrauch seiner Zeit untersucht. Dazu werden wiederkehrende grammatikalische Sequenzen (SGR, n-grams, pos-tags) mit einem ähnlichen zeitgenössischen Korpus verglichen. Nach einer theoretischen Betrachtung des Motivbegriffs, wird versucht, das Werk Góngoras stilistisch zu beschreiben, indem eine bestimmte überrepräsentierte Sequenz genauer betrachtet wird. Die zeitgleiche qualitative Analyse ermöglicht es, bestimmte wiederkehrende Sinneffekte zu erkennen, die diese Sequenz als Motiv gelten lassen. Es geht dabei um ein ausgesprochenes Annähern von heterogenen Entitäten, und um eine poetische Reflexion über die Intellektualisierung der Welt (in den Fällen, in denen die Sequenz ein vorangestelltes Komplement integriert), immer vom Besonderen auf das Allgemeine schließend.

**MOTS CLÉS :** stylométrie; motif; herméneutique numérique; idiolecte; style

**SCHLAGWÖRTER :** Stilometrie; Motiv; numerische Hermeneutik; Idiolekt; Stil

Dans cet article, nous nous proposons de mesurer la différence de la langue de Góngora à la langue poétique de son temps, grâce à une méthode de stylistique computationnelle : l'extraction de séquences grammaticales. Dès la réception des grands poèmes à partir de 1613, la langue gongorine a été per-

---

<sup>o</sup> Je tiens à remercier Mercedes Blanco, Nanette Rissler-Pipka, Bénédicte Pincemin, Corinne Mencé-Caster et Marc Douguet pour leur relecture attentive et leur aide, extrêmement précieuse.

que comme étrange, voire étrangère.<sup>1</sup> Tant dans le choix du lexique, celui des adjectifs, que dans la syntaxe, Góngora s'écartait de l'usage poétique du xvi<sup>e</sup> supposément marqué par la simplicité, la clarté, un ordonnancement « coulant » des mots dans la phrase, et définissant, combiné à d'autres éléments, la *propiedad* castillane.<sup>2</sup>

Face à cette vision d'une langue singulière, on cherchera à quantifier les spécificités syntaxiques gongorines<sup>3</sup> par rapport à l'usage poétique de son temps,<sup>4</sup> à travers l'observation des séquences de trois catégories grammaticales (ou trigrammes de POS-tags<sup>5</sup>), avec le logiciel TXM.<sup>6</sup> L'objectif de cet article sera double : proposer une nouvelle méthode de mesure des écarts stylistiques individuels à la norme linguistique dominante en poésie, et entamer une caractérisation stylistique de la langue gongorine assistée par des outils numériques (TXM et Excel pour les calculs d'écarts).

Dans un premier temps, nous nous livrerons à une clarification terminologique autour du terme « motif », et proposerons l'introduction du terme « Séquence Grammaticale Récurrente » (SGR) pour désigner les n-grams<sup>7</sup> de POS-tags, objets de notre extraction, avant de justifier le choix des séquences

<sup>1</sup> Voir Mercedes Blanco, *Góngora o la invención de una lengua* (León : Universidad de León, 2016), 296.

<sup>2</sup> Voir Marie-Églantine Lescasse, « L'emploi du concept de < propiedad > dans la polémique gongorine », *eSpania* 29 (février 2018), <http://journals.openedition.org/e-spania/27486>.

<sup>3</sup> Le terme « spécificités » désigne un calcul de statistiques textuelles inventé par Pierre Lafon en 1980. Il s'agit d'une mesure du suremploi au sein d'un corpus d'un terme dans un texte par rapport à son emploi dans les autres textes. Voir Pierre Lafon, « Sur la variabilité de la fréquence des formes dans un corpus », *Mots* 1 (1980) : 127–65. La méthode suivie ici, si elle n'utilise pas ce calcul, repose sur le même principe : mesurer la surreprésentation statistique de certains traits (ici syntaxiques) par la comparaison de la répartition de leurs fréquences entre différents textes d'un corpus.

<sup>4</sup> Nous parlons ici d'usage, mais nous justifierons plus bas l'emploi de la notion linguistique de « norme ».

<sup>5</sup> On appelle POS-tags les étiquettes « morpho-syntaxiques » attribuées à chaque forme par un logiciel étiqueteur comme Treetagger. L'acronyme POS signifie *part of speech* ou *partie de discours*, autrement dit, les catégories grammaticales.

<sup>6</sup> TXM est un logiciel de textométrie, développé à l'ENS de Lyon par l'équipe du Projet Textométrie : Bénédicte Pincemin, Serge Heiden, Matthieu Decorde, Sébastien Jacquot et Alexei Lavrentiev. Voir le site du projet : <http://textometrie.ens-lyon.fr/?lang=fr>

<sup>7</sup> Un n-gram désigne une suite, une séquence, avec *n* désignant la longueur de la suite, qui peut être une suite de lettres (l'extraction de tous les bigrammes du mot « avion » donnera : av vi io on) ou une suite de mots (dans « le gâteau est bon », les bigrammes sont « le gâteau », « gâteau est », « est bon »). Sur les n-grams, voir Efstathios Stamatatos, « A survey of modern authorship attribution methods », *Journal of the American Society for information science and technology* 60, no. 3 (2009) : 538–56, ici 541–2.

de longueur trois, et de présenter la méthode suivie. Dans un second temps, nous proposerons un parcours interprétatif autour de la séquence la plus surreprésentée chez Góngora, NC ADJ NC<sup>8</sup>, motif du rapprochement expressif entre entités hétérogènes et de la prépondérance de la perception sensorielle synthétique sur l'intellection analytique. Au sein de ce parcours, nous étudierons d'abord les effets de sens associés aux trigrammes à proprement parler (rapprochement expressif), puis nous proposerons une interprétation des nombreux cas où ceux-ci intègrent des compléments antéposés (prépondérance du sensoriel).

D'après les linguistes ayant travaillé sur la norme, celle-ci reçoit deux définitions bien distinctes.<sup>9</sup> On entendra ici le mot norme au sens quantitatif de moyenne, état régulier ou habituel, représentable graphiquement sous la forme d'une gaussienne, et non au sens qualitatif d'un état souhaitable, d'un modèle prescriptif auquel il faudrait se conformer.

Ce que nous appelons « norme poétique » s'entendra ici dans ce premier sens statistique, dénué de toute connotation axiologique : il s'agit de la moyenne d'emploi de certaines formes – on verra lesquelles – à une époque donnée.<sup>10</sup>

## Vers une quantification des spécificités gongorines

### Pour une clarification terminologique : des motifs aux SCR

La stylistique computationnelle s'intéresse depuis quelques années à ce qu'elle appelle des « motifs », n-grams de mots ou de parties du discours (POS), issus du processus TAL de *tagging* (étiquetage morpho-syntaxique).<sup>11</sup>

<sup>8</sup> NC désigne un nom commun et ADJ un adjectif selon les normes de Treetagger.

<sup>9</sup> Françoise Gadet distingue ainsi la « norme objective » de la « norme subjective », ce que nous avons appelé ici « norme quantitative » et « norme qualitative » : « Il faut distinguer entre *norme objective, observable*, et norme subjective, système de valeurs historiquement situé. Dans le premier sens, *lié à normal*, il renvoie à l'idée de fréquence ou tendance, et il peut être utilisé au pluriel, au contraire du second sens, reflété par normatif ou normé, en conformité à l'usage valorisé (la Norme, qui a pu être dite fictive) », Françoise Gadet, *La variation sociale en français* (Paris : Ophrys, 2003), 19. Alain Rey ne faisait pas autre chose en distinguant un sens descriptif et un sens axiologique du mot « norme » dans son article fondamental sur la norme linguistique. Voir Alain Rey, « Usages, jugements et prescriptions linguistiques », *Langue française* 16 (1972) : 4–28, ici 5.

<sup>10</sup> La question de l'homogénéité du corpus est alors centrale, tant du point de vue diachronique que diatopique, du genre textuel, de la tonalité, voire des thèmes traités.

<sup>11</sup> Voir les travaux de Frédérique Sitri, Agnès Tutin, Jean-Marc Leblanc, Olivier Kraif, Dominique Legallois et Thierry Poibeau, réunis dans le numéro 53 de la revue *Lidil* ; ainsi que les tra-

Or, la notion de motif, ambiguë, demande une clarification. Si le motif désigne en littérature un élément narratif concret, précis,<sup>12</sup> à distinguer du thème, sens général abstrait à partir des motifs concrets d'une œuvre,<sup>13</sup> elle prend un sens différent en stylistique computationnelle. Dans ce domaine de recherche, elle est parfois employée, d'une manière qui pourrait entraîner confusion, comme synonyme de « séquence grammaticale récurrente », ou suite de POS-tags.<sup>14</sup> Or, le terme « motif » a un sens précis en humanités numériques, qui n'est pas celui d'une simple suite d'étiquettes grammaticales. Selon la définition de Dominique Legallois, il y a motif lorsqu'un schéma syntaxique récurrent est associé à un effet de sens, une fonction (pragmatique, rhétorique) récurrente :

Quelle que soit la méthode, sont qualifiés de *motifs* les seuls segments présentant une régularité d'ordre lexical [...], et/ou fonctionnel : un motif possède une fonction sémantique, pragmatique, rhétorique discernable, voire une fonction d'organisation des plans textuels [...].<sup>15</sup>

Pour éviter ces ambiguïtés du terme motif, nous proposons donc le terme « Séquence Grammaticale Récurrente » ou SGR, pour désigner les séquences ou n-grams d'étiquettes grammaticales (POS-tags) issus de l'extraction auto-

---

vaux de l'équipe ACASA au Lip6 (Laboratoire d'informatique de l'Université Paris-Sorbonne) : Jean-Gabriel Ganascia, « Extraction automatique de motifs syntaxiques », TALN 2001, Tours ; Mohamed-Amine Boukhaled et al., « Une mesure d'intérêt à base de surreprésentation pour l'extraction des motifs syntaxiques stylistiques », *Actes de la 22<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles* (juin 2015). L'importance des « motifs » pour la caractérisation stylistique est attestée par le grand nombre de travaux en Attribution d'Autorité s'appuyant sur des n-grams de POS-tags comme marqueurs stylistiques aptes à discriminer entre des auteurs différents. Voir par exemple les travaux d'Argamon, Koppel et Avneri, Diederich et Paass, Michael Gamon, Koppel et Schler, Kukushkina, Polikarpov et Khmelev, Zhao et Zobel.

<sup>12</sup> Une « unité sémantique qui revient, avec des variations, d'un texte à l'autre [...] (le fil d'Ariane, le philtre d'amour [...] ou la voile noire », selon la définition qu'en donne Hendrik van Gorp et al., *Dictionnaire des termes littéraires* (Paris : Honoré Champion, 2005), 313.

<sup>13</sup> « Quand, par exemple, un récit contient comme motifs des péripéties matrimoniales (séparation, nouvelle relation, etc), on peut en déduire le thème du « caractère éphémère » ou de « l'impossibilité de relations amoureuses durables », même si cela n'est pas dit tel quel dans le texte », van Gorp, *Dictionnaire*, 478.

<sup>14</sup> C'est le cas dans l'article de Ganascia, et dans celui de Boukhaled et al., tous deux cités plus haut. Il ne s'agit pas de contester la méthode ni les résultats, tout à fait probants, mais l'emploi du terme « motif », qui gagnerait à notre avis à être remplacé par « séquence » pour plus de clarté, et pour revenir à une définition plus holiste du motif (séquence assortie à un effet de sens récurrent).

<sup>15</sup> Dominique Legallois, Thierry Charnois et Thierry Poibeau, « Repérer les clichés dans les romans sentimentaux grâce à la méthode des < motifs > », *Lidil* 53 (2016) : par. 18.

matique, et non encore analysées et interprétées par le chercheur. On parlera donc d'extraction de « séquences grammaticales », et non de « motifs » : le motif n'est pas donné a priori mais est le résultat d'une construction interprétative, a posteriori, de l'analyste ; après la phase d'extraction, il n'y a donc pas encore de « motifs ».

### Méthode suivie et un premier aperçu des résultats

Il s'agit à présent d'exposer la méthode suivie, avant de proposer une lecture de la séquence gongorine la plus surreprésentée par rapport au corpus-témoin.

1. Constitution d'un corpus-témoin ou étalon, de 10 poètes des XVI<sup>e</sup> et XVII<sup>e</sup> siècles, afin d'avoir un corpus de comparaison ; constitution du corpus de l'œuvre poétique gongorine.<sup>16</sup> Les dix poètes de l'étalon sont : Juan Boscán, Fernando de Herrera, Lope de Vega, les frères Argensola (Bartolomé Leonardo et Lupercio), Aldana, Timoneda, Tirso de Molina, Quevedo, fray Luis de León.
2. Import du corpus sur TXM.
3. Création d'une partition par auteurs.
4. Extraction des SGR (séquences de POS-tags) de longueur 3 et de leurs fréquences par requête CQL<sup>17</sup> sur la propriété « partie de discours ». Tests sur des séquences plus courtes (bigrammes) et plus longues (des 4-grams aux 8-grams).<sup>18</sup>
5. Export de la table de donnée au format .csv.
6. Sur Excel, élimination manuelle des lignes contenant des étiquettes de ponctuation, celle-ci n'étant pas encore fixée à l'époque classique.
7. Normalisation des fréquences de chaque SGR par rapport à la taille de chaque texte, afin de travailler sur des fréquences relatives.
8. Tracé des courbes d'emploi des SGR ou spectres fréquentiels chez Góngora et l'ensemble de l'étalon.<sup>19</sup> Cette dernière courbe correspond à la courbe

<sup>16</sup> Nous avons pris les poèmes du corpus-témoin sur le Cervantes Virtual, et les poèmes de Góngora sur le site de l'OBVIL, où ceux-ci ont reçu une édition revue par Antonio Carreira : [http://obvil.paris-sorbonne.fr/corpus/Góngora/Góngora\\_obra-poetica](http://obvil.paris-sorbonne.fr/corpus/Góngora/Góngora_obra-poetica).

<sup>17</sup> CQL (Corpus Query Language) est le langage de requêtes associé à TXM.

<sup>18</sup> Voir notre paragraphe sur le choix des séquences de 3.

<sup>19</sup> Nous appelons étalon notre corpus-témoin (ou corpus de comparaison).

moyenne d'emploi de SGR de 3 en poésie aux XVI<sup>e</sup> et XVII<sup>e</sup> siècles.<sup>20</sup> Les fréquences d'emploi des SGR y ont été moyennées à partir de leur fréquence chez chacun des dix auteurs de l'étalon.

9. On a alors voulu travailler sur les pics, c'est-à-dire les points où visiblement Góngora s'écartait de la moyenne de l'époque. Pour cela, on a établi une *méthode simple et reproductible de mesure de la significativité des pics* (ou écarts), dont nous proposons l'application systématique en stylistique computationnelle. On a créé une mesure-filtre, permettant d'estimer la fluctuation moyenne de chaque individu de l'étalon par rapport à la moyenne de l'étalon : l'erreur-type, soit l'écart-type divisé par la racine de l'effectif.

$$SE = \frac{8}{\sqrt{n}}$$

$S$  est l'écart-type,  $n$  est l'effectif, c'est-à-dire la taille de l'échantillon. Il s'agit d'une mesure de la dispersion de l'échantillon, visualisable graphiquement sous la forme de barre d'erreurs superposées aux spectres fréquentiels (Fig. 1<sup>21</sup>).

Lorsqu'une fréquence d'emploi chez Góngora sort de l'espace compris entre les barres d'erreur, cela signifie que l'écart de Góngora à l'étalon est significatif du point de vue statistique. Cette mesure vise à pallier les illusions d'optique naissant de la simple observation des décrochages dans la courbe, en écartant les pics dont la variabilité fait partie de la variabilité acceptable, attendue d'un individu par rapport à l'ensemble.<sup>22</sup>

10. Calcul de l'écart relatif entre Góngora et l'étalon pour chaque séquence :

$$\text{Écart relatif} = \frac{\text{fréquence d'une SGR chez Góngora} - \text{sa fréquence dans l'étalon}}{\text{fréquence dans l'étalon}}$$

Visualisation sous forme graphique des écarts relatifs à l'étalon (Fig. 2).

La ligne verticale partant de 0 correspond à l'emploi moyen d'une SGR dans l'étalon. Les saillies à droite représentent des suremplois, chez Garcilaso (en orange) et Góngora (en bleu), à gauche des sous-emplois par rapport à la

<sup>20</sup> On donnera ici directement les courbes superposées assorties des barres d'erreur (Fig. 1, point 9 de la méthode).

<sup>21</sup> Sur ce graphique figure la courbe d'emploi des SGR chez Garcilaso de la Vega, une première phase de notre étude portant sur une étude contrastive des SGR chez Góngora et Garcilaso.

<sup>22</sup> Dont la variabilité est comprise dans la variabilité moyenne des individus par rapport à l'ensemble de l'étalon pour une fréquence donnée.

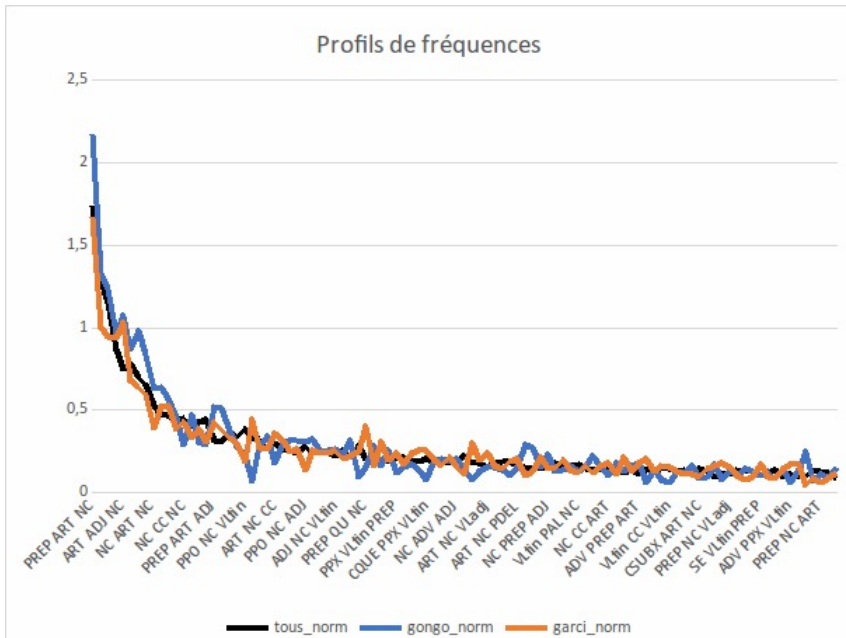


Fig. 1 : Profils de fréquences comparés Góngora/Garcilaso/étalon avec barres d'erreur

moyenne du corpus-témoin. Les chiffres donnés en abscisse correspondent au pourcentage de suremploi ou de sous-emploi de la séquence chez l'un et l'autre auteur.<sup>23</sup>

### Le choix d'extraire les SGR de longueur 3

Un premier regard sur les résultats de l'extraction automatique laisse perplexe. Les séquences de 3 paraissent non pertinentes du point de vue linguistique : ce sont des séquences hétérogènes, souvent à cheval sur deux groupes grammaticaux aux fonctions distinctes. Voici, à titre d'exemple, quelques-unes des premières instances, par ordre d'apparition chronologique dans le corpus gongorin, de la séquence NC ADJ NC extraites par TXM du corpus : *escollo duro troncos*, *estrella lucientes ojos*, *quejas purpúreos troncos*, *rayo fulminante trompa*, *aguas torpe marinero*. Par exemple, « *aguas torpe marinero* » vient de

<sup>23</sup> Par exemple, si une barre atteint 50 %, cela signifie que la séquence associée en légende est employée 50 % de fois plus chez l'auteur concerné que dans la moyenne de l'étalon.

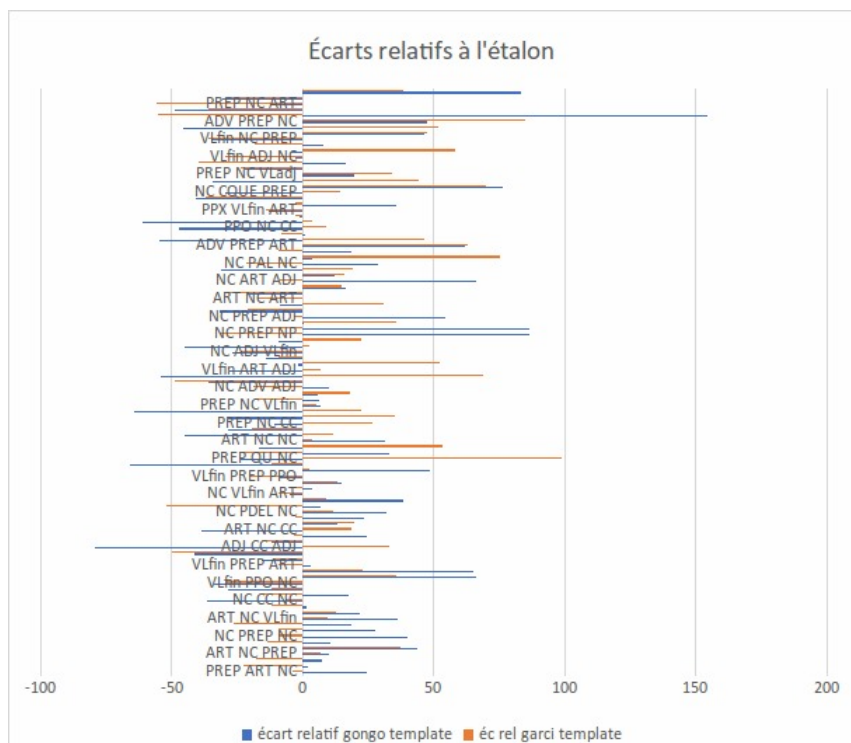


Fig. 2 : Écarts relatifs de Góngora à la moyenne de l'étalon pour chaque SGR

« de las [...] | estigias [aguas torpe marinero] », <sup>24</sup> et mêle donc dans une même séquence un morceau de complément du nom (*aguas*) et un groupe épithète-nom (*torpe marinero*).

Pourquoi alors avoir choisi d'extraire des SGR de longueur 3 ?

Les séquences de 2, trop courtes, n'apportent que peu d'informations syntaxiques, et ne sont pas directement interprétables en termes d'esthétique. <sup>25</sup> Les séquences de 4, trop longues, sont moins intéressantes, paradoxalement parce qu'elles sont plus cohérentes et homogènes : elles ne donnent pas accès

<sup>24</sup> Nous encadrerons systématiquement les SGR extraites par des crochets, pour mieux les identifier, et le passage d'un vers à l'autre par une barre oblique.

<sup>25</sup> Voici les premières instances extraites par TXM pour les bigrammes de POS-tags les plus fréquents chez Góngora, selon leur ordre d'apparition dans le corpus : ART NC (*la alba, la zampoña, los muros*), ADJ NC (*bucólica Talía, excelso conde, purpúreas horas*), PREP ART (*en las, de la, en la*), NC PREP (*son de, freno de, lebrél en, cordón de*).



à des groupes composites et donc à la syntagmatique de l'auteur.<sup>26</sup> Les séquences plus longues sont de plus en plus rares à mesure que le  $n$  du  $n$ -gram augmente. Dans les faits, elles ne moissonnent que des refrains, répétés à l'identique, et donc dépourvus d'intérêt pour notre étude.<sup>27</sup>

Les séquences de 3, en plus d'être suffisamment longues pour avoir un contenu lexical, plus facilement interprétable du point de vue de l'esthétique, nous intéressent précisément en raison de leur hybridité et de leur caractère composite. Leur hétérogénéité, qui témoigne d'un positionnement à cheval entre deux groupes linguistiques distincts, permet de détecter la surreprésentation du trigramme NC ADJ NC, qui, précisément parce qu'il ne fait pas sens au plan linguistique, est le meilleur observatoire d'une syntaxe peu orthodoxe, invitant à examiner la manière dont se suivent et s'articulent chronologiquement les groupes grammaticaux. Nous y reviendrons lorsque nous aborderons la question de la chronosyntaxe. Elles fonctionnent en ce sens comme signes, traces microtextuelles de phénomènes syntaxiques visibles à une échelle plus grande,<sup>28</sup> qui constitueront notre objet d'étude à proprement parler.

En elles-mêmes, elles indiquent déjà la tendance gongorine à la suppression de l'article et à la condensation verbale et conceptuelle.

### Étude de la séquence la plus spécifique de Góngora : NC ADJ NC

#### De l'étude des trigrammes proprement dits...

Nous proposons à présent un parcours interprétatif à travers les emplois du trigramme le plus surreprésenté chez Góngora : NC ADJ NC (154 % de fois plus que la moyenne de l'étalon! (Fig. 3)

<sup>26</sup> Par exemple, les instances de la SGR la plus fréquente sont : *son de la zampoña, lebril en el cordón, honor de la montaña, reino de la espuma*. Ils correspondent donc souvent à des ensembles grammaticaux homogènes, inintéressants pour étudier des phénomènes d'ordre syntaxique.

<sup>27</sup> Voici un exemple d'instance répétée obtenue par une requête sur les 8-grams de POS : « *degraciada, ¿ qué quiere qu'haga?* », répétée sept fois dans le poème 55 de l'édition d'Antonio Carreira.

<sup>28</sup> C'est aussi ce qu'avance Argamon dans son étude en Attribution d'Autorité visant à établir une classification stylistique automatique de journaux américains ; à la recherche d'une « méthode pour capturer des structures syntaxiques sous-jacentes » (je traduis), il évoque brièvement les avantages des trigrammes : « Trigrams are large enough to encode useful syntactic information, while small enough to be computationally manageable », Shlomo Argamon-Engelson, Moshe Koppel et Galit Avneri, « Style-based text categorization : what newspaper am I reading? », *AAAI Workshop on Learning for Text Categorization* (1998) : 1–42, ici 2.

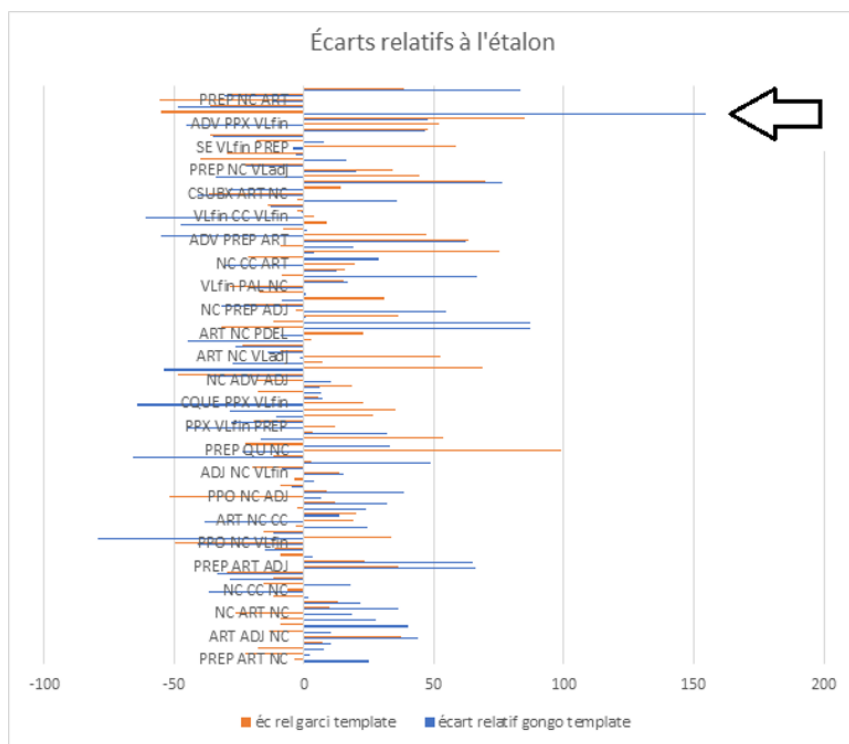


Fig. 3 : La séquence la plus surreprésentée chez Cóngora : NC ADJ NC, employée 154 % de fois plus que dans l'étalon

Nous avons mené une étude stylistique sur les 30 instances correctement étiquetées de cette séquence (sur un total de 47 instances moissonnées). Nous les avons ramenées à l'échelle plus large du groupe syntaxique dont le trigramme est le signe, et les avons classées selon quatre critères, de l'observation syntaxique à la sémantique :<sup>29</sup> la fonction grammaticale occupée par le groupe syntaxique auquel appartient le motif ; la présence ou l'absence de métaphore ; l'antéposition ou la postposition de l'adjectif par rapport au nom ; la fonction poétique ou l'« effet de sens » produit par la séquence.

Voici le résultat de cette classification : Le trigramme intègre 10 fois un complément circonstanciel de lieu, une fois un cc<sup>30</sup> de temps et une fois un

<sup>29</sup> Le choix de ces critères est né de l'étude (syntaxique et stylistique) des instances ; nous avons retenu les critères qui nous semblaient les plus pertinents pour leur analyse.

<sup>30</sup> Complément circonstanciel.

CC de cause, ce qui totalise 12 instances intégrant un CC. Il intègre 9 fois un complément du nom. Si l'on totalise CC et complément du nom, il appartient donc 21 fois à un complément. Il représente quatre fois le rapprochement d'un sujet et de son COD, une fois d'un sujet et de son épithète, deux fois d'un sujet et son attribut; il correspond une fois à un accusatif grec. Dans approximativement 38 % des cas, le trigramme se trouve donc à l'intérieur d'un complément (un CC ou un complément du nom). Il faudrait vérifier si les compléments sont effectivement suremployés par rapport aux autres auteurs, ce qui n'est pas notre propos ici.

#### LA FONCTION GRAMMATICALE OCCUPÉE PAR LE GROUPE SYNTAXIQUE AUQUEL APPARTIENT LE MOTIF

*Complément du nom* de las [...] estigias [aguas torpe marinero], de su [edad corta historias] largas, de zagalejas [cándidas voz tierna], de la amorosa [espuela duro freno], de verde [laurel caduca rama], de [estrellas inmortal corona], de este [escollo duro troncos], ve de [aromas sacros troncos], de [halcón blancas alas]

*Complément circonstanciel de lieu* en su volante [carro blancas aves], en la lengua del [agua ruda escuela], en sus [ramas inconstantes nidos], entre el néctar de [Amor mortal veneno], del siempre en la [montaña opuesto pino], en sus [riberas pacíficas banderas], sobre dos [hombros larga vara], del [pecho ardiente regüeldos], en cuyo vasto [seno duras cortezas], en mis [entrañas dulces lazos]

*Complément circonstanciel de cause* por su [edad majestüosa cano]

*Complément circonstanciel de temps* en tu [edad dorada oro]

*Sujet – COD* el [mercader nuevos soles], guardó el [heno dulces pomos], una [dama esquiva lengua], dulce [instrumento bellos efectos]

*Sujet – épithète* sus [piedras rubios metales]

*Sujet – attribut/attribut-sujet* [ejercicio competente oficina], sean tu [trofeo locas esperanzas]

*Accusatif grec* el [vulto feroz hombre]

#### LA PRÉSENCE OU L'ABSENCE DE MÉTAPHORE

*Avec métaphore* [rayo fulminante trompa], de las estigias [aguas torpe marinero], entre el néctar de [Amor mortal veneno], de la amorosa [espuela duro freno], de [estrellas inmortal corona], en tu [edad dorada oro], de [halcón

blancas alas], sean tu [trofeo locas esperanzas], en mis [entrañas dulces lazos]

*Sans métaphore* de su edad [corta historias largas], de zagalejas [cándidas voz tierna], en su volante [carro blancas aves], en sus [ramas inconstantes nidos], el [vulto feroz hombre], por su [edad majestüosa cano], de verde [laurel caduca rama], más que la [nieve blanco toro], sus [piedras rubios metales], [escollo duro troncos], del siempre en la [montaña opuesto pino], [riberas pacíficas banderas], sobre dos [hombros larga vara], guardó el [heno dulces pomos], en cuyo vasto [seno duras cortezas], ve de [aromas sacros troncos], [ejercicio competente oficina], dulce [instrumento bellos efectos], más que la [nieve blanco toro],

#### L'ANTÉPOSITION OU LA POSTPOSITION DE L'ADJECTIF PAR RAPPORT AU NOM

*adjectif – nom* de las estigias [aguas torpe marinero], en su volante [carro blancas aves], de la amorosa [espuela duro freno], de verde [laurel caduca rama], del siempre en la [montaña opuesto pino], [riberas pacíficas banderas], sobre dos [hombros larga vara], guardó el [heno dulces pomos], ve de [aromas sacros troncos], [ejercicio competente oficina], [dama esquiva lengua], en cuyo vasto [seno duras cortezas], dulce [instrumento bellos efectos], sean tu [trofeo locas esperanzas], en mis [entrañas dulces lazos], más que la [nieve blanco toro], de [halcón blancas alas],

*Nom – adjectif* de su edad [corta historias largas], de zagalejas [cándidas voz tierna], de [estrellas inmortal corona], más que la [nieve blanco toro], sus [piedras rubios metales], [escollo duro troncos], en tu [edad dorada oro], del [pecho ardiente regüeldos]

#### LA FONCTION POÉTIQUE

*Rapprochement paradoxal* de su edad [corta historias largas], del [agua ruda escuela], de [Amor mortal veneno], de la amorosa [espuela duro freno], de [halcón blancas alas],

*Rapprochement expressif* el [mercader nuevos soles], en tu [edad dorada oro], de este [escollo duro troncos], dulce [instrumento bellos efectos], en mis [entrañas dulces lazos], del [pecho ardiente regüeldos], sean tu [trofeo locas esperanzas]

*Ligne anticipatrice, suspens et tension/déplacement du centre de gravité sémantique* de zagalejas [cándidas voz tierna]; de estigias [aguas torpe marinero], [es-

collo duro troncos], [riberas pacíficas banderas], guardó el [heno dulces pomos], ve de [aromas sacros troncos],

*Primauté des circonstances, de l'adventice sur l'essentiel* en sus [ramas inconstantes nidos], por su [edad majestüosa cano], del siempre en la [montaña opuesto pino], sobre dos [hombros larga vara], en cuyo vasto [seno duras corpezas],

*Primauté des impressions sur l'intellection : fonction sensorielle/synesthésique* en su volante [carro blancas aves]; de [estrellas inmortal corona], más que la [nieve blanco toro].

*Cas plus rares (accusatif grec, apposition métaphorique/descriptive)* el [vulto ferroz hombre], [rayo fulminante trompa], sus [piedras rubios metales].

Même s'il est difficile, comme on va le voir, de donner une interprétation unifiante des instances, et donc d'identifier cette séquence comme motif à coup sûr, on verra que certains effets de sens (les cinq premiers de la liste) sont néanmoins récurrents pour cette séquence.

Voici à présent un essai d'interprétation des résultats, selon les effets de sens que prend la séquence.

### Rapprochements expressifs

La séquence NC ADJ NC est souvent le lieu de rapprochements expressifs entre entités contradictoires ou hétérogènes. Les exemples sont légion : sujet/attribut métaphorique dans « las estrellas nocturnas luminarias »<sup>31</sup> dans un jeu de rayonnement du vers, enchâssé entre deux lexèmes de la lumière, du naturel à l'artificiel, et dans le jeu surajouté de l'oxymore des luminaires nocturnes. Dans *aguas torpe marinero*,<sup>32</sup> la condensation de l'expression en « eaux marin maladroite » installe un sentiment d'inquiétude, présage de

<sup>31</sup> « Aquellas que los árboles apenas dejan ser torres hoy, dijo el cabrero con muestras de dolor extraordinarias, las *estrellas nocturnas luminarias* eran de sus almenas, cuando el que ves sayal fue limpio acero. » (*Soledad Primera*, v. 212–7.)

Le passage se situe au début de la première *Soledad* : le pèlerin, après avoir été accueilli par les chevriers, part en compagnie de l'un d'eux ; sur un promontoire, ils observent ensemble le paysage ; c'est alors que le chevrier décrit avec douleur le passage du temps sur les ruines qui se présentent à leurs yeux, autrefois des tours si hautes que les étoiles mêmes semblaient des lampions, accrochés à leurs créneaux.

<sup>32</sup> « Tú, Codicia, tú, pues, de las profundas

danger, renforcé par la métaphore macabre d'une mer où s'ouvrent partout des sépulcres et s'entassent des ossements. Le passage, qui se situe après la battue et le sommeil enlacé des jeunes villageois dans la première *Soledad*, est celui du monologue du vieillard contre la navigation. Góngora assimile ici la Convoitise, allégorisée, à Charon, psychopompe et nocher des eaux du Styx, *torpe marinero*, c'est-à-dire marin affreux, laid, abominable,<sup>33</sup> suivant en cela la prosopographie virgilienne du conducteur des morts au livre VI de l'*Énéide*.

Au troisième jour du récit, le vieil homme, contempteur du progrès, présente le pèlerin au futur époux et au père de la mariée ; les charmes de la jeune paysanne lui rappellent la beauté de celle qu'il aime, avant que la musique enjouée d'une flûte de Pan ne vienne le sortir de sa mélancolie. On trouve dans ces vers le trigramme *memoria/negras plumas* :

Este pues Sol que a olvido le condena,  
cenizas hizo las que su *memoria*  
*negras plumas* vistió, que infelizmente  
sordo engendran gusano, cuyo diente,  
minador antes lento de su gloria,  
inmortal arador fue de su pena,  
y en la sombra no más de la azucena  
[...] víbora pisa tal el pensamiento...

(*Soledad Primera*, v. 737–47)

Le passage est construit sur l'intrication de trois séries allusives : celle au mythe d'Icare, celle au mythe du Phénix, celle autour du ver du Désespoir<sup>34</sup>. La première série représente la femme aimée en Soleil inaccessible, où se brûlent les ailes de l'amoureux (*cenizas hizo las plumas*). À cette allusion mythologique se superpose l'effet joyeux de l'oubli de ses tristesses, métaphoriques par de noires plumes, à la vue de cette beauté qui lui rappelle sa dame. Le Phénix, qui devrait être une image d'espoir, se trouve ici présenté dans sa genèse répugnante, sous forme de ver, n'engendrant ni magnificence ni immortalité, mais désespoir. La signification des deux mythes est mêlée (chute

---

estigias *aguas torpe marinero*,  
cuantos abre sepulcros el mar fiero  
a tus huesos desdeñas ». (*Soledad Primera*, v. 443–6.)

<sup>33</sup> Le commentateur Díaz de Rivas, cité par Robert Jammes dans son édition des *Soledades*, indique : « *Torpe es aquí lo mismo que feo o abominable* ». Luis de Góngora, *Soledades* (Madrid : Castalia, 1994), Robert Jammes éd., 288.

<sup>34</sup> Nous suivons ici l'interprétation de Díaz de Rivas, citée par Robert Jammes. Góngora, *Soledades*, 344–6.

et espoir), et leur présentation incarnée dans quelques éléments concrets et frappants (le soleil, les plumes et les cendres ; les cendres et le ver). Ils laissent place à l'irruption du ver sourd du Désespoir, dont la dent ronge le pèlerin et creuse sur ses joues de vastes sillons<sup>35</sup> (*minador antes lento de su gloria, | inmortal arador fue de su pena*). Góngora nous livre ici l'image troublante d'un mal insidieux, lent et aveugle, et pourtant tout-puissant : l'animation progressive de cet être, qui passe de patient (*gusano est COD*) à agent (*minador, arador* sont sujets), est soulignée par le double substantif d'action en -or, et par le passage de l'adjectif *lento*, rejeté dans le passé et donc comme atténué (via l'adverbe temporel *antes*), à l'assertion emphatique d'une immortalité néfaste (dans l'antithèse *antes lento/inmortal*), avec la position première d'*inmortal*, et le double anapeste (*inmortal arador*), mimétique de ce travail de sape progressive de la mélancolie.<sup>36</sup>

Or, pour revenir au trigramme, la séparation de l'article et du substantif (*las [...] | negras plumas*) modèle une prosodie expressive, dans l'ouverture emphatique du troisième vers sur le rythme pesant d'un double trochée (*negras plumas*). Noir se trouve sous l'accent et en tête de vers, pôle d'attraction sombre de la période, chappe de plomb pesant sur toute la séquence, effet renforcé par l'homéotéleute *negras plumas*, dans un vers expressif de la lourdeur incommensurable du désespoir.

La séquence est très souvent employée à cheval sur deux vers, grâce à un rejet : ceux-ci ont pour fonction d'amener une ambiguïté volontaire, ou au contraire de résoudre l'ambivalence syntaxique. Dans le premier cas, il s'agit de *con arte/culto principio*<sup>37</sup> où l'enchâssement de l'adjectif permet l'ambiguïté

<sup>35</sup> Nous suivons ici l'interprétation d'Antonio Carreira, citée par Robert Jammes : « *El gusano, antes minador de la alegría que produjo el recuerdo de la amada, fue luego arador que abrió en su rostro los surcos de la pena* », in Góngora, *Soledades*, 346, note 742.

<sup>36</sup> Le magnifique article de Nadine Ly sur ce passage montre de manière brillante la complexité de ces vers et l'ambiguïté qu'ils construisent, à travers une double identification du référent de *Sol* (la femme aimée, la paysanne momentanément promue au rang de beauté solaire) et le redoublement de la peine causé par les deux naissances successives du Phénix, une fois (rejetée dans le passé) comme éliminateur de la joie du pèlerin (*minador antes de su gloria*), une seconde fois, comme agent actif de sa douleur (*inmortal arador de su pena*). Nadine Ly, « Poétique et signifiant linguistique. À propos d'un fragment des « Soledades » », *Bulletin Hispanique* 87, no. 3-4 (1985) : 447-70.

<sup>37</sup> « Bajaba entre sí el joven admirando armado a Pan, o semicapro a Marte, en el pastor mentidos, que con arte culto principio dio al discurso, [...] » (*Soledad Primera*, v. 233-6.)

syntaxique : l'épithète *culto* se rattache-t-elle à *arte* ou à *principio*? Sans doute à *principio*, l'origine de l'anomalie étant ici l'ordre verbe-complément.

Au début du *Polyphème*, après la description de l'abondance fructifère de la Sicile polyphémique et des effets dévastateurs de la musique du géant, intervient une description de la beauté de Galatée :

Son una y otra luminosa *estrella*  
*lucientes ojos* de su blanca pluma ;  
 si roca de cristal no es de Neptuno,  
 pavón de Venus es, cisne de Juno. (*Polifemo*, v. 101-4.)

Dans ce passage, le jeu syntaxique s'étend sur les deux premiers vers. Le sujet et l'attribut semblent interchangeables (*estrellas-ojos*), jeu compliqué par la syllepse de *ojos* (yeux et ocelles), et par la variation entre pluriel (*ojos*) et singulier distributif (*una y otra estrella*). La plasticité du vers, qui semble réversible, en fait une sorte de palindrome syntaxique, en conformité avec l'essence ambivalente de Galatée : l'indétermination de l'être, la labilité des attributs se retrouve dans l'inversion paradoxale des compléments du nom cygne de Junon/paon de Vénus.

Le rejet peut aussi être résolutif en ce qu'il tranche une ambiguïté préalable : dans *sueño/política alameda*,<sup>38</sup> le sujet est rejeté sur un seul vers, heptamètre, et mis en valeur par l'épithète incongrue *política*, signifiant « sociable », et par dérivation, « resserré », en parlant des ormes.<sup>39</sup> *Política* implique aussi

---

Le passage se situe juste avant le blâme de la navigation déjà évoqué. Dissimulé dans le tronc creux d'un chêne, le pèlerin regarde l'assemblée des paysans et paysannes.

<sup>38</sup> « Cabo me han hecho, hijo,  
 deste hermoso tercio de serranas ;  
 si tu neutralidad sufre consejo,  
 y no te fuerza obligación precisa,  
 la piedad que en mi alma ya te hospeda  
 hoy te convida al que nos guarda *sueño*  
*política alameda*,  
 verde muro de aquel lugar pequeño  
 que, a pesar de esos fresnos, se divisa » (*Soledad Primera*, v. 516-24.)

Le passage se situe juste après le discours sur la navigation : le pèlerin vient d'être invité par le vieillard à se joindre au groupe de jeunes paysans qui se rendent au village pour les noces à venir. Apercevant le hameau de loin (*que, a pesar de esos fresnos, se divisa*), le vieil homme invite le protagoniste à y séjourner, protégé par cette enceinte végétale, rempart sécurisant pour le sommeil de l'hôte.

<sup>39</sup> Jammes glose « tan hermosa y bien dispuesta como las de las ciudades », Góngora, *Soledades*, 302.



qu'ordre et civilisation règnent dans ce village, dont l'enceinte régulière protège le sommeil de l'hôte comme une muraille.

Le rapprochement couplé à un rejet peut avoir une visée paradoxale, comme dans [*seno blando/piélago*] *duro*,<sup>40</sup> avec le parallélisme antithétique *blando/duro*. Le vers se situe dans le passage sur les concours sportifs des jeunes villageois après le banquet de noces et l'épithalame : un villageois robuste mais agile s'élançait, puis retombe pesamment par terre, dans le sein mou de l'herbe, – mou, mais dur comparé à celui de la mer où est tombé Icare –. On a ici un jeu de redoublement de la première antithèse *ágil/robusto*, accentuant le décalage burlesque entre le caractère trivial de la scène et son expression solennelle, via la périphrase antonomastique *Ícaro montañés* et la prosodie épique du vers (endécasyllabe héroïque), dessinant une description héroï-comique de la chute cocasse d'un « Icare montagnard ». <sup>41</sup>

La séquence crée enfin des effets de prosodie expressive, comme dans la chronographie printanière qui ouvre la première *Soledad* :

Era del año la estación florida [...]
 cuando el que ministrar podía la copa
 a Júpiter mejor que el garzón de Ida,
 náufrago y desdeñado, sobre ausente,
 lagrimosas de *amor dulces querellas*
da al mar, que condolido,
 fue a las ondas, fue al viento
 el mísero gemido,
 segundo de Arión dulce instrumento.

(*Soledad Primera*, v. 1–14.)

<sup>40</sup> « [...] bien que impulso noble de gloria, aunque villano, solicita a un vaquero de aquellos montes, grueso, membrudo, fuerte roble, que, ágil a pesar de lo robusto, al aire se arrebata, violentando lo grave tanto, que lo precipita, Ícaro montañés, su mismo peso de la menuda hierba el *seno blando piélago* duro hecho a su ruina ». (*Soledad Primera*, v. 1002–11.)

<sup>41</sup> Jáuregui, ennemi littéraire de notre poète, s'offusqua de ces vers « plébéiens », surtout à cause de l'emploi de l'adjectif *grueso*. Francisco Fernández de Córdoba, érudit, ami et commentateur de Góngora, lui répondit que *grueso* ne signifie pas ici gros, mais grand, robuste. Voir Góngora, *Soledades*, 402, note 1004.

Ici, dans « *lagrimosas de [amor dulces querellas]* », l'assemblage permet de placer *amor* au centre du vers et sous l'accent, enchâssé entre *lagrimosas* et *querellas*, syntaxe expressive de l'amour malheureux et comme enfermé, enkysté dans la douleur.

### ...à celui de leur environnement syntaxique

De nombreuses instances ne se laissent pas aisément classer dans cette première catégorie (rapprochements expressifs, condensation de la pensée et de l'écriture), et nécessitent la prise en compte de leur environnement syntaxique afin d'être interprétées. C'est le cas des nombreuses instances (38 % des cas) intégrant des compléments antéposés (compléments circonstanciels et compléments du nom), dont nous proposons à présent une interprétation. Il ne s'agit pas d'étudier tous les cas d'antéposition de compléments chez Góngora, mais seulement ceux qui accueillent la séquence NC ADJ NC. En d'autres termes, le trigramme a fonctionné pour notre recherche comme indice d'une structure récurrente, associée à un effet de sens répété, celui de la recreation poétique du processus mental d'appréhension, d'abord sensorielle, puis intellectuelle, du monde, depuis le particulier vers le général. Nous livrons ici les premiers résultats de cette recherche préliminaire, qu'il faudrait approfondir par une étude systématique de l'antéposition des compléments chez Góngora et chez les auteurs de son temps.

Dámaso Alonso, dans un chapitre de *La lengua poética de Góngora* portant sur l'hyperbate, classait l'antéposition du complément comme sous-type possible de cette figure de construction.<sup>42</sup> L'auteur rappelle l'origine latine de l'antéposition du génitif, son emploi récurrent dans la poésie et la prose du xv<sup>e</sup> siècle, puis au xvi<sup>e</sup> chez Herrera, le statut de précurseurs des poètes andalous dans son extension en poésie, et signale donc qu'il n'est pas nouveau chez Góngora : « *cuando Góngora se lanza por la senda de los hipérbatos pisa un terreno ya largamente hollado durante dos siglos* ». Le poète en aurait simplement exploité la valeur expressive.

Voici quelques exemples : *de las estigias [aguas torpe marinero]*, *de su [edad corta historias] largas*, *de zagalejas [cándidas voz tierna]*, *de la amorosa [espuela duro freno]*, *de verde [laurel caduca rama]*, *de [estrellas inmortal corona]*. Le poète donne alors le propos avant le thème : il y a un déplacement du centre de gravité sémantique du vers, repoussé vers la droite (*marinero*, *historias*, *voz*, *freno*, *rama*, *corona*). Góngora créé ainsi une tension, un suspens, un effet d'attente, tout

<sup>42</sup> Dámaso Alonso, « Hipérbaton » in *Obras completas*, V, *Góngora y el gongorismo* (Madrid : Gredos, 1978), 188–226.

en insufflant une cinétique particulière au vers, comme prêt à basculer, précipité vers sa résolution lexicale, dans une déstabilisation à la fois syntaxique (dans la subversion de l'ordre standard des mots dans la phrase) et sémantique (le lecteur attend le thème), ce que j'appelle la « ligne anticipatrice » de la poésie gongorine, à la manière de la syntaxe de l'allemand décrite par Louis Truffaut dans ses *Problèmes linguistiques de traduction*.<sup>43</sup> S'intéressant à ces effets de suspens et d'incomplétion provisoire du sens dans le déroulé chronologique des textes, Yves Macchi a montré la manière dont l'ordre d'apparition des mots dans un texte change radicalement les représentations mentales générées à la lecture.<sup>44</sup> Le texte, « tout entier fait de temps », se laisse appréhender dans sa linéarité par la chronosyntaxe.

Or, l'attente créée par l'antéposition des compléments et la perplexité qui en découle sont tous deux des traits de la figure appelée en grec *plagiasmos*, typique du style héroïque, à laquelle Le Tasse fait allusion sans la nommer dans ses *Discorsi* et que définit Hermogène de Tarse dans *Sur les formes du style*.<sup>45</sup> Le phénomène a été décrit en détail par Mercedes Blanco dans *Góngora o la invención de una lengua* : il s'agit de commencer une période par des cas obliques (circonstanciels, compléments, COD) pour reporter la nomination de l'essentiel en fin de phrase.<sup>46</sup> Mercedes Blanco parle ainsi des « acrobaties » mentales auxquelles Góngora oblige son lecteur, de la ligne sinueuse

<sup>43</sup> « La séquence française, allant du général au particulier, est dite séquence progressive ou linéaire. La séquence allemande est dite régressive, anticipatrice, centripète ou antilinéaire, selon le point de vue auquel on se place », cité par Michaël Oustinoff, « L'adjectif antéposé de l'anglais : problème de traduction ? », *Palimpsestes* 14, (2001) : 99–117.

<sup>44</sup> Yves Macchi, « Sculpture du sens et sens d'une sculpture (chronosyntaxe VIII) », *Chréode* 2 (2018) : 143–85, ici 146–52.

<sup>45</sup> Voici le passage portant sur le *plagiasmos* : « Les cas obliques, puisque leur emploi implique des propositions subordonnées, introduisent des pensées contingentes, secondaires. Si vous utilisez la subordination et dites « Lorsque Crésus était » ou « Depuis que Candaule est », le style n'est plus clair ni pur. Il y a d'emblée une confusion, car il est nécessaire qu'une autre pensée suive », Hermogène de Tarse, *Hermogenes' On types of style*, trad. par Cecil W. Wooten (Chapel Hill : University of North Carolina Press, 1987), 231 ; je traduis à partir de la traduction anglaise du grec de Cecil Wooten.

<sup>46</sup> « *Entendamos que la frase, o el verso, no comienza con el sujeto ni con el predicado sino con algún componente verbal 'oblicuamente' conectado con ellos, por lo tanto por lo que en griego o en latín sería un dativo, o mejor un genitivo o ablativo [...]. A nivel no ya sintáctico, sino pragmático, empezar con casos oblicuos significa no empezar nombrando aquello de qué se nos va a hablar, que los lingüistas llaman el TEMA, sino adelantando otro término asociado con este de modo lateral o circunstancial* », Blanco, *Góngora*, 286.

du vers, et des mille détours circonstanciels dans lesquels nous emmène cette « chasse herméneutique ». <sup>47</sup>

À partir de cette observation de l'importance de l'antéposition des compléments chez Góngora, nous aimerions proposer une interprétation unifiée de ce trait idiolectal, <sup>48</sup> qui vaudrait à la fois pour le complément du nom et le complément circonstanciel. <sup>49</sup> Dans son article sur le sujet, Domingo Yndurain, après un recensement des emplois dans le *Poema del Cid*, propose son interprétation : <sup>50</sup> les langues romanes, analytiques dans leur syntaxe, font apparaître d'abord les éléments essentiels, comme le nom, puis les éléments secondaires, comme le complément du nom ou l'adjectif, qui viennent apporter un renseignement supplémentaire sur la chose nommée. <sup>51</sup> Inverser cet ordre normal reviendrait dès lors à chercher à reproduire dans le discours l'ordre sensoriel de perception du monde, qui va du particulier au général, là où les langues romanes disent le monde du général au particulier. <sup>52</sup>

Yndurain prend l'exemple de Garcilaso : « de verdes sauces hay una espesura » : où l'on perçoit d'abord le prédicat de la couleur verte, puis l'arbre, enfin son ordonnancement en fourré, selon un mouvement qui part d'une épithète sensorielle (*verdes*) pour aller ensuite vers un substantif conceptuel (*espesura*). Nous aimerions faire nôtre cette analyse et l'appliquer dans les lignes qui suivent à la compréhension de l'emploi que fait Góngora de l'antéposition des compléments où figurent la séquence NC ADJ NC.

<sup>47</sup> Blanco, *Góngora*, 290.

<sup>48</sup> L'idiolecte désigne l'emploi préférentiel ou exclusif par un individu d'un ensemble de traits linguistiques (lexicaux, syntaxiques). Voir Christophe Gérard, « L'individu et son langage : idiolecte, idiosémie, style », *PhiN* (2010) : 1–40.

<sup>49</sup> Bien que nous n'ayons travaillé que sur les 21 instances où figurait le trigramme NC ADJ NC à l'intérieur d'un complément.

<sup>50</sup> Domingo Yndurain, « De verdes sauces hay una espesura : anteposición de complemento con < de > », *Vox Romanica* 30–31 (1971) : 98–105. <http://www.cervantesvirtual.com/nd/ark:/59851/bmcrv164>

<sup>51</sup> « *Las lenguas románicas, en concreto, parecen reflejar una concepción de la realidad que correspondería a las categorías aristotélicas o tomistas [...]. Las cosas son definidas por su esencia, [...] y de aquí se separan analíticamente los accidentes. Por otra parte, la realidad es concebida en planos o categorías en las que se acepta un orden que va de lo general a lo particular, del todo a las partes* », Yndurain, « De verdes sauces hay una espesura ».

<sup>52</sup> « *Desde otro punto de vista puede suceder [...] que el hablante quiera reproducir el proceso de percepción de la realidad, la gestación de su conocimiento en un camino que va desde las categorías particulares a las generales, del detalle individual a la concepción unitaria y totalizadora* », Yndurain, « De verdes sauces hay una espesura ».

En donnant à son lecteur les circonstances de l'action avant l'action elle-même, Góngora crée une véritable *poétique des circonstances*, dans une primauté paradoxale de l'adventice sur l'essentiel. Dans le vers « *por su [edad majestuosa cano]* », <sup>53</sup> le propos (*cano*) est postposé à sa circonstancialisiation, par le complément de cause « *por su edad* ». Le vers décrit, dans la *Soledad segunda*, le lendemain de la rencontre avec la famille de pêcheurs, un château ancien illuminé par les rayons du soleil levant : sa blancheur s'explique autant par la pureté de son marbre que par son âge avancé. Le trigramme présente ici une homogénéité sémantique, rapprochant la vieillesse et son attribut chromatique traditionnel, la blancheur. La séquence *quejas/purpúreos troncos* intègre elle aussi un syntagme où chromatisme et forme apparaissent avant le matériau lui-même :

Sorda hija del mar, cuyas orejas  
 a mis gemidos son rocas al viento :  
 o dormida te hurten a mis *quejas*  
*purpúreos troncos* de corales ciento,  
 o al disonante número de almejas  
 –marino, si agradable no, instrumento–  
 coros tejiendo estés, escucha un día  
 mi voz, por dulce, cuando no por mía. (*Polifemo*, octava 48.)

Il s'agit de l'adresse de Polyphème à Galatée, la suppliant de le suivre, après que le géant a vu les ébats de celle-ci avec Acis. Le trigramme rapproche de manière étrange *quejas* de *purpúreos troncos*, situant le discours amoureux dans un paysage inattendu, non celui de la forêt, mais au fond de la mer, parmi les coraux et les coquillages aux claquements dissonants. Ce faisant, Góngora se livre à une recontextualisation inédite du topos pétrarquiste de la plainte amoureuse, dans un jeu de déplacement du familier (plaintes

<sup>53</sup> « Dejaron pues las azotadas rocas,  
 que mal las ondas lavan  
 del livor aún purpúreo de las focas,  
 y de la firme tierra el heno blando  
 con las palas segando,  
 en la cumbre modesta  
 de una desigualdad del horizonte,  
 que deja de ser monte  
 por ser culta floresta,  
 antiguo descubrieron blanco muro,  
 por sus piedras no menos  
 que *por su edad majestuosa cano* » (*Soledad segunda*, v. 686–97.)

de l'amant devant l'aimée) dans un environnement étrange et exotique (les fonds maritimes, corallins et conchyliens), forme de paradoxographie érotique.

Cette poétique des circonstances est aussi celle de la prédominance du sensoriel sur l'intellection. Dans les vers finaux de la première *Soledad*, après le mariage des deux villageois et le banquet tant rustique qu'abondant qui s'ensuit, les époux se rendent chez eux, accompagnés d'un cortège de jeunes gens. Vénus a préparé leur lit de noces, le couvrant des plumes les plus douces des cygnes tirant son char. Or, dans les vers « [el lecho ha prevenido] de las plumas que baten, más süaves/en su volante carro blancas aves », <sup>54</sup> le bruissement des plumes et la douceur des ailes sont nommés avant le concept d'oiseau. Dans cette poésie synesthésique, où son et matière (*baten, süaves*), adventices, précèdent l'essence conceptuelle (*aves*), l'antéposition de la circonstance donne à l'ensemble un fonctionnement synthétique. Dans ce passage, empreint d'un érotisme latent, la post-position du thème renforce l'impression de douceur et de bien-être, par l'animation mystérieuse des plumes, semblant battre seules, dénuées de tout support corporel, comme le vers de tout support thématique.

L'antéposition des compléments, dont le trigramme est ici la trace, pourrait ainsi fonctionner comme l'antéposition de l'épithète en linguistique : en anglais, « a red car » implique la perception synthétique du syntagme comme un tout cohérent, au contraire de « voiture rouge » qui implique la vision analytique d'un objet auquel on applique un qualificatif secondaire. De la même manière, *de estrellas corona* n'est pas équivalent à *corona de estrellas* : la perception est sensorielle et synthétique dans le premier cas, conceptuelle et analytique dans le second. Observons ce qu'il en est dans le sonnet à Juan Rufo (« A Juan Rufo, jurado de Córdoba ») :

---

<sup>54</sup> « Llegó todo el lugar, y despedido,  
 casta Venus, que el lecho ha prevenido  
 de las plumas que baten más süaves  
 en su volante carro blancas aves,  
 los novios entra en dura no estacada ;  
 que, siendo Amor una deidad alada,  
 bien previno la hija de la espuma  
 a batallas de amor campo de pluma. » (*Soledad Primera*, v. 1084–91.)

Culto jurado, si mi bella dama,  
 en cuyo generoso mortal manto  
 arde, como en cristal de templo santo,  
 de un limpio amor la más ilustre llama,  
  
 tu musa inspira, vivirá tu fama  
 sin envidiar tu noble patria a Manto,  
 y ornarte ha, en premio de tu dulce canto,  
 no *de verde* [*laurel caduca rama*],  
 sino *de* [*estrellas inmortal corona*].  
 Haga, pues, tu dulcísimo instrumento  
 bellos efectos, pues la causa es bella,  
  
 que no habrá piedra, planta, ni persona,  
 que suspensa no siga el tierno acento,  
 siendo tuya la voz, y el canto de ella.

Il s'agit d'un sonnet dédié au poète épique de la *Auñtríada*. Le poème est construit tout entier sur un leurre : s'ouvrant sur l'apostrophe au *culto jurado*, il dessine la feinte d'une thématique littéraire (la louange d'un ami poète), alors qu'il s'agit de la demande triviale à celui-ci de prendre sa dame pour muse. La thématique amoureuse domine (mais de manière oblique) le sonnet, dans le rejet expressif de *arde*, la métaphore paradoxale de la flamme-cristal, et le retour final sur l'aimée. Le poème dessine ainsi une structure circulaire (de la dame à la dame),<sup>55</sup> camouflée par le leurre de l'apostrophe initiale (*culto jurado*). Góngora va même jusqu'à énoncer la supériorité de la beauté de l'aimée à celle des vers de son ami (« bellos efectos, pues la causa es bella »). Les promesses péremptives du second quatrain, au futur, font office de chantage poétique, assurant gloire et immortalité au thuriféraire de sa dame.

Étudions à présent l'effet du double trigramme, incorporé à des compléments du nom, dans les vers « no de verde [*laurel caduca rama*]/sino de [*estrellas inmortal corona*] ». L'antéposition du lexème *estrellas* par rapport à *corona* entraîne la prédominance de la sémantique stellaire sur la sémantique royale. Les sèmes de hiératisme, pouvoir suprême, sagesse inclus dans *corona* sont écrasés par les sèmes d'éclat lumineux, immatérialité éthérée, rayonnement cosmique de *estrellas* placé en première position. Ce dernier mot domine le vers, sémantiquement et rythmiquement (il est placé sous

<sup>55</sup> Le premier quatrain est entièrement dédié à la dame, véritable sujet du sonnet, et le dernier tercet quasiment entièrement.

l'accent), avec l'emphase prosodique de l'anisochronie, dans le passage de l'accentuation en 3, 6, 8, 10 à 4, 8, 10, et du rythme dactylo-trochaïque à dactylo-iambique.<sup>56</sup> Góngora dessine ainsi l'apothéose cosmologique du poète, dans le passage de la métaphore banale (couronne de laurier) à la métaphore audacieuse (couronne d'étoiles), du matériel à l'immatériel, du possible à l'impossible, du mat au brillant, du végétal au stellaire. Le trigramme *laurel caduca rama* pourrait ainsi constituer une mise en abyme métaphorique du processus de renouvellement poétique mis en œuvre par Góngora : de même que le laurier, symbole du poète, est dit caduc, la métaphore du laurier poétique, rebattue, est remplacée par l'image nouvelle, audacieuse, de la couronne d'étoiles, figurant le poète comme inspiré des cieux, et possesseur d'une connaissance supérieure.

### Bilan de l'étude et perspectives

Nous avons rencontré plusieurs difficultés au cours de notre travail. Les trigrammes, SGR hétérogènes, se révèlent difficiles à appréhender et à interpréter. Nous avons été confrontés à un problème d'échelle : afin de comprendre le rôle des trigrammes au sein des périodes, il a fallu revenir à l'échelle plus grande du groupe grammatical qui les accueillait. L'objet de l'étude s'est alors déplacé des trigrammes eux-mêmes aux structures syntaxiques qu'ils incorporent. Ce déplacement se justifie par l'appartenance répétée (38 % des cas) des trigrammes à une structure avec antéposition des compléments. On peut en ce sens envisager les SGR comme traces microtextuelles d'un phénomène plus large et récurrent, dont il faudrait vérifier s'il est typiquement gongorin dans son usage (du point de vue quantitatif et qualitatif). Si l'effet poétique du trigramme proprement dit est souvent celui d'un rapprochement expressif, voire paradoxal, entre entités hétérogènes, signe d'une condensation de l'expression et de la pensée, la fonction poétique de l'antéposition du complément intégrant la séquence NC ADJ NC, qu'il faudrait analyser de manière plus exhaustive, serait celle de la recréation poétique du processus mental de conceptualisation, du particulier au général.

<sup>56</sup> Nous nous fondons sur l'analyse de Tomás Navarro Tomás du *período rítmico interior*, commençant par la première syllabe accentuée (incluse), et terminant sur la dernière accentuée (incluse), le comput rythmique des trochées, dactyles et autres pieds s'effectuant uniquement au sein de la période rythmique intérieure, laissant de côté les syllabes non accentuées de début et de fin de vers. Voir la description qu'en donnent Madeleine et Arcadio Pardo, *Précis de métrique espagnole* (Paris : Armand Colin, 2010), 35–40.



Cette étude a également été l'occasion de réfléchir aux notions de style et d'idiolecte, et à la pertinence de l'emploi de ces termes en stylométrie. À y regarder de plus près, ce que quantifient les algorithmes au sein de tables de données fréquentielles, ce n'est aucunement le style, celui-ci étant inséparable de l'effet de sens que produit le texte et des fonctions pragmatiques, rhétoriques de ses éléments :<sup>57</sup> ce que l'on trouve dans ces tableaux de chiffres n'est autre que l'ensemble des données *linguistiques* d'un *idiolecte*.<sup>58</sup> Idiolecte et style seraient ainsi deux objets d'étude correspondant aux deux étapes successives de l'herméneutique numérique, définie par Damon Mayaffre :<sup>59</sup> d'abord l'extraction automatique et les calculs de distance se font sur des unités *linguistiques*, l'ensemble desquelles, assorties de leurs fréquences, constitueraient un *idiolecte* – et l'idiolecte serait en ce sens objectivable sous la forme d'un tableau de chiffres –, puis, l'analyse qualitative et l'interprétation, par le retour au texte, la contextualisation et le recours aux méthodes

<sup>57</sup> Nous suivons en ce sens Alain Rabatel et André Petitjean dans leur introduction au numéro de la revue *Pratiques* consacrée au style : le style « implique une < vision du monde >, c'est-à-dire des < valeurs politiques, idéologiques, éthiques, esthétiques > », André Petitjean et Alain Rabatel, « Le style en questions », *Pratiques* 135/136 (Décembre 2007) : 3–14, ici 7.

<sup>58</sup> On peut en effet considérer que les usages linguistiques sont quantifiables : chaque trait (proportion de tel mot, adjectif, tendance ou non à l'inversion adjectif-nom, fréquence de « mots-outils », fréquence de termes vulgaires) peut être assorti de sa fréquence. Les traits linguistiques sont en ce sens des observables statistiques, on peut les délimiter, les décrire, les mesurer.

<sup>59</sup> Selon Damon Mayaffre, l'herméneutique numérique suppose une démarche épistémologique inversée, où les hypothèses de travail sont objectives, émergeant du corpus lui-même : « La démarche épistémologique face au texte se trouve donc inversée : là où traditionnellement le chercheur interrogeait le texte sur la base d'hypothèses de travail construites, c'est le texte qui interroge le chercheur *sans tabou et sans a priori*. Par une lecture différente (hypertextuelle plutôt que linéaire, nous l'avons vu, mais aussi paradigmatique plutôt que syntagmatique, quantitative plutôt que qualitative), l'ordinateur voit autre chose pour *déranger nos certitudes* et élargir l'horizon étroit de nos modes (aux deux sens du terme) d'interrogation. » Il s'agit de détecter des faits objectivement marquants, c'est-à-dire saillants du point de vue quantitatif, pour ensuite laisser place à l'analyse qualitative par le retour au texte : « Avec la lexicométrie, ce sont les *informations objectivement pertinentes* du corpus qui *remontent*, en bon ordre, jusqu'au chercheur. D'une certaine manière, l'ordinateur dégrossit le texte pour *attirer notre attention sur les faits objectivement marquants* », (je souligne) dans une forme de nouveau positivisme. L'interprétation est toujours aussi importante, mais celle-ci se fait désormais sur « des bases irréfutables ». Le nouveau modèle épistémologique proposé par Damon Mayaffre sous l'appellation « herméneutique numérique » correspond tout à fait à notre approche en deux temps, quantitatif puis qualitatif, linguistique puis stylistique, de détection automatique puis d'interprétation humaine et littéraire. Damon Mayaffre, « L'herméneutique numérique », *L'Astrolabe : recherche littéraire et informatique* (2002) : 1–11, ici 5.

de la stylistique traditionnelle, visent à interpréter chaque trait quantifiable de cet idiolecte, en le caractérisant de manière qualitative via son *effet de sens*, son rendement expressif ou pragmatique, afin de parvenir à la caractérisation d'un *style*. L'idiolecte ne serait que le squelette linguistique, mesurable, du style, lui-même irréductible à un tableau de fréquences.

\*

\*\*

L'extraction de SGR constitue donc un apport prometteur à l'étude du style gongorin. Les méthodes de la stylistique computationnelle apportent une nouvelle objectivité grâce à l'extraction automatique, entièrement informatisée, sans intervention humaine et donc sans a priori, des séquences syntaxiques les plus employées. Ici, l'extraction des trigrammes de POS-tags et les mesures de comparaison à un corpus-témoin ont permis d'identifier une séquence massivement surreprésentée chez Góngora. L'analyse qualitative, second temps de ce que Damon Mayaffre a appelé l'« herméneutique numérique », <sup>60</sup> nous a conduit à définir plusieurs effets de sens récurrents pour la séquence NC ADJ NC. Cette séquence, tant en elle-même que comprise comme trace d'un phénomène syntaxique plus large, l'antéposition des compléments, pourrait constituer un motif gongorin : dans le premier cas, celui du rapprochement expressif d'éléments discordants, dans le second, celui de la recreation poétique du mouvement d'appréhension intellectuelle du monde, du particulier au général.

---

<sup>60</sup> Voir Mayaffre, « L'herméneutique numérique », 5.

# In Search of a New Language

## Measuring Style of Góngora and Picasso

Nanette Rißler-Pipka (Karlsruhe)

**ABSTRACT:** Both authors are considered to be ‘unreadable’, both are separated by more than 300 years, both authors are said to have invented a new language or to use language in a mannered way. It is well known that Picasso, like many of his contemporaries, was fascinated by Góngora and made broadly use of other artworks and literature. But how can one compare both authors using digital methods? The language of the 16<sup>th</sup>/17<sup>th</sup> century cannot be compared with that of the 20<sup>th</sup> century. Nevertheless, there are ways and means of opening up new perspectives on the subject of author and style. Here are examples of different digital methods presented and used: Part-of-speech-tagging, stylometry with *stylo* for R and Topic Modeling. In addition, the connection between stylistics and stylometry is discussed.

**ZUSAMMENFASSUNG:** Beide Autoren hält man für ‚unlesbar‘, beide trennen mehr als 300 Jahre. Von beiden Autoren sagt man sie hätten eine neue Sprache erfunden oder die alte Sprache verkünstelt. Dass Picasso wie viele seiner Zeitgenossen von Góngora fasziniert war und sich in der Kunst- wie in der Literaturgeschichte bediente, ist bekannt. Wie soll man aber auf Ebene des digitalisierten Werks beide Autoren vergleichen? Die Sprache des 16./17. Jahrhunderts lässt sich nicht mit derjenigen des 20. Jahrhunderts vergleichen. Dennoch gibt es Mittel und Wege, neue Perspektiven auf das Thema Autor und Stil zu eröffnen. Hier werden exemplarisch verschiedene digitale Methoden vorgestellt und verwendet: Part of Speech Tagging, Stilometrie mit *stylo* für R und Topic Modeling. Außerdem wird die Verbindung zwischen Stilistik und Stilometrie diskutiert.

**SCHLAGWÖRTER:** Picasso, Pablo; Góngora, Luis de; Autorschaft; Stil; Stilometrie; Topic Modeling

Today, concepts of intertextuality make it possible to compare authors from completely different times and style, because every author takes his ideas, language and some parts of his style from his predecessors. Picasso as a painter is particularly known for recreating masterpieces of art history (Velázquez, El Greco, Goya, Manet, etc.), but he is rather unknown for his poetry, and for reinventing Góngora’s style of obscure and dazzling poetry.<sup>1</sup>

<sup>1</sup> Picasso also illustrated Góngora’s sonnets in 1948, for an analysis in context with his writings see Nanette Rißler-Pipka, *Picasso’s schriftstellerisches Werk: Passagen zwischen Bild und Text* (Bielefeld: transcript, 2015), 297–313.

About both poets, Góngora and Picasso, we can read in contemporary and actual literature that they are “unreadable”, treating language as a puzzle to be solved or at least as game of combinatorics (trying new syntaxes, new words, new grammatical orders). Speaking about combinatorics, there are many points that link mathematics, digital humanities and literary criticism. Borges comments on the importance of combinatorics, games and literature:<sup>2</sup>

hacer de la metafísica, y de las artes, una suerte de juego combinatorio. Quienes practican ese juego olvidan que un libro es más que una estructura verbal, o que una serie de estructuras verbales; es el diálogo que entabla con su lector y la entonación que impone a su voz y las cambiantes y durables imágenes que deja en su memoria<sup>3</sup>

Borges is criticizing in the context of the Avant-garde a certain degree of arbitrary which was also responsible for the end of Avant-garde literature and art. Plus, he is pointing out, that the process of reading is not countable or measurable in all its parts, because it is a dynamic dialog between each reader and the text. Nevertheless, the idea of playing a game of combinatorics when constructing a text is quite common in the context of Avant-garde literature as well as in the context of emblematic art/poetry and mannerism. If it is a game, it should have rules that can be described.

From a mathematical point of view, combinatorics do not model rational sequences but countable structures. Applying this model on the literary text, it seems possible to detect the rules behind the structure. If we define a text as a combination of elements (letters, words, lemma, interpunction, POS, n-grams, etc.) we can count these elements and compare texts to find a pattern which might be characteristic for authorial style or genre. That is in a very reduced way an explanation of what stylometry is about. What we are not looking at, is the sense of the text. The semantic part of it is explicitly taken out of our view. That point is very important when talking about authors known for their pleasure in obscuring sense. Probably on another level, in the structure of combining elements, there is a pattern telling us some-

---

<sup>2</sup> See also on this subject: Christoph Strosetzki, “Von der Metaphysik-und Sprachkritik des Wiener Kreises zu Jorge Luis Borges und Ernesto Sábato”, in *Wort und Zahl = Palabra y número* (Heidelberg, 2015), 289–300 and Volker Roloff, “Streifzüge durch surreale Bibliotheken: von Rabelais und Borges zum Internet”, in *Durchquerungen: Ralf Schnell zum 65. Geburtstag*, eds. Iris Hermann und Anne-Maximiliane Jäger-Gogoll (Heidelberg: Winter, 2008) 225–34.

<sup>3</sup> Jorge Luis Borges, “Nota sobre (hacia) Bernhard Shaw”, in *Otras inquisiciones* (1952), 93, [https://apuntesliterarios.files.wordpress.com/2013/09/borges\\_otras\\_inquisiciones.pdf](https://apuntesliterarios.files.wordpress.com/2013/09/borges_otras_inquisiciones.pdf).

thing about the way these authors are constructing their texts ingeniously. The appalling effect of the “unreadable” text is already known but we would like to detect a rule (if there is one), like a new grammar, a solution of the combinatoric game. One could reasonably object that authors like Góngora and Picasso are probably pushing language intentionally to its limits where no rule or grammar is detectable, so why looking for it. But even if we take this as a hypothesis it would be nice to prove it with more evidence than our pure reading. Second hypothesis would be, that Góngora is rather using a kind of grammar or rule (we know already that he uses hyperbaton etc. to create new effects of language and grammar)<sup>4</sup> than Picasso. That is no revolting news to be proved but in comparison to other authors of their time it tells us about the influences and depending on the method also about how the poetry is constructed (apart from semantic riddles).

Nevertheless, it is obvious that no analysis can be free of the semantic level. When we look at the word which is used most in a corpus, we are reflecting about the “what does that mean?” question. And even before that: when we select our texts for the corpus decisions are made, based on the context knowledge, based on the semantics of each text. There is no claim of objectivity by using digital methods but it is in fact a completely different kind of looking at our texts and usually the path of knowledge becomes traceable.

Without the intention of presenting a solution, but rather with the intention of presenting various possibilities of stylometric analysis and discussing the advantages and disadvantages, I tested three different methods.

1. For microanalysis there would be POS (Part-of-speech-tagging) which tells us the concrete grammatical order of each sentence and how many sentences are structured the same way, etc.
2. For macroanalysis and a comparative view: Stylometry. Here it does not make sense to compare Góngora and Picasso directly because the corpora should be homogeneous regarding the period, number of words and genre.<sup>5</sup> But, we will see if it is possible to create two corpora for each author in its time and then to compare the results.

<sup>4</sup> For a detailed analysis of Góngora inventing a new language see Mercedes Blanco, *Góngora o La invención de una lengua, Lectura y signo* (León: Universidad de León, 2012).

<sup>5</sup> Cf. Christof Schöch, “Corneille, Molière et les autres: stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik”, *PhiN* 7 (2014): 130–57, <http://web.fu-berlin.de/phn/beiheft7/b7to8.pdf>; Maciej Eder, “Does size matter? Authorship attribution, short samples, big problem”, in *Digital Humanities 2010: Conference Abstracts*, 2010, 132–5.

3. For a look into the sematic level, but without the usual given preferences, one could try Topic Modeling or Word Embeddings (word2vec). Here only Topic Modeling is tested rudimentarily for Picasso.

Before comparing and testing these three digital methods we should have a look on how traditional stylistics would proceed. In a rather one-sided, reductive way Sowinski is looking on Spitzer and stylistics from an actual point of view. His most important reproach is the intuitive way of analyzing literature:

1. intuitive Detailbetrachtung;
2. Feststellung von Gemeinsamkeiten im scheinbar Zufälligem;
3. Rückschluss auf den Seelenzustand des Verfassers beim nochmaligen Lesen des Ganzen<sup>6</sup>

Reading means here: intuitive micro-analysis and observation of details, which catches our individual interest and next step would be to connect some parts of the text based on arbitrary found similarities. In a last step all this is transferred to the complete work and soul of the author in describing his personal style. We know that Spitzer did not analyze texts in this simple way, and that stylistics today are used to distinguish our 'digital' selves from an intuitive way of reading. Though this image of stylistics might be the reason why it is no longer the state of the art in literary criticism.

Nevertheless, there are quite similar reproaches against stylistic analysis today:

[stylistics] still operates according to the principle prove by examples. It should by now be clear that under such a regimen any theory can be proven right: one can always find an example that illustrates a theory, however outlandish it maybe.<sup>7</sup>

The solution for van Peer, Zyngier and Hakemulder seems to be statistical tests which are quite easy to elaborate for stylistics. But, proving theories with statistical values does not automatically mean that they are more valid than those who are using other methods and models. It always depends on the choice you make what to count and measure. It is true though that the

<sup>6</sup> Bernhard Sowinski, *Stilistik: Stiltheorien und Stilanalysen*, 2<sup>nd</sup> rev. ed., Sammlung Metzler 263 (Stuttgart: Metzler, 1999), 138.

<sup>7</sup> Willie van Peer, Sonia Zyngier, and Jèmeljan Hakemulder, "Foregrounding, Estrangement and Pattern", in *Stylistics: Prospect & Retrospect*, ed. by David L. Hoover, PALA Papers 3 (Amsterdam: Rodopi, 2007), 1–22, here 5.

chance to test arguments with statistics makes your hypotheses verifiable and if providing the used data also intersubjectively verifiable.

Still, we should not forget that in literary criticism one could never measure statistically all factors of theory. For example, for the concept of intertextuality each text is the result of many others plus the predisposition of the reader. For Borges the identity of the author, his soul or his style is not at all important because:

El hecho es que cada escritor crea sus precursores. Su labor modifica nuestra concepción del pasado, como ha de modificar el futuro. En esta correación nada importa la identidad o la pluralidad de los hombres.<sup>8</sup>

### Measuring and Tagging Picasso

Both, Picasso and Góngora are using prepositions (most frequently “de”) to construct sequences that seem to be readable backwards as well as in the given order (forwards). A very often cited example for Góngora is “montes de agua y piélagos de montes”<sup>9</sup> which describes the disordered perception of the young shipwrecked who was just washed up the beach at the beginning of *Soledades*. We are not sure if he describes the giant waves as mountains or the giant mountains as waves or sea in general.<sup>10</sup> How could we ever represent (or model) an effect of reader-irritation and uncertainty of sense in a quantitative and measurable way? Probably not the effect itself, but the significant frequency of the same grammatical construction could show us the pattern of a strategy used by the author all over his work. While, the examination and interpretation of several examples explains the effect itself and its meaning.<sup>11</sup>

For Picasso we do not have very well analyzed examples. Therefore, we draft a possible way of how methods like POS-tagging could work but with-

<sup>8</sup> Jorge Luis Borges, “Kafka y sus precursores”, in *Obras completas* (Buenos Aires 1974), 711–12.

<sup>9</sup> Luis de Góngora, *Soledades*, ed. by John Beverley (Madrid: Catedra, 142009), V. 44, 77.

<sup>10</sup> See for more details Bernhard Teuber, “Curiositas et crudelitas: das Unheimliche am Barock bei Góngora, Sor

Juana Inés de la Cruz und José Lezama Lima”, in *Diskurse des Barock: dezentrierte oder rezentrierte Welt?*, ed. by Joachim Küpper and Friedrich Wolfzettel (München: Fink, 2000), 615–52 and Reißler-Pipka, *Picassos schriftstellerisches Werk*, 261–91.

<sup>11</sup> Both parts of analysis, the quantitative and the hermeneutic qualitative are shown in the study by Marie-Eglantine Lescasse, “Góngora hors norme? Étude stylométrique d’un motif góngorin”, in this volume, p. 91.

out really proceeding them seriously to come to results.<sup>12</sup> A starting point for the argumentation is the hypothesis that Picasso's writings are extremely unconventional even for his period, the avant-garde. That is one of the common ways of arguing in the context of literary history that the author X has an outstanding individual style which needs more detailed analysis than others and is worth a redefinition of the canon.

As we already know, that Picasso does not write in sentences (he is rejecting punctuation and other grammatical order). There is the possibility that he creates his own grammatical rule or his own language. To detect the order on the level of microanalysis Part-of-speech-tagging is a common method in linguistics. What we would like to count is the number of prepositions, verbs, nouns, adverbs, personal pronouns, etc. to compare this number with other authors. Picasso's system of connecting the disparate in a baroque manner is completely dependent on prepositions and pronouns:<sup>13</sup>

la ilusión del drama que se juega la vida sobre la última carta  
y hace su castillo de naipes sobre el filo de la navaja del son  
del canto de cisne del clarín que se muere de pena nadando  
buzo ciego por las tripas verdes del lago por donde Ofelia va  
buscando en el fondo un trozo de papel y un lápiz

Reading this rather randomly chosen verses of a longer prose poem (dated 02.02.1936), Picasso's style is intuitively detectable as a form of adding new aspects to a poetic image by enlarging a never-ending sentence. We understand motives like the Ophelia myth (the lake, the swan, swimming, diving) or theatre (illusion, drama, playing, life, song, horn ("clarín"), probably also the house of cards and dying of pain) or writing/painting (letter, poetry ("canto" can mean song and poetry as well), the colour "green", sheet of paper and pencil). The remaining nouns which are more difficult to understand are knife edge ("el filo de la navaja") and tripe ("tripas"). They are part of the kitchen and food vocabulary which has an important role in Picasso's writings and at once belong to the leitmotif of bullfight.<sup>14</sup> Analyzing his motives and trace them back to predecessors like Góngora can be done by examples

<sup>12</sup> That would be another study to do. This paper is a reflection about the possible methods of stylometry in comparison to known theories of authorship and stylistics.

<sup>13</sup> Pablo Picasso, *Écrits*, ed. by Christine Piot and Marie-Laure Bernadac (Paris: Gallimard, 1989), 100–1 [Highlighting of prepositions and pronouns by N. R-P.].

<sup>14</sup> Particularly the tripe ("tripas") of the horse in the bullfight which is visible when the bull cuts the stomach of the horse with his horn is a very known and discussed motive in Pi-



like this citation, because the very significant point about Picasso's writing is, that he is repeating these motives and vocabulary over and over again in 200 prose poems (60.805 words) written in Spanish and 346 prose poems (56.518 words) written in French.<sup>15</sup>

Though, what we cannot detect by this kind of hermeneutic, stylistic (qualitative) analysis is the function of the quantitative significant number of prepositions and pronouns. As we see in the few verses, the prepositions and pronouns are responsible for the loss of sense. There is no semantic link between the motives like "castillo de naipes" and "el filo de la navaja" and "son del canto" and "cisne" and "clarín" – if we cut the prepositions and pronouns out ("su, sobre, de, del") and leave only those who are part of a collocation. There is still a link between "son del canto" and "clarín" but because of the insertion of "de cisne del" it becomes: the sound of the song of the swan of the horn – and this is obviously a very broken and crooked image. We are literary forced to try to make sense out of the bizarre combination of motives Picasso is presenting us because he collates them via prepositions and pronouns – a technique very similar to collage and montage in plastic arts as well as in avant-garde literature. All this should be reason enough for Part of Speech Tagging. To test it, without much technical effort and knowledge in linguistic tools, I tried the demo version of FreeLing<sup>16</sup> with the same verses (cited above).

For the few lines of verse (67 words = token) we get as an output a list of part of speech-tags for each token and an even more detailed XML-file (11 pages) where each token gets a description, see fig. 1.

---

casso's painting and writing. See Rißler-Pipka, *Picassos schriftstellerisches Werk*; Marie-Laure Bernadac, "Le Gazpcho de la Corrida", in *Picasso, toros y toreros* (Paris: RMN, Seuil, 1993), 47–59.

<sup>15</sup> For an overview and comparison see Nanette Rißler-Pipka, "Picasso et son esthétique numérique", in *Zwischen Genres und Medien: Formen moderner Prosa in Frankreich*, ed. by Christof Schöch and Nanette Rißler-Pipka, *PhiN Beiheft* 16 (2019), <http://web.fu-berlin.de/phin/beiheft16/b16to4.pdf>.

<sup>16</sup> FreeLing is created in Spain and therefore probably more adopted to Spanish language than other POS tools. For further information about the tool: Lluís Padró and Evgeny Stanilovsky, "FreeLing 3.0: Towards Wider Multilinguality", in *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)* (Istanbul, Turkey: ELRA, 2012), <http://nlp.lsi.upc.edu/publications/papers/padro12.pdf>.

## Analysis Results

## ▼ Language Identification

Identified language is: Spanish (es)

## ▼ Sentences

## Sentence 1

la	ilusión	de	el	drama	que	se	juega	la	vida	sobre	la	última	carta	y	hace
DA0FS0	NCF5000	SP	DA0MS0	NCMS000	PR0CN00	PO0CN00	VMP3S0	DA0FS0	NCF5000	SP	DA0FS0	AO0FS0	NCF5000	CC	VMP3S0

▼ CONLL format

1	la	el	DA0FS0	DA	pos=determiner type=article gen=feminine num=singular	-	-	-	-	-	-	-	-	-	-
2	ilusión	ilusión	NCF5000	NC	pos=noun type=common gen=feminine num=singular	-	-	-	-	-	-	-	-	-	-
3	de	de	SP	SP	pos=adposition type=preposition	-	-	-	-	-	-	-	-	-	-
4	el	el	DA0MS0	DA	pos=determiner type=article gen=masculine num=singular	-	-	-	-	-	-	-	-	-	-
5	drama	drama	NCMS000	NC	pos=noun type=common gen=masculine num=singular	-	-	-	-	-	-	-	-	-	-

Figure 1: Screenshot of FreeLing-Results (<http://nlp.lsi.upc.edu/freeling/demo/demo.php>) and a transcription of line 1-5

1	la	el	DA0FS0	DA	pos=determiner type=article gen=feminine num=singular	-	-	-	-	-	-	-	-	-	-
2	ilusión	ilusión	NCF5000	NC	pos=noun type=common gen=feminine num=singular	-	-	-	-	-	-	-	-	-	-
3	de	de	SP	SP	pos=adposition type=preposition	-	-	-	-	-	-	-	-	-	-
4	el	el	DA0MS0	DA	pos=determiner type=article gen=masculine num=singular	-	-	-	-	-	-	-	-	-	-
5	drama	drama	NCMS000	NC	pos=noun type=common gen=masculine num=singular	-	-	-	-	-	-	-	-	-	-

```

<document>
<wordcount>67</wordcount>
<putime>0.017303</putime>
<paragraph>
  <sentence id="1">
    <token begin="0" ctag="DA" end="2" form="la" gen="feminine" id="t1.1"
lemma="el" num="singular" pos="determiner" tag="DA0FS0" type="article"/>
    <morpho>
      <analysis ctag="DA" gen="feminine" lemma="el" num="singular"
pos="determiner" selected="1" tag="DA0FS0" type="article"/>
      <analysis case="accusative" ctag="PP" gen="feminine" lemma="lo"
num="singular" person="3" pos="pronoun" tag="PP3FSA0" type="personal"/>
      <analysis ctag="NC" gen="masculine" lemma="la" num="singular"
pos="noun" tag="NCMS000" type="common"/>
    </morpho>
  </sentence>
</paragraph>
</document>

```

Table 1: xml-document, result of POS-Tagging via FreeLing (detail for the first word “la”)

All this is to describe the document, the sentence (if we can call it one) and the first tiny word “la”. We now have metadata about the word “la” and in whole about the few verses of Picasso we analyzed by using POS-Tagging. Let’s assume, that we would have all poetry by Picasso in XML-format and already tagged, we could look out for the pattern of grammatical construction – even if each poem (some of them consist of more than 6000 words) is one long sentence. Questions could be answered like: How long are the sequences of noun preposition noun (probably added a DA (determiner, article)? Yet, by looking a bit deeper in the analysis of our few verses, we can already make some significant observations.

Sequences look here like: DA NC SP DA NC PR PO (see the example above). That means there are many long sequences without a single verb. Out of the 67 words in the cited verses, we find only 4 conjugated verbs (juega, hace, muere, va – all 3. person sing.) and 2 in gerund forms (nadando, buscando). At least this is what we understand as a human being reading the text. The machine shows us other possibilities of understanding the pure grammatical form and produces without intention an alienation effect.<sup>17</sup> Of course, we know that words can represent different grammatical types, but are usually clearly marked by their context. For example, the difference between “el juego” (the game) and “juego al fútbol” (I play soccer) is easily detectable by the article “el”.

Through the process of POS-Tagging, we get each word analyzed separately and shown all possible grammatical meanings. You can see it above in the simple example of “la”: tagged correctly as DAØFSØ: determiner, article, feminine, singular. But, in the following rows it shows also the other possible meanings: PP3FSAØ: personal pronoun, third person, feminine, singular, accusative, or another possible meaning: NCMSØØ: noun, common, masculine, singular.<sup>18</sup> All these alternative meanings are excluded for our human reader-eye because we cannot ignore context and its meaning and probably do not know all possible grammatical meanings at once. For Picasso and the poetics of Avant-garde, it would be real fun to be able to imagine every possible meaning of every single word at once.

The fact, that Picasso as an author of avant-garde poetry is getting nearly that far can be proved by POS-Tagging because it is his authorial style of constructing sequences with prepositions that produces failures in the functionality of the tool. I rather would not have recognized it because reading all the 11 pages of tags in the XML-document is not that amusing. But, when I was looking at the possible forms of verbs (because they are obviously rare as we see by reading the 67 words: 4 verbs, 2 gerund forms), I was surprised by the tag for “son”: VSIP3PØ: verb, semiauxiliary, indicative, present, 3. person, plural.

<sup>17</sup> This is exactly the same effect, Stephen Ramsay is speaking about in, *Reading machines: toward an algorithmic criticism, Topics in the digital humanities* (Urbana: University of Illinois Press, 2011).

<sup>18</sup> For all, who are like myself not Spanish native speaker: “la” has also the meaning of the “A”, a musical note. See “La = Cf. fa. 1. m. Sexta nota de la escala musical.” (*Diccionario de la lengua española*, Real Academia Española, <http://dle.rae.es/?id=MixNIFB|MixQLcB|MiZ5vEt>).

```

</token>
  <token begin="124" ctag="VSI" end="127" form="son" id="t1.29"
lemma="ser" mood="indicative" num="plural" person="3" pos="verb" tag="VSI3P0"
tense="present" type="semiauxiliary">
  <morpho>
    <analysis ctag="VSI" lemma="ser" mood="indicative" num="plural"
person="3" pos="verb" selected="1" tag="VSI3P0" tense="present"
type="semiauxiliary"/>
    <analysis ctag="NC" gen="masculine" lemma="son" num="singular"
pos="noun" tag="NCMS000" type="common"/>
  </morpho>
</token>

```

Table 2: xml-document, result of POS-Tagging via FreeLing (detail for the 27<sup>th</sup> word “son”) and transcription of the tag “VSI3PO VSI”

The context of the word “son” makes it very clear, that Picasso is talking about sound: “el son del canto de cisne del clarín”. The result of POS-Tagging is, that “son” means: 3. person plural of the verb “ser”, which is obviously wrong – even if we assume that Picasso is linking words in an alienated way. So, why did the POS-Tagging identify the word wrongly? Probably because Picasso’s grammatical construction, the sequence of noun, preposition, noun, etc. is not the usual way how language is used.<sup>19</sup>

Three consequences for a possible POS-Tagging-analyses of Picasso’s writing should be noted: First, the reliance of the tool cannot be guaranteed. Nobody would recognize failures like this one in about 60.000 words. Though, for an overview of the quantity of long sequences of preposition-noun-etc. the tool will certainly give reliable results. Second, the unexpected and very powerful result is the detection of these failures as evidence for Picasso’s outstanding style. Naturally, this could also be the effect of a failure in the programming of the tool which I could not test but other examples show it is not an exception (see Tab. 3). Third, the listed possibility of meaning due to the grammatical function of each word show us the uncertainty and vagueness of sense in Picasso’s writings. For this last point we also find an example in our few verses:

<sup>19</sup> The ambiguity of POS-Tagging in poetry is also verifiable in Spanish Golden Age sonnets, see Borja Navarro-Colorado, “A Metrical Scansion System for Fixed-Metre Spanish Poetry”, *Digital Scholarship in the Humanities* 33, no. 1 (1. April 2018): 112–27, <https://doi.org/10.1093/llc/fqx009>.

```

</token>
  <token begin="187" ctag="NC" end="191" form="buzo" gen="masculine"
  id="t1.44" lemma="buzo" num="singular" pos="noun" tag="NCMS000" type="common">
    <morpho>
      <analysis ctag="NC" gen="masculine" lemma="buzo" num="singular"
      pos="noun" selected="1" tag="NCMS000" type="common"/>
      <analysis ctag="VMI" lemma="buzar" mood="indicative" num="singular"
      person="1" pos="verb" tag="VMIP1S0" tense="present" type="main"/>
    </morpho>
  </token>
  <token begin="192" ctag="AQ" end="197" form="ciego" gen="masculine"
  id="t1.45" lemma="ciego" num="singular" pos="adjective" tag="AQOMS00"
  type="qualificative">
    <morpho>
      <analysis ctag="AQ" gen="masculine" lemma="ciego" num="singular"
      pos="adjective" selected="1" tag="AQOMS00" type="qualificative"/>
      <analysis ctag="VMI" lemma="cegar" mood="indicative" num="singular"
      person="1" pos="verb" tag="VMIP1S0" tense="present" type="main"/>
      <analysis ctag="NC" gen="masculine" lemma="ciego" num="singular"
      pos="noun" tag="NCMS000" type="common"/>
    </morpho>
  </token>

```

Table 3: xml-document, result of POS-Tagging via FreeLing (detail for the 44<sup>th</sup> and 45<sup>th</sup> words “buzo” and “ciego”) and transcription of the tags “NCMS000 NC”, “AQOMS00 AQ”

Reading the part of the cited verses: “nadando buzo ciego por las tripas...” we cannot be sure if a blind diver is swimming to find the green tripe or if some kind of narrator (yo buzo) is diving blind and at once swimming to find the green tripe. The word “buzo” can mean the first person singular of the verb “buzar” as well as the noun “el buzo”. In the context of the poem and the construction of the sentence (or sequence) we are simply not able to decide which one of the meaning should be favored. The POS-tagging gives us all possible meaning of each words – at least regarding its grammatical function. That Picasso is intentionally playing with the grammatical scope and indeterminateness is now not only an intuitive hypothesis while reading his text but can easily be argued with the help of POS-tagging.<sup>20</sup>

## Stylometric experiments for Góngora (and Picasso)

Stylometry is an important and increasing field in quantitative literary analysis.<sup>21</sup> It is one of the main subjects of this volume (see the contributions of Fradejas Rueda, Hernández Lorenzo, Calvo Tello, Lescasse). For the Spanish speaking world José Calvo Tello explained the functionality of Delta and

<sup>20</sup> From a linguistic point of view, the semantics of poetry are rarely analyzed and can also be a connection between close reading and digital methods, see Aurélie Herbelot, “The Semantics of Poetry: A Distributional Reading”, *Digital Scholarship in the Humanities* 30, no. 4 (Dezember 2015): 516–31, <https://doi.org/10.1093/llc/fqu035>.

<sup>21</sup> See the Special Interest Group (SIG-DLS) for Digital Literary Stylistics: <https://dls.hypotheses.org/> and the new priority programm for Computational Literary Studies of the DFG: [http://www.dfg.de/en/service/press/press\\_releases/2018/press\\_release\\_no\\_07/index.html](http://www.dfg.de/en/service/press/press_releases/2018/press_release_no_07/index.html)

Christof Schöch explained in the introduction into *Digital Humanities* quantitative literary analysis in its different methods.<sup>22</sup> Historically born out of authorship contribution and forensic linguistics,<sup>23</sup> stylometry is now used for different kind of literary research questions (re-writing literary history, changes of style within the works of one author, multilingual comparisons, genre classification, collaborative authorship). Instead of supporting the positivistic thesis that every author has an individual style, that can be proved like a fingerprint, more and more analysis in the literary context show that stylistic differences can be due to genre signals and others.

To compare Góngora and Picasso each of them to contemporary authors of approximately the same genre I used *stylo*<sup>24</sup>, a package for the statistical program R, that calculates the statistical value for the ‘stylistic’ distance for each text of the corpus to another (based on the frequencies in the whole corpus). That means, the result depends enormously on the corpus composition, on the chosen parameters (how many frequent words are taken into consideration, etc.) and on the chosen delta (Euclidian, cosine, Eder’s, Burrow’s, etc.).<sup>25</sup> The advantage of such a tool is the very transparent decision (all parameters are given in the visualization of the results) and that it is rather easy to understand – even if the mathematics behind it are quite complicated.

While conceptually straightforward, stylometry is rather advanced when it comes to its mathematical background; simultaneously comparing several frequencies of function words requires techniques which are referred to as

<sup>22</sup> José Calvo Taller, “Entendiendo Delta desde las Humanidades”, *Caracteres, Estudios culturales y críticos de la esfera digital*, 5(1): 140–76, <http://revistacaracteres.net/revista/vol5n1mayo2016/entendiendo-delta/>; Christof Schöch, “Quantitative Analyse”, in *Digital Humanities*, ed. by Fotis Jannidis, Hubertus Kohle and Malte Rehbein (Stuttgart: J.B. Metzler, 2017), 279–98, [https://doi.org/10.1007/978-3-476-05446-3\\_20](https://doi.org/10.1007/978-3-476-05446-3_20).

<sup>23</sup> For early spanish literature see Javier Blasco and Cristina Ruíz Urbón, “Evaluación y cuantificación de algunas técnicas de ‘Atribución de autoría’ en textos españoles”, *Castilla. Estudios de Literatura* (15. Oktober 2009): 27–47, <https://doi.org/10.24197/cel.o.2009.27-47>; Javier Blasco Pascual, Patricia Marín Cepeda and Cristina Ruiz Urbón, Ed., *Hos ego versiculos feci: estudios de atribución y plagio* (Madrid and Frankfurt am Main: Iberoamericana and Vervuert, 2010); Nanette Rißler-Pipka, “Digital Humanities und die Romanische Literaturwissenschaft: der Autorschaftsstreit um den *Lazarillo de Tormes*”, *Romanische Forschungen* 128, no. 3 (15. September 2016): 316–42.

<sup>24</sup> Maciej Eder, Mike Kestemont and Jan Rybicki, “Stylometry with R: a suite of tools”, in *Digital Humanities 2013. Conference Abstracts* (Lincoln: University of Nebraska-Lincoln, 2013), 487–89.

<sup>25</sup> For a list of all parameters, see Maciej Eder, Jan Rybicki and Mike Kestemont, “Stylo: A Package for Stylometric Analyses”, last modified June 2017, <https://goo.gl/oZ5bBh>.

multivariate, or multidimensional, because they involve multidimensional geometry to compute similarities between texts.<sup>26</sup>

In contrast to traditional stylistics, the choice of what should be the basis of comparison is not made by the researcher (for example the comparison of a very typical word or sequence) but is simply based on the MFW (Most Frequent Words). Of course, by changing the parameters and picking only the significant results, one can still manipulate the outcome. Yet, the changing of parameters has also the character of experimenting and if every decision is transparent and discussed it can be part of a positive research process.

In the era of digitalization, it is still not at all natural to find a digitized corpus with reliable texts and metadata. For the 17<sup>th</sup> century Borja Navarro provides a valuable resource of Spanish baroque sonnets, including Góngora.<sup>27</sup> The problem is, each sonnet is given as a single xml-file. While, for stylometry I need more words and txt-format (xml is also possible, but the real problem are the small single files with limited number of token (words)). That is why in the end I took the texts directly from the Biblioteca Virtual Miguel de Cervantes (BVMC) to fold the sonnets for each author into one single document, so we get files of 3000-5000 words per author. This is about the minimum we need to proceed with stylometry with R. What really could be a minimum of words used in *stylo* is discussed by Maciej Eder. He also points out, that short literary forms like sonnets or aphorisms in his example have a very genre driven style and it is problematic to treat many short texts of this kind as one long one.

Short literary forms are often masterpieces of concise language, with a domination of verbs over adjectives, particles and so on, with a proverbial witty

<sup>26</sup> James O'Sullivan, Katarzyna Bazarnik, Maciej Eder and Jan Rybicki, "Measuring Joycean Influences on Flann O'Brien", *Digital Studies = Le Champ Numérique* 8, no. 1 (27. März 2018), <https://doi.org/10.16995/dscn.288>.

<sup>27</sup> Borja Navarro, "A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects", in *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 2015, 105–113, <http://aclweb.org/anthology/W15-0712> and for the corpus itself: *Corpus of Spanish Golden-Age Sonnets*, ed. by Borja Navarro Colorado, María Ribes Lafoz and Noelia Sánchez, 2015. <https://github.com/bncolorado/CorpusSonetosSigloDeOro> : see also the review: José Calvo Tello, "Corpus of Spanish Golden-Age Sonnets", in *RIDE: a review journal for digital editions and resources* 6 (September 2017), <https://ride.i-d-e.de/issues/issue-6/corpus-of-spanish-golden-age-sonnets/>. The recent project Diachronic Spanish Sonnet Corpus (DISCO) enlarges this work by providing more than 3000 sonnets from 15<sup>th</sup>–19<sup>th</sup> century: Ruiz Fabo, Pablo, Helena Bermúdez Sabel, Clara Martínez Cantón and José Calvo Tello, *Diachronic Spanish Sonnet Corpus (DISCO)*, (Madrid: UNED, 2017). <https://github.com/pruizf/disco>.

style, and with a strong tendency to compression of content. Thus, a collection of aphorisms will certainly have different word frequencies than a long essay written by the same author<sup>28</sup>

The next problem is, that now we have one single ‘text’ per author and the typical dendrogram as a statistical output of *stylo* tries to cluster at least pairs of texts. Usually, it is better to have at least two texts by the same author to allow a correct clustering. So, if we have 7 authors (that was the number of authors with enough text, easily accessible via BVMC, and from the baroque period) one must be the outlier. And it is not surprisingly Góngora in the first experiment (fig.2<sup>29</sup>).

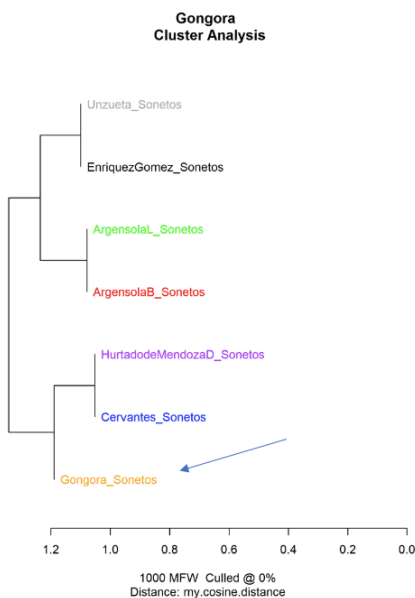


Figure 2: Dendrogram of the Corpus of Spanish Golden Age Sonnets (1000 MFW, cosine Delta), arrow added to point out Góngora

<sup>28</sup> Maciej Eder, “Does size matter? Authorship attribution, short samples, big problem”.

<sup>29</sup> The result didn’t change when changing the parameters from 100-1000 MFW or using Eder’s distance, but as the cosine distance was tested to perform best with literary texts in various languages, I used only cosine in the following experiments, see: Fotis Jannidis et al., “Improving Burrows’ Delta: an empirical evaluation of text distance measures”, in *Digital Humanities 2015: Conference Abstracts* (Sydney, 2015), <http://dh2015.org/abstracts/index.php>.



Looking out for more texts to enlarge the corpus, I added 25 authors to have an even number of 32 in total to pair the authors in the dendrogram. But, still there is only one text per author (see fig. 3). The choice of the additional authors was made by the help of Borja Navarro's collection of Golden Age sonnets and the precondition of finding enough sonnets (at least 40) in BVMC. Furthermore, the difference between the number of words for each text (collection of sonnets) varies between 1850 (Tassis y Peralta) and 13061 (Lope de Vega). Another problem is the wide range of age: Some of the authors are more than hundred years younger/older than others (therefore I highlighted the dates of authors who died in 16<sup>th</sup> century in the table 4).

Table 4: Table of used corpus

Author	Number of words	Biographical data	File-name
Hernando de Acuña	4093	<u>1518–1580</u>	AcunyaF_Sonetos
Francisco de Aldana	4497	<u>1537–1578</u>	AldanaF_Sonetos
Juan de Almeida	4292	<u>1530–1573</u>	AlmeidaJ_Sonetos
Bartolomé Leonardo de Argensola	3334	1562–1631	ArgensolaB_Sonetos
Lupercio Leonardo de Argensola	2830	1559–1613	ArgensolaL_Sonetos
Francisco de Borja y Aragón	2935	1577–1658	BorjayAragonF_Sonetos
Juan Boscán Almagóver	4975	<u>1492–1542</u>	BoscánAlmagóverJ_Sonetos
Luis Carrillo y Sotomayor	2551	1582–1610	CarrilloSotomayorL_Sonetos
Miguel de Cervantes Saavedra	3604	1547–1616	Cervantes_Sonetos
Gutierre de Cetina	3262	<u>1510–1554</u>	CetinaG_Sonetos
Antonio Enriquez Gómez	4216	1600–1661	EnriquezGomez_Sonetos
Francisco de Trillo y Figueroa	4871	1618–1680	FigueroaF_Sonetos
Garcilaso de la Vega	3787	<u>1501–1536</u>	GarcilasodelaVega_Sonetos
Luis de Góngora y Argote	2936	1561–1627	Gongora_Sonetos
Fernando de Herrera	6165	<u>1534–1597</u>	HerreraF_Sonetos
Diego Hurtado de Mendoza	5361	<u>1503–1575</u>	HurtadodeMendozaD_Sonetos
Fray Luis de León	6330	<u>1527–1591</u>	LeonFrayLuis_Sonetos+Eglogas
Joseph de Litala y Castelví	4891	17 <sup>th</sup> century	LitalayCastelviJ_Sonetos
Lope de Vega	<u>13061</u>	1562–1632	Lopedevga_Sonetos
Francisco de Medrano	5188	1570–1607	MedranoF_Sonetos

Author	Number of words	Biographical data	File-name
Pedro de Padilla	4245	<u>1540–1595</u>	PadillaP_Sonetos
Bernardino de Rebolledo	2418	1597–1674	RebolledoB_Sonetos
Agustín de Salazar y Torres	3342	1642–1675	SalazarYTorresA_Sonetos
Sor Juana Inés de la Cruz	3430	1648–1695	SorJuanaInesdelaCruz_Sonetos
Pedro Soto de Rojas	3635	1584–1658	SotodeRojasP_Sonetos
Juan de Tassis y Peralta	1850	1582–1622	TassisyPeraltaJ_Sonetos
Juan Timoneda	2688	<u>1520–1583</u>	TimonedaJ_Sonetos
Tirso de Molina	5615	1579–1648	TirsodeMolina_Sonetos
Francisco de la Torre	3723	<u>1521–1582</u>	TorreF_Sonetos
Luis de Ulloa Pereira	3478	1584–1674	UlloaPereiraL_Sonetos
Gabriel Bocángel y Unzueta	3965	1603–1658	Unzueta_Sonetos
Diego Ximénez Ayllón	2660	<u>1530–1590</u>	XimenezAyllon_Sonetos

In the comparison of 32 authors Góngora is not at all an outlier (at least not in the genre of sonnets). He is prominently clustered with Lope de Vega, and on a sidetrack with the rather unknown Joseph (or José) de Litala y Castelvi (a knight in the Order of Calatrava).<sup>30</sup> The clustering of Lope and Góngora is not that surprising even if they were literary enemies. Both wrote sonnets as satire to fight on the field of literature against each other. The different genres of sonnets (love, satire, religion, honor) are also a problem for a stylistic comparison via MFW.<sup>31</sup> More important than this aspect seems to be the distance of different periods: all authors on the lower arm of the dendrogram are 16<sup>th</sup> century authors. Only Francisco de Aldana and Fray Luis de León (clustered together), Francisco de la Torre and Fernando de Herrera (clustered together) are clustered on the upper arm of the dendrogram and died already at the end of 16<sup>th</sup> century. This phenomenon can be observed all along the following experiments (also in the series of 100–1000 MFW which

<sup>30</sup> In 1941 Homero Serís is trying to establish José Delitala y Castelvi as “un clásico olvidado”, obviously without much success. Homero Serís, “Libro raro y curioso. Poesías de José Delitala y Castelvi (1672). Un clásico olvidado”, *Bulletin Hispanique* 43, no. 2 (1941): 171–81, <https://doi.org/10.3406/hispa.1941.2908>.

<sup>31</sup> As all sonnets are treated as one long text, the different genres are mixed together in one text. Therefore, also in the sonnet-collection of Góngora and Lope de Vega not only the satire sonnets are compared.

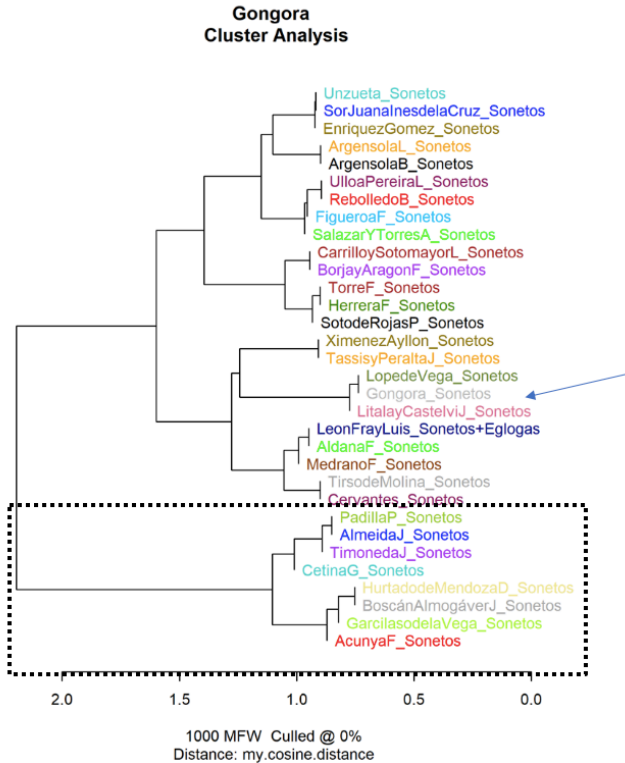


Figure 3: Dendrogram for an enlarged corpus for 32 authors of Golden Age sonnets, same parameters, arrow added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets

are not discussed here) and therefore is very consistent under different parameters.

For the next experiment, I changed the corpus another time: deleted Fray Luis de León, because he did not write enough sonnets (before I added his Eglogías to get the minimum number of words) and replaced him with Francisco de Quevedo, because he is also very well known for the satire sonnets in the literary battle between Góngora and other poets. Another important change is to split the sonnets for each author in at least two texts (of 1000–3000 words). The results now show the problems of the corpus and probably the difficulty of treating sonnets as one long text. In the series of 100–1000 MFW the clustering was not consistent and never could cluster all authors

correctly. The example of 1000 MFW here is picked out to have the same parameters as before (fig. 2–4).

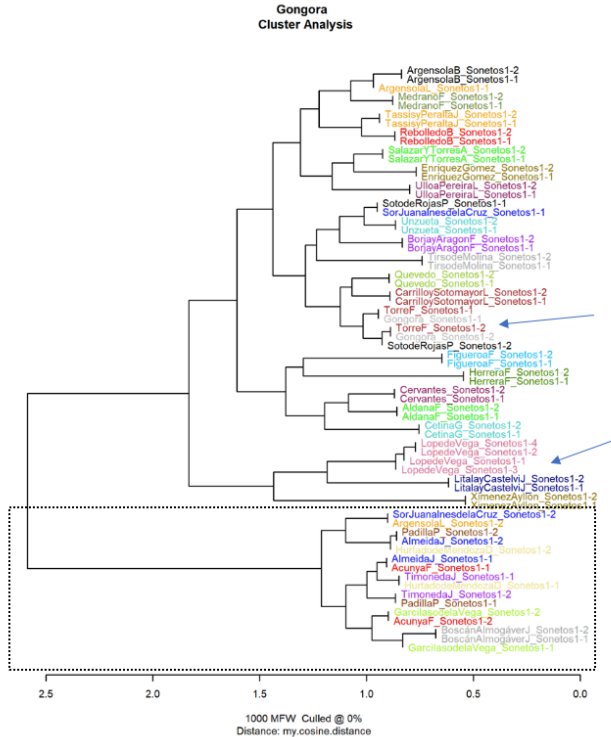


Figure 4: Dendrogram for sonnet-corpus split in at least two texts, same parameters; arrows added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets

Góngora is now split up and clustered with Francisco de la Torre and Pedro Soto de Rojas but sharing the arm of the dendrogram also with Quevedo (here similarities in style are known). In the whole picture, most of the authors are clustered correctly.<sup>32</sup> For example, Lope de Vega was split up in four texts (because of the enormous number of words) and all four found together on the same arm with the unknown Litala y Castelevi as before (fig. 2 and 4).

<sup>32</sup> “Correctly” means here: all parts of sonnets by the same author are clustered on one single arm of the dendrogram together.

Obviously wrong is the clustering of Sor Juana Inés de la Cruz and Lupericio Leonardo de Argensola, because one part of their sonnets was clustered in the period of 16<sup>th</sup> century authors and the other as before in the younger group. All in all, the 16<sup>th</sup> century arm of the dendrogram with Garcilaso de la Vega now is still detectable as a different style, but in itself only Juan Boscán Almagáver was clustered correctly.

This last experiment leads to two conclusions and further experiments: First, if the clustering of authorship basically works – even for a complex source like poetry – it might tell us something about stylistic difference of each author to another. Second, the inconsistencies in the corpus are probably due to the wide range of period and size (number of words). That means if we delete all texts of the 16<sup>th</sup> century and the ones less than 1000 words the results should be consistent. Before, we should have a look in the series of experiments *stylol* did already from 100 to 1000 MFW. If the number of words is part of the problem, with less MFW the performance is probably more consistent. But, as we see this is only true for the 17<sup>th</sup> century texts (fig. 5).

Now only 4 authors are not clustered correctly (in fig. 4 with 1000 MFW 10 authors are not clustered correctly) and Góngora is back on the arm of the dendrogram with Lope de Vega and Litala y Castelevi. This reappearance in different experiments gives more evidence for a stylistic similarity between the authors. But, it tells us nothing about the question if Góngora is the stylistic outlier who invented a new language. Also, in the following experiment, when we delete all sonnets from authors died before the 17<sup>th</sup> century and with less than 1000 words Góngora seems to be less an outlier than expected (fig. 6).

Now the outliers are Lope de Vega and Litala y Castelevi (both before clustered with Góngora). Apart of Sor Juana Inés de la Cruz and Lupericio Leonardo de Argensola all split sonnet texts are clustered correctly with their authors. Both, Sor Juana and L. Argensola were also in the first experiment with split sonnet-corpus wrongly clustered (fig. 4–5). This is probably due to the extremely mixed themes in their sonnets: Argensola's sonnets in the collection 1-1 are more about history and heroes and the collection 1-2 has religious subjects. For Sor Juana, we know that of course, many sonnets have religious subjects but are silently mixed with love-themes. In the collection of sonnets 1-2 of both authors (here clustered together) the word "pecho" (sin) is one of the most used content words. But, statistics reveal that it is

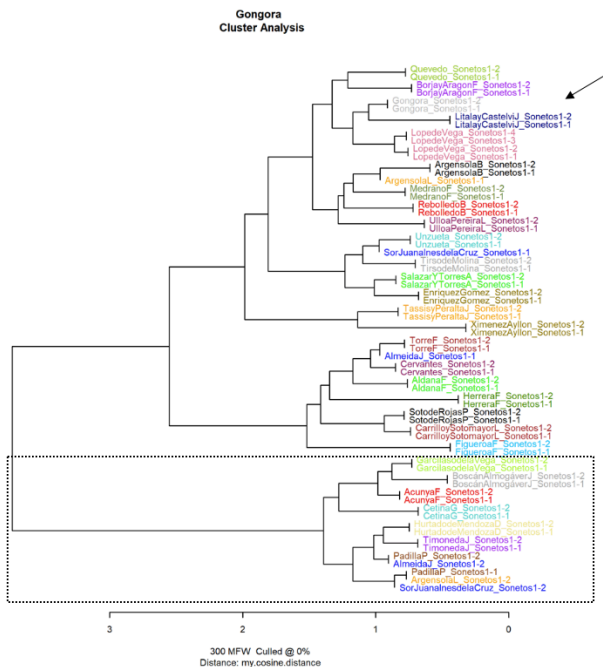


Figure 5: Dendrogram of the split sonnet-corpus (cf. fig. 4), only 300 MFW, arrow added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets

still used only 5 times out of approximately 1500 words.<sup>33</sup> Plus, we know that *stylo* takes into consideration the MFW of the whole corpus and here “pecho” is ranked on 85.<sup>34</sup>

All these experiments give us valuable insight about the literary history and a comparison of more authors than the usual canon. Proceeding with new experiments or following the surprisingly new path of Sor Juana Inés de la Cruz and the older Argensola brother would be another study. A result of the stylometric experiments until now is: Góngora is not the expected outlier

<sup>33</sup> For the calculation of words, I used *voyant-tools*: see for each corpus, Argensola (<http://voyant-tools.org/?corpus=b96dd2492f53e3ed1a6d5c6663c32159>) and Sor Juana (<http://voyant-tools.org/?corpus=52bc60da5092b9ed8e707a8585a427d2>). Sinclair, Stéfán and Geoffrey Rockwell, 2016. Voyant Tools. Web. <http://voyant-tools.org/>.

<sup>34</sup> To give an overview of the word list for the last experiment: first content word is “amor” (22.), followed by “vida” (43.), “sol” (45.), “alma” (48.), “cielo” (51.) – all of them could be used in every genre of sonnet: love, religion, satire, history.

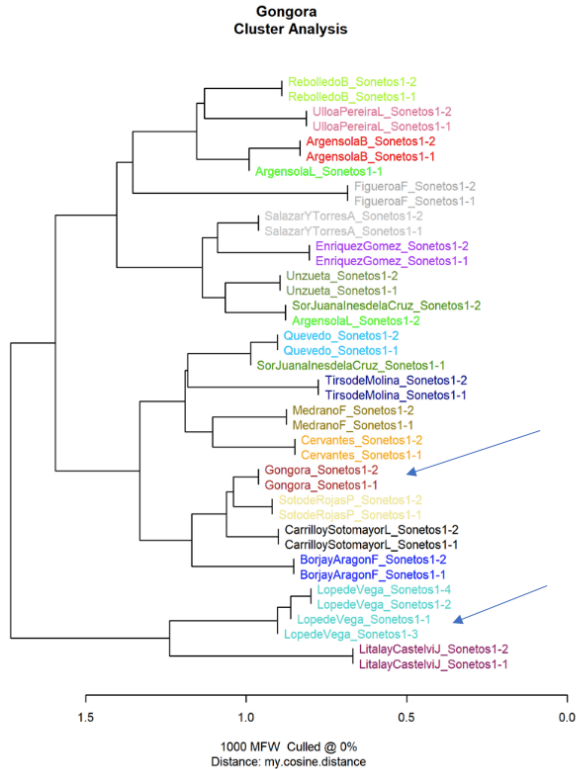


Figure 6: Dendrogram of the split sonnet-corpus, without 16<sup>th</sup> century-authors and Tassis y Peralta (less than 1000 words), arrows added to point out clustering of Góngora, Lope and Litala.

in the genre of sonnets and for contemporary authors. His role as a model and rebel for avantgarde literature is constructed by Lorca, the 27<sup>th</sup> generation and others<sup>35</sup> but defeated in recent studies, for example by Mercedes Blanco:

La difficulté que nous éprouvons (ou devrions éprouver) à pointer tel trait poétique comme étant spécifiquement gongorin a en effet un caractère structural : cette œuvre est tout le contraire d'une météorite ou d'un OVNI, et son

<sup>35</sup> See Federico Garcia Lorca, "La imagen poetica de Don Luis de Góngora", in *Obras completas*, Vol. 3 (Madrid: Aguilar, 1986), 223–47; Hans-Ulrich Gumbrecht, "Warum gerade Gongora? Poetologie und historiales Bewusstsein in Spanien zwischen Jahrhundertwende und Bürgerkrieg" in *Lyrik und Malerei der Avantgarde*, ed. by Rainer Warning and Winfried Wehle (München: Fink, 1982), 145–92; Elsa Dehennin, *La résurgence de Góngora et la génération poétique de 1927* (Paris: Didier, 1962).

auteur l'opposé d'un solitaire et d'un extravagant. Góngora, en phase avec son temps, à l'aise dans son monde, offre des solutions convaincantes bien qu'étonnantes à une recherche où tous sont embarqués, dans ce moment d'essor inégalé de la poésie espagnole qui se place entre les dernières décennies du XVI<sup>e</sup> et les premières du XVII<sup>e</sup> siècle.<sup>36</sup>

The presented stylometric experiments support the thesis of Blanco, that Góngora was not at all the bizarre author for which he was admired by the early Spanish avantgarde and other European modern authors like Picasso. Though, she does not contradict the assumption that Góngora tries to invent a new language. Instead, she tries to replace him in his period and points out his predecessors as well as his imitators.<sup>37</sup>

In the various recent studies on Góngora using digital methods or precisely stylometry we can see the new perspectives, but also its limits. As Blanco rightly points out, there exists no tool to explain the difficulty in understanding grammatical constructions like the hyperbaton or metaphors, poetic images, etc.

En outre ce n'est pas par ce biais qu'il sera le plus facile de se servir des humanités numériques. Les logiciels de stylométrie sont pour l'heure basés sur le lexique et non sur la syntaxe, de sorte que le repérage et la quantification des entorses à l'ordre des mots standard (hyperbates), que l'on tient depuis toujours pour typiques du style de Góngora, semblent actuellement hors de la portée d'une observation automatisée.<sup>38</sup>

Here, we have to contradict her today by citing the work of her own scholar: Marie-Eglantine Lescasse is analyzing not only lexis but also syntax, using POS-Tagging combined with statistical tests and hermeneutic analysis.<sup>39</sup> Staying on the semantic level of lexis it is also possible to compare Góngora to his contemporaries. Antonio Rojas Castro presented in two recent studies the comparison of key-words in a corpus of 16-17<sup>th</sup> century *Fábulas* (fables) and for Góngora alone in three different periods (1580–1610, 1611–1618,

<sup>36</sup> Mercedes Blanco, "Sor Juana Inés et le programme Phœbus : tester sur le gongorisme un logiciel d'exploration de l'intertextualité", *e-Spania. Revue interdisciplinaire d'études hispaniques médiévales et modernes*, no. 29 (1. Feb. 2018), <https://doi.org/10.4000/e-spania.27677>.

<sup>37</sup> As the author of the study *Góngora, o, La invención de una lengua* Mercedes Blanco is trying to establish the grammatical and semantic new paths Góngora is taking as a model for other Spanish Golden Age authors. See for example her chapter on sonnets: 24–30.

<sup>38</sup> Blanco, "Sor Juana Inés et le programme".

<sup>39</sup> See Marie-Eglantine Lescasse, "Góngora hors norme? Étude stylométrique d'un motif gongorin", p. 91 in this volume.



1619–1625).<sup>40</sup> Arguing on the level of content words involves always more hermeneutic interpretation than arguing with the MFW. The conclusion, that out of the keywords it might be possible to define precisely the esthetics behind the poetry (“de identificar dos dominantes estéticas bien claras: por un lado, el registro cortesano; por el otro, el registro culto y sublime”)<sup>41</sup> seems difficult, thinking about the various meanings of words in different contexts of poetry.<sup>42</sup> Though, it is not at all a “danger” of interpretation but a necessary part of stylometric analysis (“On remarque un saut périlleux entre les données qu’il obtient des logiciels et l’interprétation, séduisante, qu’il en propose.”)<sup>43</sup> At this point we are back on literary criticism, when out of the different methods, different results are discussed and interpreted. To participate in the debate of literary criticism, to combine and to compare stylistics and stylometry is one of the biggest challenge for digital Humanists, because there are two parties to convince: the one of statistical evidence (using the proper method?) and the one of hermeneutics (does the interpretation in the end have some impact in the literary debate?).

Stating that the results of stylometry contradict my hypothesis, that Góngora is an outlier of his period and at the same time support the thesis of Blanco and others, that Góngora is well established in his period, we have to change the research question: Góngora (and Picasso) try to invent a new language – yes, but at least Góngora (and probably Picasso too) are not alone on this project.

It is not possible to test this in the same way with Picasso’s writings. Simply, because I could not find enough Spanish prose poems for the 20<sup>th</sup> cen-

---

<sup>40</sup> Antonio Rojas Castro, “Luis de Góngora y la fábula mitológica del Siglo de Oro: clasificación de textos y análisis léxico con métodos informáticos”, *Studia Aurea* 11 (22. Dec 2017): 111–42, <https://doi.org/10.5565/rev/studiaaurea.260>; Antonio Rojas Castro, “¿Cuántos “Góngoras” podemos leer? Un análisis contrastivo de la poesía de Luis de Góngora”, *e-Spania. Revue interdisciplinaire d’études hispaniques médiévales et modernes*, no. 29 (1. Feb. 2018), <https://doi.org/10.4000/e-spania.27448>.

<sup>41</sup> Rojas Castro, “¿Cuántos “Góngoras” podemos leer?”.

<sup>42</sup> As Blanco and Navarro point out, it is probably more important to know hows the used words are ingeniously combined, than to know the words themselves, see Navarro, “A computational linguistic approach to Spanish Golden Age Sonnets”, 112.

<sup>43</sup> Mercedes Blanco about the study of Rojas Castro in her article “Sor Juana Inés et le programme”.

tury.<sup>44</sup> The experiments I made show rather an effect of genre clustering than of authorial style (fig. 7).

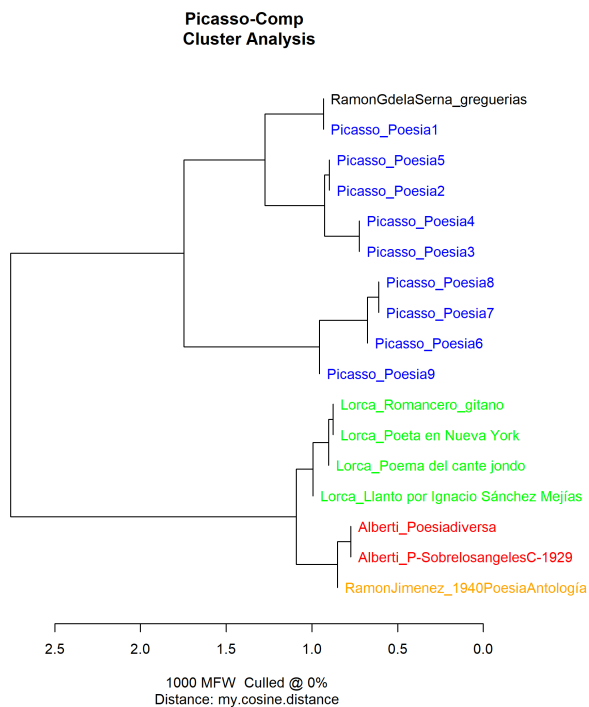


Figure 7: Dendrogram for the Picasso-corpus, generated with *stylo*

Here, I split the prose poems Picasso wrote in the years of 1935–1957 into 9 texts to have approximately the similar number of words as for the anthologies of poetry by Federico Garcia Lorca, Rafael Alberti, Juan Ramón Jiménez and Ramón Gómez de la Serna. Again, we can state that the attribution of authorship via *stylo* is correct and that prose (poems) and verse poetry is clustered together (genre signal). But, apart from those rudimentary observations every interpretation of style is not based on reliable statistic results. The

<sup>44</sup> There are anthologies of 20<sup>th</sup> century poetry in the BVMC, but most of them are verse poems, see: José Luis García Martín, ed., *Poetas del Novecientos: entre el Modernismo y la Vanguardia (Antología)*. Tomo I + II, edición digital basada en la de [Madrid, Fundación Santander Central Hispano, 2001] (Alicante: Biblioteca Virtual Miguel de Cervantes, 2004), <http://www.cervantesvirtual.com/obra-visor/poetas-del-novecientos-entre-el-modernismo-y-la-vanguardia-antologia-tomo-i-de-fernando-fortun-a-rafael-porlan--o/html/>.

number of authors compared is far too small for concluding anything about Picasso's standing in literary history. Plus, the genre difference between Picasso's prose poems, the modern verse poems of Lorca, Alberti and Ramón Jiménez and the *greguerías* by Gómez de la Serna influences the statistics far too much. In a direct hermeneutic comparison, we could prove similarities between Lorca, Alberti, Picasso and others.<sup>45</sup> But, they are due to topics, humour and for example the adoration for Góngora which is hard to measure in stylometry. Let me show with a very simple example how complex the relations via vocabulary and topics can be: Comparing three citations of Góngora, Picasso and Gómez de la Serna, we see that they are somehow talking about a swan, but all three in a completely different way and style:

GÓNGORA: <i>Soledades</i> , Soledad II, V. 799–805 (1618) <sup>46</sup>	PICASSO: <i>Écrits</i> , 02.02.1936 <sup>47</sup>	GÓMEZ DE LA SERNA: <i>Greguerías</i> (Madrid 1936) <sup>48</sup>
Can de lanas prolijo, que animoso <u>buzo</u> será, bien de profunda <u>ría</u> , bien de serena playa, cuando la fulminada prisión caya del Nebli, a cuyo vuelo tan vecino a su <u>cielo</u> el <u>Cisne</u> perdonara, <u>luminoso</u>	... del son del canto de <u>cisne</u> del clarín que se muere de pena nadando <u>buzo ciego</u> por las tripas verdes del <u>lago</u> por donde Ofelia va buscando en el fondo un trozo de <u>papel</u> y un <u>lápiz</u> ...	La luna ilumina la cifra de almanaque de los <u>cisnes</u> .  Cuando el cielo se haya jugado todas las estrellas se jugará la Luna a cara o cruz.  En la Luna se han visto revolver papeles de un picnic antediluviano.

Table 5: Examples of Close Reading for Góngora, Picasso, Ramón Gómez de la Serna

The 'danger' of distant reading using stylometry and other digital methods is not at all the part of interpretation, but to be careful not to compare apples and oranges. In the previous experiments we have seen how much the composition of the corpus influences the performance of the clustering on the whole. Not until experiments with different parameters show the same effect, we can speak about reliable results. That is why I did not proceed with the stylometry on Picasso, because the corpus for comparison appears to be too poor for further analysis.

<sup>45</sup> See Rifler-Pipka, *Picassos schriftstellerisches Werk*, 238–60.

<sup>46</sup> Luis de Góngora y Argote, *Soledades*, Soledad II, V. 799–805, ed. by John Beverley (Madrid: Catedra, 142009), 158.

<sup>47</sup> Picasso, *Écrits*, 100–1.

<sup>48</sup> Ramón Gómez de la Serna, "Greguerías", *ABC (Madrid)* (22.07.1969): 96, <http://hemeroteca.abc.es/nav/Navigate.exe/hemeroteca/madrid/abc/1969/07/22/096.html>.

## Trying Topic Modeling on Picasso's writings in place of a conclusion

Topics are extremely important for Picasso's writings. Reading his texts from 1935 to 1957 the reappearing topics are probably the only stable and 'readable' in terms of 'understanding' that serves us as Ariadne's thread out of the labyrinth. Marie-Laure Bernadac tried even to make a dictionary or vocabulary for his writing.<sup>49</sup> But, to reduce and to explain Picasso's poetry by his topics simply cannot be a solution. It would deprive us from the pleasure of being lost in his word-labyrinth, of the humorous non-sense of his sequences (of words, numbers, signs), of the synesthetic perception and so on. Nearly every entry in Bernadac's dictionary is an ineligible reduction of sense:

écriture [...] L'écriture est au départ, en 1935, un recours quasi thérapeutique qui lui permet de surmonter sa dépression [...] éléments [...] enface [...] ennui<sup>50</sup>

Difficult not to suspect an interpretation motivated one-sided by Picasso's biography as a celebrated artist and painter, reducing all the effort and literary power of his poetry to a psycho-analytic pathologic explanation. The raising question and simple falsification of this kind of interpretation is to detect Picasso's topics by using Topic Modeling on his writings without preliminary assumptions and without interpretation of the topic itself. The advantage of the method is its absolutely blindness regarding the person of the author (and his possible psychological problems). The disadvantage is that the methods and the model behind the method naturally assumes common texts to be analyzed. When John Firths explains distributional semantics as "a word is characterized by the company it keeps" – it is correct for the usual context and language.<sup>51</sup> But the poet, who intentionally changes the usual semantic context, probably destroys also the functionality of distributional semantics. Exactly this failure of the method can be an advantage for the literary scholar as Lisa Rhody points out:

Topic modeling as a methodology, particularly in the case of non-figurative language texts like poetry, can help us to get to new questions and discoveries

<sup>49</sup> Marie-Laure Bernadac, "The Poetry of Picasso (with an Abridged Dictionary)", in *Écrits*, ed. by Christine Piot and Marie-Laure Bernadac (Paris: Gallimard, 1989), XIII-XXV.

<sup>50</sup> Bernadac, "The Poetry of Picasso", XIX.

<sup>51</sup> See for the reference of John Firth and an introduction into Topic Modeling using french examples: Christof Schöch, "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama", *Digital Humanities Quarterly* 011, no. 2 (22. May 2017).

– not because topic modeling works perfectly, but because poetry causes it to fail in ways that are potentially productive for literary scholars.<sup>52</sup>

That means, we are able to test two aspects by trying Topic Modeling on Picasso's writings: First, to show the interpretation of topics by Bernadac are rather driven by a biographical look on the texts than by an analysis of the text itself. Second, to show that Topic Modeling probably is going to fail as a method because it is created to detect topics in common text collections. A Topic is built by the statistical comparison of the use of words in a defined paragraph of text. That means all words of the corpus are compared to all other words. Then, statistics counts how often words appear together in the same paragraph (context). This is an extensive process because every single word has to be calculated in relation to every other word in the corpus and in relation to the distribution of words all over the corpus.<sup>53</sup> It is easy to imagine that words like "doctor, pharmacy, medicine, illness, temperature, etc". are more used together in one paragraph than mixed up with the word "landscape". In some corpuses these words belonging to the same lexical field are probably mixed with words like "child, night, car" or other terms that describe the situation in which the topic is used in this specific corpus.<sup>54</sup> Important to state though, that the results out of Topic Modeling are not necessarily the same as literary topics.

This test is far from being seriously using Topic Modeling as a method to really analyze topics represented in the text collection of Picasso. But, it is rather an 'experiment' to test if the algorithm 'reads' Picasso's poetry differ-

<sup>52</sup> Lisa M. Rhody, "Topic Modeling and Figurative Language", *Journal of Digital Humanities* Vol. 2, no. 1 (2012). <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/>.

<sup>53</sup> For a very clear explanation of Topic Modeling for Spanish examples, see recently Borja Navarro-Colorado, "On Poetic Topic Modeling: Extracting Themes and Motifs From a Corpus of Spanish Poetry", *Frontiers in Digital Humanities* 5 (2018). <https://doi.org/10.3389/fdigh.2018.00015>.

<sup>54</sup> For prose and drama in most cases we can get sensible results, for example see the analysis by Ulrike Henny u. a., "Topic, Genre, Text: Topics im Textverlauf von Untergattungen des spanischen und hispanoamerikanischen Romans (1880–1930)", in *Jahrestagung des DHd-Verbands 2016: Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma* (dhd2016, Leipzig, 2016), <http://www.dhd2016.de/abstracts/vorträge-055.html>; Nanette Rifler-Pipka und Christof Schöch, "Topic Modeling als Perspektive auf das spanische und französische Theater des 17. Jahrhunderts", in *Von Daten zu Erkenntnissen: Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation, Jahrestagung des DHd-Verbands* (dhd2015, Graz, 2015), 20, <http://gams.uni-graz.at/o:dhd2015.v.015>.

ently than a literary scholar like Bernadac (or myself). For a serious analysis I would have needed much more texts and probably a text collection for comparison. But, as a first step and thanks to the new Dariah Topics Explorer,<sup>55</sup> I tried my small collection of Picasso’s Spanish texts.

Corpus size, in documents	174
Corpus size (raw), in tokens	48321
Corpus size (clean), in tokens	13623
Size of vocabulary, in tokens	1213
Number of topics	30
Number of iterations	5000
The model log-likelihood	-97656
Passed time, in minutes	3

As you can see, your corpus is much smaller after cleaning. You either defined a threshold for most frequent words, or selected an external stopwords list. In addition so-called *hapax legomena* have been removed. In corpus linguistics, a *hapax legomenon* is a word that occurs only once within a context. So, if a word occurs only once in a document, it is very likely that the word is semantically insignificant – meaning not useful for the topic modeling algorithm.

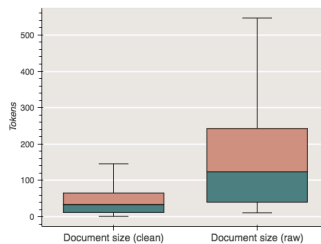


Figure 8: Screenshot: Result of DARIAH Topics Explorer: corpus and document size

One of the first results you get after using the tool is a description of the input (corpus), parameter and the difference between the document size with and without stop-words (functional words). Here, most of the words are functional words which is typical for prose, not necessarily for poetry. Further steps for deeper analysis could be the comparison to other corpora (prose or poem or prose-poems) of the same period and language.

The next result, that is the most interesting, is the table of topics for the corpus (defined before by number, here: 30, see fig. 4). For comparison, I tried a well-prepared corpus of Hispanic novels and it seems easy to detect topics like “war” or “family”, “rural life”, “high society”, “church”, “virtue” etc. (fig. 9).

When I first tried the tool with the above given parameters and texts, the results were rather frustrating, because they did not much vary from the original texts. They did not form topics properly as shown for the corpus of hispanic novels of the 19<sup>th</sup> century (fig. 9):

<sup>55</sup> See Steffen Pielström et al., <https://dariah-de.github.io/TopicsExplorer/>.

	Key 1	Key 2	Key 3	Key 4	Key 5	Key 6	Key 7	Key 8
<b>Topic 1</b>	paisano	partida	perro	compadre	toldo	cuero	especie	animal
<b>Topic 2</b>	fondo	pared	carne	papá	ocasión	valle	mozo	espalda
<b>Topic 3</b>	habitación	recuerdo	ángel	placer	sonrisa	sombra	presencia	rayo
<b>Topic 4</b>	mamá	chico	baile	negro	libro	patio	corredor	mozo
<b>Topic 5</b>	dama	convento	negro	fraile	cárcel	merced	justicia	palacio
<b>Topic 6</b>	ejército	tropa	guerra	ministro	emperador	comandante	presidente	oficial
<b>Topic 7</b>	virtud	cargo	naturaleza	do	concepto	término	impulso	intento
<b>Topic 8</b>	sociedad	salón	coche	fortuna	época	teatro	país	minuto
<b>Topic 9</b>	hacienda	cura	favor	juez	conde	interés	ropa	marqués
<b>Topic 10</b>	español	sacerdote	anciano	dios	piedra	país	bosque	rey

Figure 9: Screenshot: Result of DARIAH Topics Explorer: Topics (here limited to 10) in the corpus of CLIGS-textbox, Hispanic novels of the 19<sup>th</sup> century<sup>56</sup>

Obviously, the Picasso-corpus was not that well prepared as the one cited above in comparison for Hispanic novels. We have a French text or parts of a French texts mixed up in the corpus (which is not that surprising because Picasso wrote some texts in both languages) (see Topic 6 in fig. 10). The elimination of stop-words did not work perfectly, because there are still words like “te”, “tú”, etc. Nevertheless, there are topics making sense in the context of Picasso’s writings: for example, Topic 7 (corrida): “colores, caballo, abre, manos, toro, vientre, enamorado, bombo”. But to be serious, we have to state, that the corrida-motif is spread out all over the topics in words like “traje” (for “traje de luces”, the costume the torero wears), “aplausos” (in the arena de toros), even the innocent word “tarde” (afternoon) is strictly used in terms of the afternoon when the corrida begins. Looking at the topic of color and its lexical field it is even more striking that there is no real topic to identify (even Topic 11 contains only 3 color words), but it is spread all over the topics (see the highlighted words in fig. 10).

These, not very enlightening results, may lead to two conclusions. First, the tool did not work, because the parameters and input (not enough texts) were inappropriate. Or, second, that Picasso’s style is refusing the rules of semantical distribution, because he refuses semantics at all. In Topic Modeling – as in most tools – there is always the danger that the observed effect is due to other reasons than we thought.

In practice, collections are often constructed by combining documents from multiple sources, which may have distinctive style and vocabulary. This heterogeneity of sources leads to a serious but rarely studied problem: the

	Key 1	Key 2	Key 3	Key 4	Key 5	Key 6	Key 7	Key 8
Topic 1	color	cuchillo	tira	lleva	aun	ciego	necesario	sentado
Topic 2	cara	oro	cuentas	día	sol	higos	puestas	tejado
Topic 3	blanco	puesta	trapo	abajo	traje	sopa	punta	máquina
Topic 4	sombra	mar	medio	lluvia	corazón	canta	deja	pico
Topic 5	azul	plaza	gota	mañana	brazo	recoge	arriba	mira
Topic 6	et	du	qui	sa	ce	dans	au	par
Topic 7	colores	caballo	abre	manos	toro	vientre	enamorado	bombo
Topic 8	color	leche	grano	lleno	oreja	peso	cal	cuelga
Topic 9	tarde	medio	debajo	hechos	echando	cartero	niñas	caracoles
Topic 10	canta	seda	vestido	tiempo	olor	jamón	puerta	sartén
Topic 11	verde	azul	dedos	amarillo	medio	piel	alas	cielo
Topic 12	aceite	pared	papel	agua	plata	estrellas	hilos	nubes
Topic 13	silencio	mano	fondo	nido	verano	hora	canto	arde
Topic 14	iris	cristal	oro	arco	aire	escalera	deseo	cuerpo

Figure 10: Screenshot: Result of DARIAH Topics Explorer: Topics for Picasso-corpus as cited in fig. 8 – here cut by topic number 14 (of 30). Highlighted color words by N. R.-P.

strongest, most prominent patterns in a collection may simply repeat the known structure of the corpus. Instead of finding informative, cross-cutting themes, models simply repeat the distinctive vocabulary of the individual sources.<sup>57</sup>

To find out if it is due to Picasso’s individual distinctive style or a limitation of tokens, I increased the number of words artificially by quadruplicating Picasso’s poems. Now the parameters are 20 topics, and 100 iterations (instead of 30 topics and 5000 iterations before) and the raw corpus consists of 209754 tokens (instead of 48321) (the corpus is cleaned up of French elements and stop words list is improved) (fig. 11)

Through the improved corpus the performance of the tool is obviously better. For example, the color theme is now concentrated in Topic 13 (“verde, azul, color..”) and we find topics which consist of semantically linked words. In Topic 8 we have the food and bullfight combined: “sopa, cuchillo, harina” (soup, knife, flour) are clearly food and kitchen words and “tarde, traje” (afternoon, costume) belong to the bullfight in Picasso’s vocabulary. Other examples of meaningful Topic-words are the rhyme words like “gritos, fritos” (screams, fried) in Topic 7 or “campana, ventana” (field, window) in Topic 10,

<sup>57</sup> Laure Thompson and David Mimno, “Authorless Topic Models: Biasing Models Away from Known Structure”, (COLING 2018) <http://www.cs.cornell.edu/~laurejt/papers/authorless-tms-2018.pdf>.



	Key 1	Key 2	Key 3	Key 4	Key 5	Key 6	Key 7	Key 8
Topic 1	noche	alas	do	canta	cama	vestido	caballo	mil
Topic 2	azul	detrás	manos	luz	sangre	sol	mano	hilo
Topic 3	don	azul	medio	cura	puesto	tarde	digo	mar
Topic 4	gori	caramelo	miel	tarde	quema	gotea	silencio	hora
Topic 5	tiene	pájaro	recuerdo	jamón	pone	cuenta	pasar	equilibrio
Topic 6	claro	día	verde	noche	después	silencio	salta	mesa
Topic 7	noche	después	par	día	borde	seda	gritos	fritos
Topic 8	tarde	patas	debajo	hechos	sopa	cuchillo	traje	harina
Topic 9	tejado	cara	abre	ojos	toro	borde	amor	dedos
Topic 10	papel	campana	ventana	pedra	oreja	agua	rosa	manos
Topic 11	ala	cuarto	tantas	aroma	aire	blancura	tira	silencio
Topic 12	sol	tío	luz	cuentas	cama	hora	ventana	vestido
Topic 13	verde	azul	color	lleno	medio	cielo	amarillo	piel
Topic 14	trigo	calor	trae	coche	azul	tarde	da	olor
Topic 15	mano	cae	lleva	tira	cara	envuelto	cuchillo	necesario
Topic 16	mano	lleva	cuerva	noche	reloj	hora	esconde	atraviesa
Topic 17	tarde	cielo	monte	cristal	arco	aire	iris	oro
Topic 18	sol	plumas	ruedas	negro	llena	olor	carne	pliegues
Topic 19	blanco	pie	frio	plata	oro	colgando	cartero	pared
Topic 20	gritos	mar	sartén	sol	lágrimas	tinta	par	medio

Figure 11: Screenshot: Result of DARIAH Topics Explorer: Topics for Picasso-corpus (quadrupled), Highlighted boxes added by N. R.-P.

where also “peidra, oreja, agua, rosa” would follow the same rhyme, but less intensive only on the last letter. The preparation of the corpus is far from being perfect, there are still functional words like “debajo” (under) and “después” (then, later) and forms of verbs like “tiene” (has) or “pone” (put) which should be lemmatized and/or cleaned up to get a corpus of semantic words.

The next important question and experiment should be: How is Picasso’s poetry performing in comparison to other poets of his tongue and time. To answer this question with the limited means I had, I took the same corpus as in stylometry for Picasso and added some of the same period, but with different style and poetics (Juan Ramón Jiménez, Ramón Gómez de la Serna, Rafael Alberti, Federico García Lorca, Jorge Guillen, Antonio Marchado) (see fig. 12 without Picasso and fig. 13 including Picasso).

For a comparison between Picasso and his contemporary Spanish colleagues, this is a simple method to point out differences in vocabulary. Adding Picasso in the experiment (fig. 13) shows that the vocabulary in the

	Key 1	Key 2	Key 3	Key 4	Key 5	Key 6	Key 7	Key 8
<b>Topic 1</b>	ángel	parece	ángeles	mar	alma	agua	tierra	viento
<b>Topic 2</b>	tiempo	sol	azul	luz	vida	hombre	mar	mundo
<b>Topic 3</b>	sangre	cielo	mundo	niño	ojos	luna	amor	caballo
<b>Topic 4</b>	mujer	vida	parece	cielo	tarde	tiene	rosas	corazón
<b>Topic 5</b>	poeta	tierra	abel	camino	mar	mañana	oh	mairena
<b>Topic 6</b>	noche	luna	hombre	nieve	alto	lejos	frío	pena
<b>Topic 7</b>	amargo	agua	verde	amor	tierra	noche	jinete	luna
<b>Topic 8</b>	agua	tarde	corazón	sombra	tiene	tierra	piedra	ojos
<b>Topic 9</b>	martín	amor	campo	día	mano	sol	alma	blanca
<b>Topic 10</b>	mis	luz	mar	aire	ojos	sombra	alma	tus

Figure 12: Screenshot: Result of DARIAH Topics Explorer: Topics for corpus of Spanish poetry 20<sup>th</sup> century (without Picasso), words in the corpus doubled, 10 Topics, 100 iterations. High-lighted boxes added by N. R.-P.

	Key 1	Key 2	Key 3	Key 4	Key 5	Key 6	Key 7	Key 8
<b>Topic 1</b>	alma	mar	sombra	viento	luna	ojos	mundo	piedra
<b>Topic 2</b>	silencio	xxxv	sangre	aire	ventana	toro	noche	reloj
<b>Topic 3</b>	día	mano	después	sol	crystal	decir	corazón	rosa
<b>Topic 4</b>	par	tarde	cara	debajo	capa	medio	copa	sartén
<b>Topic 5</b>	tierra	poeta	agua	martín	camino	oh	corazón	campo
<b>Topic 6</b>	don	luz	azul	blanco	verdes	día	frío	ventana
<b>Topic 7</b>	luz	luna	agua	noche	cielo	muerte	mujer	sangre
<b>Topic 8</b>	verde	sol	azul	cielo	negro	piel	agua	alás
<b>Topic 9</b>	tarde	amor	verde	vida	azul	sol	cielo	jardín
<b>Topic 10</b>	azul	fuego	da	pan	color	colores	borde	olor

Figure 13: Screenshot: Result of DARIAH Topics Explorer: Topics for corpus of Spanish poetry 20<sup>th</sup> century (with Picasso), words in the corpus doubled, 10 Topics, 100 iterations. High-lighted boxes added by N. R.-P.

whole corpus gets more disturbing, inconsistent, harder to understand. While in fig. 12 the first topic is clearly attributed to the angel poems by Alberti and looking into the distribution of topics in document would show where other poets are using this element. Another prominent and celebrated topic is the landscape and the new pilgrim in Machado's poetry which is here visible in Topic 5. Other topics are hinting to Lorca with the importance of "luz" (light) for example. But, adding Picasso's texts has the effect that not very well-defined topics out of the bullfight and color themes are mixed together (Topic 4 and 8). The other strong topics from the experiment before (fig. 12) are still visible (topic 5 with "tierra, poeta, ..., camino, corazón, campo"), but it is interesting how Picasso uses some of the same words like "azul" (together with Lorca) to combine them with strange motifs like "tarde" (afternoon), "ventana" (window) or "pan" (bread). Before, in Topic 2 (fig. 12) "azul" was part of a transcendental reflection on life, man and the world ("tiempo, sol, azul, luz, vida, hombre, mar, mundo") which can be called a common topic for modern poetry.

Regarding the few tests, it can only be an attempt, concluding that Picasso is writing in a fragmented style, a montage of words, not in sentences, not in paragraphs, not in chapters, not in verses. Still after comparing Picasso's poetry to others, one can read the topics as a poem based on Picasso's box of words. The results are showing us the way Picasso writes, but they do not give us a one-sided explanation of the content. It helps us to understand the principle of montage – without filling the gaps between. For me, this effect is much more helpful and gives more insight than the hermeneutic explanation (for example by Bernadac). We can be curious about to hear what more can be done by trying algorithms on Picasso's poetry as Luis Meneses and Enrique Mallen presented on the subject at the DH2017 in Mexico.<sup>58</sup> Though, I would not agree, that the method could give us insights in how was "Picasso's 'vision' of the world".<sup>59</sup> There is no vision to be captured, but a way of creation that inspires and surprises his readers and spectators every time anew.

---

<sup>58</sup> Luis Meneses and Enrique Mallen, "Distinctions between Conceptual Domains in the Bilingual Poetry of Pablo Picasso". Mexico City, 2018. <https://dh2018.adho.org/en/distinctions-between-conceptual-domains-in-the-bilingual-poetry-of-pablo-picasso/>.

<sup>59</sup> Meneses and Mallen, "Distinctions".



# Delta Inside Valle-Inclán

## Stylometric Classification of Periods and Groups of His Novels

José Calvo Tello (Würzburg)

**ABSTRACT:** In this article, I analyze the internal periodization and group of works in a corpus of novels by Valle-Inclán. The technique used is supervised Machine Learning (classification, K-Nearest Neighbors), using the stylometric Delta distances (cosine Delta, 5000 MFW) as features. The evaluation shows that this technique classifies correctly the undisputed texts by literary scholars. The prediction about the disputed texts establish that the year of greatest stylistic change in Valle was 1920 and not 1905. This technique also allows us to classify the works by Valle in only three groups: the *sonatas*, a *social-bélico* group, and the later *esperpentic* works.

**ZUSAMMENFASSUNG:** Im vorliegenden Beitrag analysiere ich die interne Periodisierung und Werkgruppen in einem Valle-Inclán Roman-Korpus. Die angewandte Methode ist überwachtes Machine Learning (classification, K-Nearest Neighbors). Dabei werden stilometrische Delta Distanzen (cosine Delta, 5000 MFW) als Parameter verwendet. Die Überprüfung zeigt, dass mithilfe der angewendeten Methode die nicht-umstrittenen Romane Valle-Incláns richtig klassifiziert werden. Die These bezüglich der umstrittenen Texte wird dahingehend bestätigt, dass der bedeutende stilistische Wechsel innerhalb des Romanwerks von Valle-Inclán 1920 stattfand und nicht wie bisher behauptet 1905. Die Methode erlaubt es uns außerdem, Valles Werk in drei Gruppen einzuteilen: die *sonatas*, eine *social-bélico*-Gruppe und später dann die *esperpentic*en Werke.

**KEYWORDS:** stylometry; Valle-Inclán; novel; Machine Learning; classification

**SCHLAGWÖRTER:** Stilometrie; Valle-Inclán; Roman; überwachtes Machine Learning; Klassifizierung

### 1. The Novels by Valle-Inclán

Valle-Inclán is one of the most influential Spanish writers of the 20<sup>th</sup> Century. Among other genres, he published 13 novels during his life: four *Sonatas*, three historical novels about the *Carlista* War, three avant-garde-historical novels<sup>1</sup> (part of *El ruedo ibérico*) and three other novels (*Flor de Santidad*, *La media noche* and *Tirano Banderas*).

<sup>1</sup> The last one, *Baza de Espadas*, was only published partly in a magazine before Valle's death.

The traditional studies of Valle-Inclán's works show some dispute about the classification of his novels as far as periodization and grouping his works is concerned. While it is clear that the early novels (*Sonatas*) and the late novels (*El ruedo ibérico*) are part of two different periods of Valle's writing, it is unclear which group the rest of the novels should be assigned to.

In a similar way, while some novels belong clearly to a certain group of works (the four *sonatas*, the three novels about the *Carlista* war, the three avant-garde-historical novels), the classification of the other three is unclear.

## 2. Stylometric Clustering and Classification

Since Burrows proposed Delta (2002), it has become one of the most widely used methods of Digital Humanities for authorial recognition. Many researchers have already pointed out that other aspects (signals) can affect its results, like genre,<sup>2</sup> gender,<sup>3</sup> narrative perspective, or period.<sup>4</sup> For that reason corpora tend to be created trying to neutralize these phenomena. Hoover<sup>5</sup> and Jannidis and Lauer<sup>6</sup> have used Delta with corpora of a single author in order to analyze the internal structure of the works clustering the Delta values.

In my proposal for the Hispanic Digital Association Conference in 2017<sup>7</sup> I explored to what extent the homogeneity of the clusters based on Delta values correlate with the metadata in single author corpora. The results of this

<sup>2</sup> John Burrows, "Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship," *Literary and Linguistic Computing* 17, 3 (2002): 267–87, here 271; Christof Schöch, Ulrike Henny, José Calvo Tello and Stefanie Popp, *The CLiGS Textbox*, (Würzburg: University of Würzburg, 2015) <https://github.com/cligs/textbox>.

<sup>3</sup> Shlomo Argamon, Moshe Koppel, Jonathan Fine and Shimoni Anat Rachel, "Gender, Genre, and Writing Style in Formal Written Texts," *Text and Talk*, 23 (2003): 321–46. <https://doi.org/0165-4888/0023-0321>.

<sup>4</sup> David L. Hoover, "Testing Burrows's Delta," *Literary and Linguistic Computing* 19, 4 (2004): 453–75, here 455.

<sup>5</sup> David L. Hoover, "A Conversation Among Himself: Change and the Styles of Henry James" in *Digital Literary Studies*, edited by David L. Hoover, Jonathan Culpeper and Kieran O'Halloran, (New York & London: Routledge, 2014) 90–119.

<sup>6</sup> Fotis Jannidis and Gerhard Lauer, "Burrows's Delta and Its Use in German Literary History" in *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, edited by Matt Erlin and Lynne Tatlock, (Rochester: Camden House, 2014) 29–54, [gerhardlauer.de/index.php/download\\_file/view/335/1/](http://gerhardlauer.de/index.php/download_file/view/335/1/).

<sup>7</sup> José Calvo Tello, "What Does Delta See inside the Author?: Evaluating Stylometric Clusters with Literary Metadata," in *III Congreso de La Sociedad Internacional Humanidades Digitales Hispánicas: Sociedades, Políticas, Saberes* (Málaga, 2017), 153–61. <http://hdh2017.es/wp-content/uploads/2017/10/Actas-HDH2017.pdf>.

research show that form, period, and the group of works tend to be differentiated by the clusters of Delta.

Although more present in the stylometric practice, clustering, an unsupervised machine learning method, has some disadvantages in compare to supervised methods: first, the known and unknown cases are not clearly divided. Second, the evaluation of the algorithm is more complicated than in classification tasks. Third, the amount of information in a dendrogram is strongly reduced: we see the closeness between the most similar texts but the rest of the distances to the other texts are only represented indirectly by the distances between the subclusters. And finally, the dendrogram used normally in stylometry does not show clear groups, making the evaluation even harder.

However, the results for the clustering of the corpus of the novels of Valle (stylo version 0.6.5, cosine Delta, 5000 MFW) based on the Delta matrix are:

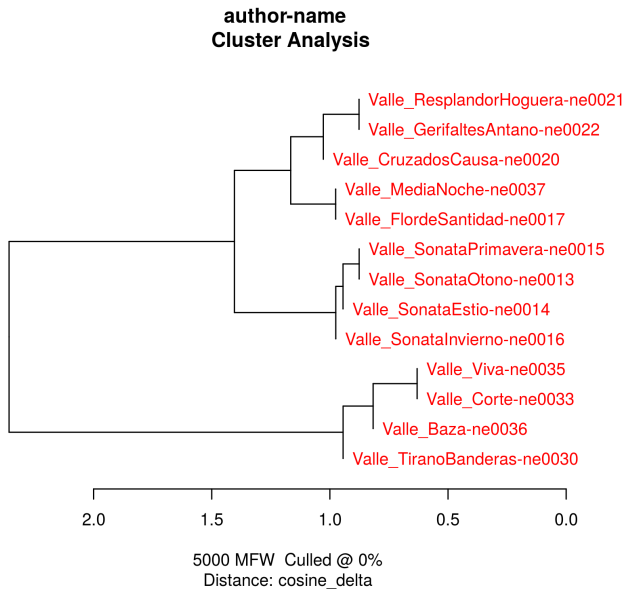


Figure 1: Dendrogram of Valle's novels

Although the dendrogram contains a clear *sonatas* subcluster in the middle, the rest is rather imprecise: do the three novels on the top constitute their own subcluster or should they be considered in conjunction with the

following two? What is the correct number of subclusters to be considered in this cluster? Two? Seven? Something in between?

In general, the kind of tasks for which stylometry is used can be fitted in a supervised task, such as classification, which helps to better answer the questions mentioned in the last paragraph. That is why in this article I use the Delta matrix as features for the classification task. The idea is to conserve the features that we know are the best for the authorial research (Delta values), but instead of deleting information to cluster them, we will use all the distances to the rest of the texts to classify them.

The corpus is composed of 13 works by Valle-Inclán, all of them part of the *Corpus of Novels of Spanish Silver Age* (CoNSSA). This collection is encoded in XML-TEI, validated and enhanced with metadata. The texts were already digitized by projects such as Gutenberg Project, Archive.org and ePubLibre. The CoNSSA will be available at the end of the CLiGS project and some of Valle's works are already part of the Textbox.<sup>8</sup> In any case, all the texts, metadata and code (as Jupyter Notebook) for this article are available in a GitHub repository<sup>9</sup>. In order to clarify the specific steps, the analysis is reproducible and gives the opportunity to reuse the data for other purposes.

### 3. Periods of Valle-Inclán's writing

I will first analyze the various periods in which Valle-Inclán's novels are inscribed. Some scholars have pointed out that the year of 1920 entails an important change in Valle's style.<sup>10</sup> After this year, he produces the *esperpentic* works in two different genres: novels (*Tirano Banderas*, *Corte de los milagros*....) and theater (*Cara de Plata*, *Luces de bohemia*). This hypothesis would group nine novels in the first period, and four in the second one.

Nevertheless, this date is not completely unproblematic. A contemporary intellectual to Valle, Maeztu, assigns the change of period not to 1920 but to 1905.<sup>11</sup> Although less accepted, some information about the subgenres and

---

<sup>8</sup> Schöch et al., *The CLiGS Textbox*.

<sup>9</sup> [https://github.com/morethanbooks/publications/tree/master/Delta\\_Valle](https://github.com/morethanbooks/publications/tree/master/Delta_Valle).

<sup>10</sup> José-Carlos Mainer, "Ramón Del Valle-Inclán," in *Historia y crítica de la literatura española* 6, (Barcelona: Crítica, 1980) 289–97, here 292; José-Carlos Mainer, *La Edad de Plata (1902–1939). Ensayo de interpretación de un proceso cultural* (Madrid: Cátedra, 2009), here 166; Felipe B. Pedraza Jiménez and Milagros Rodríguez Cáceres, *Manual de literatura española. 8: Generación de fin de siglo: introducción, líricos y dramaturgos* (Pamplona: Cénlit Ed., 1986, 2. ed.), here 613–16.

<sup>11</sup> Adolfo Sotelo Vázquez, "Valle-Inclán y Ramiro de Maeztu (Dos semblanzas de Valle por Maeztu: 1899 y 1936)" *Cuadernos Hispanoamericanos* 438 (1986): 83–114, here 90.



even characters of the works would support this theory: after 1905 Valle wrote two different series of novels that can be considered historical: *Las guerras carlistas* (1908–1909) and *El ruedo ibérico* (1927–1932). Valle even started a collection of plays called *Las comedias bárbaras* in 1907, which was finished only in 1923, a collection that would surpass the frontier of 1920. The protagonist of the theater play of 1923, *Cara de Plata*, is also the main character in a novel of 1909. The hypothesis of change in 1905 would then group five novels in the first period (the *sonatas* and *Flor de santidad*) and eight in the second one.

Both hypotheses can be condensed in the following table:

TITLE	YEAR OF PUBL.	HYPOTHESIS 1905	HYPOTHESIS 1920	CERTAIN FIRST PERIOD	CERTAIN SECOND PERIOD
<i>Sonata de otoño</i>	1902	1 <sup>st</sup> period	1 <sup>st</sup> period	yes	no
<i>Sonata de estío</i>	1903	1 <sup>st</sup> period	1 <sup>st</sup> period	yes	no
<i>Sonata de primavera</i>	1904	1 <sup>st</sup> period	1 <sup>st</sup> period	yes	no
<i>Flor de santidad</i>	1904	1 <sup>st</sup> period	1 <sup>st</sup> period	yes	no
<i>Sonata de invierno</i>	1905	1 <sup>st</sup> period	1 <sup>st</sup> period	yes	no
<i>Los cruzados de la Causa</i>	1908	2 <sup>nd</sup> period	1 <sup>st</sup> period	no	no
<i>El resplandor de la hoguera</i>	1909	2 <sup>nd</sup> period	1 <sup>st</sup> period	no	no
<i>Gerifaltes de antaño</i>	1909	2 <sup>nd</sup> period	1 <sup>st</sup> period	no	no
<i>La media noche</i>	1917	2 <sup>nd</sup> period	1 <sup>st</sup> period	no	no
<i>Tirano Banderas</i>	1926	2 <sup>nd</sup> period	2 <sup>nd</sup> period	no	yes
<i>La Corte de los Milagros</i>	1927	2 <sup>nd</sup> period	2 <sup>nd</sup> period	no	yes
<i>Viva mi dueño</i>	1928	2 <sup>nd</sup> period	2 <sup>nd</sup> period	no	yes
<i>Baza de espadas</i>	1932	2 <sup>nd</sup> period	2 <sup>nd</sup> period	no	yes

Table 1: Works of Valle arranged in two competing chronological groups

As we see, both hypotheses diverge on the four novels published after 1905 and before 1920, but they do concur regarding the other nine. That means that we can approach this problem from a supervised classification task in Machine Learning: we take the novels whose period is certain, put them in a train corpus and evaluate the model. If acceptable results are achieved, the model can be applied to the novels for which the hypothesis differ and let the model predict their periods.

### 3.1. Methodology

For this work the Delta matrix is created with *stylo*<sup>12</sup> (5000 MFW, cosine Delta) using the entire corpus of the novels. After that, the Delta Matrix is split in two matrices: one with the distances of the texts whose periodization is clear (the train corpus) and another with the disputed texts (the test corpus). That is, the distances of each text to the rest of Valle's corpus are treated as features and the metadata about period as labels.

Different classifiers were tested: Decision Trees, Logistic Regression, Random Forest, Supported Vector Machines and K-Nearest Neighbors (KNN). The latter two showed the best results and finally KNN was chosen since it does not require the different classes to be equally represented by data points (that means, the corpus does not have to contain equal amount of texts for, in this case, each period).

### 3.2. Evaluation and Results

To evaluate the model, I used cross validation, a technique in which “the data is instead split repeatedly and multiple models are trained”.<sup>13</sup> In this case the cross validation was repeated four times since the second period only had four different undisputed texts. All of the four times, the accuracy scores were 1.0: the algorithm achieved every time a perfect classification in the train corpus. The evaluation is also correct if we add to the corpus Valle's other prose texts:<sup>14</sup> three collections of short stories (*Femeninas*, *Jardín umbrío* and *Corte de Amor*) and an essay (*Lámpara maravillosa*).

If the algorithm classified the texts with an unchallenged period correctly, what is its prediction for the four disputed texts published after 1905 and before 1920? The algorithm provides the following results:

TITLE	YEAR	PREDICTED PERIOD
<i>Los cruzados de la Causa</i>	1908	first
<i>El resplandor de la hoguera</i>	1909	first
<i>Gerifaltes de antaño</i>	1909	first
<i>La media noche</i>	1917	first

Table 2: Predicted period of disputed text

<sup>12</sup> Maciej Eder, Mike Kestemont and Jan Rybicki, “Stylometry with R: A Package for Computational Text Analysis,” *The R Journal* 16,1 (2016): 1–15.

<sup>13</sup> Andreas C. Müller and Sarah Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists* (Beijing: O'Reilly, 2016), here 254.

<sup>14</sup> See texts and code in the subfolder of GitHub “/data/prose/”.

As we observe, each disputed text is classified by the algorithm as part of the first period. That means, the algorithm agrees with the hypothesis of 1920 as the year of change for Valle over the one that sets the change in 1905. Still, we don't know to what extent the hypothesis of 1905 is correct or wrong. To quantify the agreement of each hypothesis with the algorithm, we can evaluate both hypotheses with cross validation (four times) and show the accuracy scores as box plots:

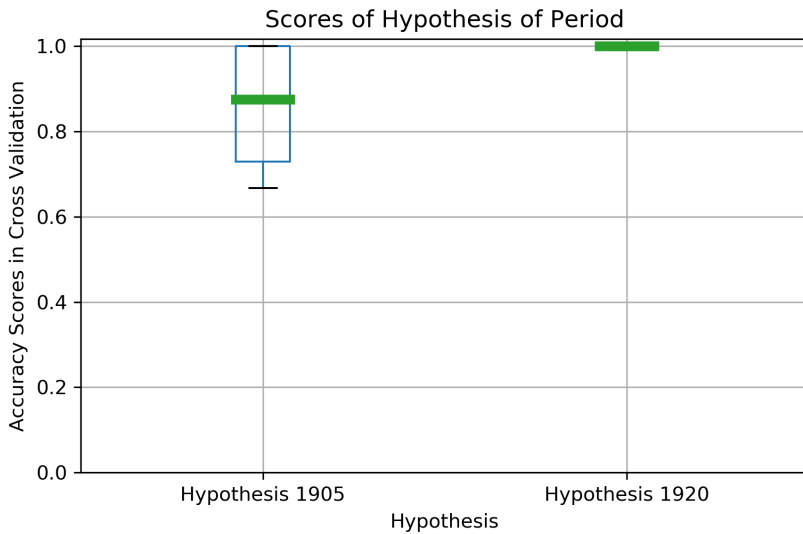


Figure 2: Accuracy scores for hypothesis of periodization in Valle's novels

Here we can conclude that while the median of accuracy scores of the hypothesis of 1905 is 0.88, all the scores of 1920 achieve perfect results of 1.0. That means, 1905 does represent a change of Valle's style, but 1920 defines clearer periods. The shift of Valle's style towards the *esperpento* is stronger than the similarity between the two groups of historical novels.

#### 4. Groups of Valle-Inclán's Works

A more complicated classification of Valle's works regards how to group his texts in relative smaller collections of texts (in Spanish *series*), which will be referred here as *group of works*. All the *sonatas* and the *carlista* novels share subtitles ("Memorias del Marqués de Bradomín" in the first case, "La guerra carlista" in the second), while it is undisputed that the last three novels belong

to “El ruedo ibérico”, a group that should have contained nine works but only the first three were published during Valle’s life.<sup>15</sup>

Different researchers have proposed alternative ways to group the other three texts (*Flor de santidad*, *Tirano Banderas* and *La media noche*). In the following table three different classifications of groups of works, proposed by literary scholars such as García de Nora (1963), Pedraza Jiménez and Rodríguez Cáceres (1986) and de Juan Bolufer (2000), are summarized:

TITLE	YEAR	SUBTITLE	PEDRAZA-RODRIGUEZ	NORA	BOLUFER
<i>Sonata de otoño</i>	1902	Memorias del Marqués de Bradomín	sonatas	modernista	sonatas
<i>Sonata de estío</i>	1903	Memorias del Marqués de Bradomín	sonatas	modernista	sonatas
<i>Sonata de primavera</i>	1904	Memorias del Marqués de Bradomín	sonatas	modernista	sonatas
<i>Flor de santidad</i>	1904	Historia milenaria	flor	modernista	flor
<i>Sonata de invierno</i>	1905	Memorias del Marqués de Bradomín	sonatas	modernista	sonatas
<i>Los cruzados de la Causa</i>	1908	La guerra carlista	guerra carlista	carlista	bélica
<i>El resplandor de la hoguera</i>	1909	La guerra carlista	guerra carlista	carlista	bélica
<i>Gerifaltes de antaño</i>	1909	La guerra carlista	guerra carlista	carlista	bélica
<i>La media noche</i>	1917	Visión estelar de un momento de guerra	otras	not analyzed	bélica
<i>Tirano Banderas</i>	1926	Novela de tierra caliente	tirano	esperpéntica	esperpéntica
<i>La Corte de los Milagros</i>	1927		ruedo	esperpéntica	esperpéntica
<i>Viva mi dueño</i>	1928		ruedo	esperpéntica	esperpéntica
<i>Baza de espadas</i>	1932		ruedo	esperpéntica	esperpéntica

Tab. 3: Works of Valle arranged in several competing groups of works

Even if the different scholars use different specific labels (for example “guerra carlitas”, “carlista”, or “bélica” for the *carlista* novels), they all agree that eleven novels constitute three clear groups: four *sonatas*, three *carlistas*

<sup>15</sup> The first two as book, the last one only in magazines.

and three *El ruedo ibérico*. The remaining three are classified in different ways:

- *Flor de santidad* is classified by García Nora together with the *sonatas*, while the other two form an ad hoc class<sup>16</sup>
- *La media noche* is treated by de Juan Bolufer together with the *carlista* novels;<sup>17</sup> García Nora does not analyze this text and Pedraza-Rodríguez<sup>18</sup> use another ad hoc class
- Pedraza-Rodríguez assign an ad hoc class for *Tirano Banderas*, while the other two group it within the esperpentic group.

#### 4.1. Evaluation and Results

Like in the previous section, this problem can be analyzed as a classification task in which the undisputed texts are placed in a train corpus and the disputed texts are placed in a test set. For this, I used the exact same methodology that I explained for the previous case. Now the cross validation could only be run three times (because the *carlista* novels only had three novels). Again, the accuracy scores (based on cosine Delta matrix, 5000 MFW, KNN), for the evaluation in the train corpus achieve a perfect value of 1.0 every time.

The prediction of the corresponding group for the three disputed texts is as follow:

TITLE	YEAR	KNN PREDICTED GROUP OF WORKS
<i>Flor de santidad</i>	1904	carlista
<i>La media noche</i>	1917	carlista
<i>Tirano Banderas</i>	1926	esperpéntica

Table 4: Predicted groups of disputed text

The results of the prediction concur with Nora and Bolufer to classify *Tirano Banderas* as an esperpentic novel and also with Bolufer in considering *La media noche* along with the *carlista* novels. The new nuance that the prediction provides is the classification of *Flor de santidad* in this same group of the *carlista* novels as well, an unexpected association.

<sup>16</sup> Eugenio García de Nora, *La novela española contemporánea* (Madrid: Ed. Gredos, 1963).

<sup>17</sup> Amparo de Juan Bolufer, *La técnica narrativa en Valle-Inclán* (Santiago de Compostela: Universidade de Santiago de Compostela, 2000).

<sup>18</sup> Pedraza /Rodríguez, *Manual de literatura española*.

Let us reflect on this new group of novels by Valle that Delta and the classification method have revealed: a group of works published between 1904 and 1917 and therefore with a start date before the *sonatas* are finished (published between 1902 and 1905). This group of works has a thematic and chronological core with the three novels of the *carlista* war, another war novel (about World War I) and a *sui generis* novel (*Flor de santidad*) in a timeless Galicia about a female character who suffers hallucinations and has sexual relationships with a man who claims to be Jesus Christ. The relation between the three *carlista* novels is very clear, and also the closeness to the other war-novel is predictable. But how do we fit in *Flor de santidad* with the rest of these novels? Although very different, *Flor de santidad* does share some literary phenomena with these novels: its story shares settings with the *carlista* *Los cruzados de la Causa*. It can be argued that the action of *Sonata de invierno* also takes place in a rural Galicia, one of the reasons that makes Nora group the disputed novel together with the *sonatas*. But other aspects do differentiate *Flor de santidad* from the *sonatas*: they have an autodiegetic narrator and share an upper-class protagonist: Marqués de Bradomín. In contrast, all the novels of this new group have a heterodiegetic narrator and the conditions of lower-class characters are portrayed in detail. That is why I am proposing to use the following label (in Spanish) for this group: *social-bélico*.

In addition to the already existing three hypotheses from literary scholars, I am proposing a new way of classifying Valle's work based on the results of the Delta matrix and the prediction of the algorithm. Now, the different hypothesis (the already existing ones and the new one) can be evaluated and compared to the results of the accuracy scores of a cross validation (4 times).

This shows that Pedraza/Rodríguez, Bolufer and Calvo Tello have a median of accuracy of 1.0. Bolufer and Pedraza/Rodríguez mismatch some texts in the cross validation evaluation, while the new hypothesis with the *social-bélico* group is correctly classified every time. From the already existing way of classifying the groups, Bolufer achieves the best results, an outcome that was easily predictable, since she is the only expert on the specific field of Valle (the other two address bigger collections of texts).

Given the possible existence of other ways of classification proposed by other literary scholars, which have not been considered, here I would like to encourage scholars to report any missing information through the issues of GitHub and I will update the results in the newest versions of the Jupyter notebook cited before. Likewise, if new digital editions of the texts with bet-

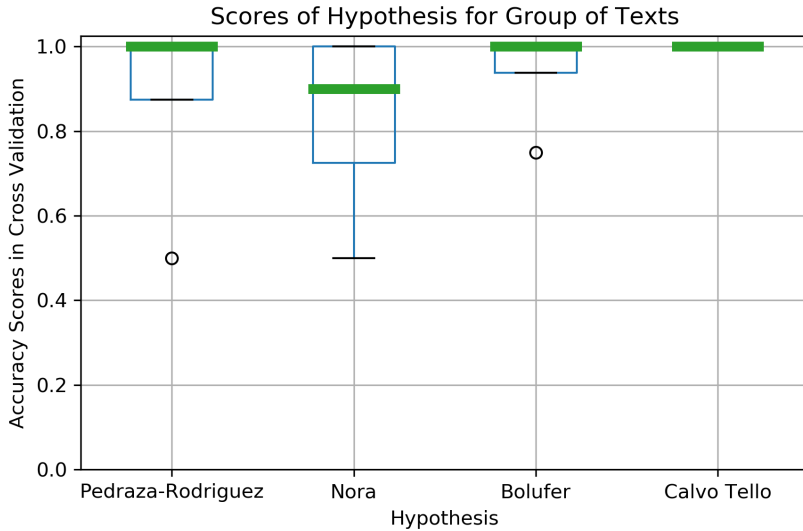


Figure 3: Accuracy scores in cross validation (KNN) of different hypotheses

ter quality or a more precise philological control are published, for example with the publication of the archive GIVIUS<sup>19</sup> the analysis can be run again.

If we go back to the cluster, now we can represent the color of the subclusters with the labels from the classification, as it is shown in the figure 4.

## 5. Conclusion

In this article, I have analyzed two specific questions regarding the corpus of Valle-Inclán's novels: first, what is the internal periodization of Valle's novels? And second, what is the internal classification within groups of works? I have applied and evaluated different hypotheses proposed by literary scholars. To answer these questions, I have used Delta distances from the novels as features in a classification task, evaluating whether this methodology sorts the undisputed texts correctly. For both cases the algorithm achieved perfect results.

The results show that 1920 is the frontier between the first and the second period of Valle's novels.

<sup>19</sup> Margarita Santos Zas and Carmen Elena Vilchez Ruiz, "Valle-Inclán en red: el Archivo Digital del GIVIUS," in *Sociedad, políticas, saberes*, (Málaga: HDH, 2017) 85–87, <http://hdh2017.es/wp-content/uploads/2017/10/Actas-HDH2017.pdf>.

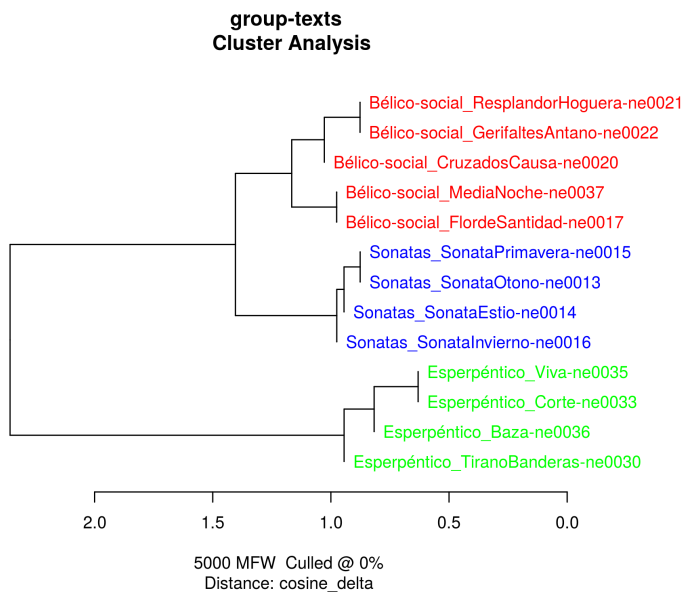


Figure 4: Dendrogram of Valle's Novels with groups from the classification

The predictions of the algorithm for the three disputed cases of group of works match in two cases with hypotheses from literary scholars and also establish an unexpected and new twist: the fact that *Flor de santidad* belongs to the war novels group. With these results, I am proposing a new classification for Valle's novels with only three groups: *sonatas* (four novels), *bélico-sociales* (five novels) and *esperpénticas* (four novels). These groups not only share lexical features (captured in the form of Delta distances), but also literary phenomena and chronological closeness.

This article has only analyzed Valle's novels, although his oeuvre is more extensive and contains different genres. Other scholars are currently working on the digitization of Valle's works,<sup>20</sup> so I hope that in the near future we will be able to reanalyze these questions with more data that will make our work even more solid. In any case, in this article I am proposing a way

<sup>20</sup> Santos Zas, "Valle-Inclán en red"; Concepción María Jiménez, Elena Martínez Carro, María Teresa Santa María, José Calvo Tello, María Simón Parra, Roxana Beatriz Martínez Nieto and María García Sánchez, "BETTE: Biblioteca Electrónica Textual Del Teatro En Español de La Edad de Plata" in *Sociedad, políticas, saberes*, (Málaga: HDH, 2017) 88–91, <http://hdh2017.es/wp-content/uploads/2017/10/Actas-HDH2017.pdf>.



---

of mutual evaluation between traditional literary studies and digital humanities: evaluation of digital methods through solid literary knowledge, on the one hand, but also evaluations of diverging literary hypotheses through solid digital methods, on the other.



# Psychologie des Stils

## Autopsyche, Erzählanalyse und die Poetik Antonin Artauds

Andrea Gremels und Maren Scheurer (Frankfurt am Main)

**ZUSAMMENFASSUNG:** Aus einer interdisziplinären Perspektive denkt der Beitrag über neue Anregungen zu einer allgemeinen Theorie des individuellen Autorstils nach. Ausgehend von klassischen Stilbegriffen von Georges Buffon, Ludwig Börne und Roland Barthes steht dabei die Frage im Vordergrund, inwiefern die erzählanalytischen Methoden der psychologischen Wissenschaften dazu beitragen können, den Zusammenhang zwischen Autorpsyche und Stil zu erhellen. Dass eine Wechselwirkung unbestreitbar ist, zeigt das Beispiel des surrealistischen Schriftstellers Antonin Artauds, dessen Werk oft in einem biographischen und pathologischen Rahmen untersucht wird. Artauds Stil gründet jedoch weit darüber hinausgehend auf einer komplexen, das Unbewusste und Triebhafte ausschöpfenden, surrealistischen Poetik. Nicht nur im Hinblick auf die enge Verschränkung von Somatik, Poetik und Psyche in Artauds Werk, sondern auch in Bezug auf eine allgemeine Theorie zeigt sich der Stil als ein überdeterminiertes und komplexes Phänomen, das sich einer absoluten auktorialen und analytischen Kontrolle entzieht.

**ABSTRACT:** From an interdisciplinary perspective, we think about new remarks on a general theory of individual and authorial style. Based on the traditional definition of style set by Georges Buffon, Ludwig Börne and Roland Barthes, we focus on the question, to what extent the narrative analytical methods of psychology can help to illuminate the relationship between the psyche of the author and style. The example of the surrealist writer Antonin Artauds shows clearly this named interdependency. His work was until now often analysed in a biographical and pathological framework. Artaud's style, however, is founded far beyond that on a complex, exhausting the unconscious and impulses, surrealist poetics. Artaud's style is an over-determined and complex phenomenon that eludes absolute authorial and analytical control as can be shown by focusing on the close entanglement of somatics, poetics and psyche in Artaud's writings as well as in relation to general theory.

**SCHLAGWÖRTER:** Artaud, Antonin; Poetik; Stil; Psychoanalyse

**KEYWORDS:** Artaud, Antonin; poetics; style; psychoanalysis

## Einleitung

„Le style est l'homme même“: Mit dieser Formel hat Buffon seit 1753 die Debatte über den Zusammenhang zwischen Autor und Stil geprägt.<sup>1</sup> Buffon stellt zwar in seiner Rede vor der französischen Akademie den Stil in erster Linie als hart zu erarbeitendes Produkt rhetorischer Raffinesse dar und nicht als „Manifestation der Subjektivität des Autors“, wie sein Ausspruch gerade in der Romantik interpretiert und weiterverbreitet wurde.<sup>2</sup> So wendet sich Ludwig Börne 1826 explizit *gegen* Buffon, wenn er zu bedenken gibt, dass man „der erste Grobian“ sein und trotzdem „die zartesten Lieder“ dichten könne.<sup>3</sup> Auch dieser Einspruch dokumentiert aber nur umso deutlicher, dass sich spätestens mit Buffon ein Problem in die Stildebatte einschleicht, das sich nicht leicht lösen lässt: Auf der einen Seite steht die Intuition, dass Stil unmittelbar etwas mit der Seele des Autors zu tun hat und somit paradoxerweise ein Stück weit seiner Kontrolle entzogen ist, und auf der anderen Seite die Erfahrungen, dass guter oder außergewöhnlicher Stil erlernbar und damit kontrollierbar ist, obwohl oder gerade weil andere Faktoren wie Sprache, Gattung und Diskurs auf ihn einwirken. Vor dem Hintergrund dieses widersprüchlichen Zusammenhangs behilft sich Roland Barthes in seiner frühen Schrift *Le degré zéro de l'écriture* (1953) damit, das Phänomen Stil auf drei Ebenen aufzuspalten, die Ebene der *langue*, die das historisch gegebene Sprachmaterial betrifft, die Ebene des *style*, der eine unbewusste körperlich-psychische Dimension des Sprachgebrauchs betrifft, und die *écriture*, die zwar noch in der gesellschaftlichen Position des Schriftstellers verankert bleibt, ihm aber die Freiheit lässt, als Kompromiss zwischen diesen Gegebenheiten eine eigene Schreibform zu entwickeln.<sup>4</sup>

Später hat Barthes zusammen mit anderen Poststrukturalisten wesentlich dazu beigetragen, die Bedeutung des Autors so weit aus der literaturwissenschaftlichen Debatte auszutragen, dass autorzentrierte Einflussfaktoren auch in der Stildebatte kaum noch Berücksichtigung finden konnten. Doch das Problem ist damit noch nicht gelöst und die Frage nach der Bedeutung

<sup>1</sup> Georges Buffon, „Discours prononcé à l'académie françoise par M. de Buffon, le jour de sa réception“, in *Œuvres philosophiques de Buffon*, hrsg. von Jean Piveteau (Paris: Presses Univ. de France, 1954), 500–4, hier 503.

<sup>2</sup> Wolfgang G. Müller, „Der Topos ‚Le Style est L'Homme Même‘“, *Neophilologus* 61, Nr. 4 (1977): 481–94, hier 485.

<sup>3</sup> Ludwig Börne, „Bemerkungen über Sprache und Stil“, in *Ludwig Börnes gesammelte Schriften: erster Band* (Berlin: Carl Hermann Otto, o. J.), 5–10, hier 6.

<sup>4</sup> Vgl. Roland Barthes, *Le degré zéro de l'écriture* (Paris: Éditions du Seuil, 1953), 19–28.

der Autorpsychie für den Stil nicht aus der Welt geschafft. Weitere Hinweise dazu liefern die Forschungen der *Digital Humanities*. Auch wenn in den fachinternen Debatten zwischen hermeneutischen, poststrukturalistischen, kulturwissenschaftlichen und stilometrischen Ansätzen noch nicht geklärt ist, mit welchen Stilbegriffen die unterschiedlichen Verfahren operieren und in welchen Punkten sie konvergieren, ist es doch ein bemerkenswerter Fund, dass man auf Basis der häufigsten Wörter in einem Textkorpus relativ zuverlässige Autorattributionen vornehmen kann<sup>5</sup> oder dass sich in solchen Analysen die Eigenarten des Autors gegenüber Genremerkmalen zuverlässig durchsetzen.<sup>6</sup> Ausgehend von diesen Befunden wollen wir uns noch einmal mit der theoretischen Frage, wie Stil und Autor zusammenhängen, auseinandersetzen und eine weitere Disziplin in unsere Überlegungen mit hineinholen, welche die persönlichen oder psychologischen Bedingungen von Stil, die hier doch zur Debatte stehen, genauer untersucht. Wenn wir im Folgenden auf die psychologischen Wissenschaften Bezug nehmen, so tun wir dies explizit nicht, um deren Verfahren als stilanalytische Methode für Lektüreverfahren fruchtbar zu machen. Uns geht es auch nicht darum, für einen ‚psychologisierenden‘ Zugriff auf literarische Werke zu plädieren oder zu postulieren, dass eine Diagnose der Autorpsychie das Stilproblem lösen könnte. Das Gegenteil wird sogar der Fall sein. Im Mittelpunkt der Analyse steht die Fragestellung, welche Anregungen die methodischen Ansätze aus Psychologie und Psychoanalyse für eine interdisziplinär geführte, theoretische Debatte über den Zusammenhang von Autor und Stil bieten können.

Psychologie und Psychoanalyse befassen sich schon seit einiger Zeit mit dem Erzählstil ihrer Probanden. Transkriptionen von Patientenerzählungen werden im Austausch mit Linguistik und Literaturwissenschaft auf Erzählstrukturen und -perspektiven sowie auf grammatikalische und lexikalische Besonderheiten hin untersucht. Solche Erzählanalysen sollen der Diagnostik von Persönlichkeitsmerkmalen dienen und therapeutische Herangehenswei-

---

<sup>5</sup> Vgl. Nanette Rißler-Pipka, „Avellaneda y los problemas de la identificación del autor: propuestas para una investigación con nuevas herramientas digitales“, in *El otro Quijote: la continuación de Avellaneda y sus efectos*, hrsg. von Hanno Ehrlicher (Mesa Redonda und Universität Augsburg, 2016), 27–51. <https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/3704>; und Christof Schöch, „Corneille, Molière et les autres: stilometrische Analysen zu Autorschaft und Gattungszugehörigkeit im französischen Theater der Klassik“, *PhiN* 7 (2014): 130–57, hier 134.

<sup>6</sup> Franco Moretti u. a., „Quantitative Formalism: an Experiment“, *Literary Lab* 15.01.2011: 1–26, hier 13, <https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf>.

sen an eine Reihe psychiatrischer Erkrankungen schaffen. Auch wenn diese Methoden zur literarischen Analyse ungeeignet sind, so scheint es uns doch von besonderem Interesse, sie für eine Präzision unserer Annahmen für eine allgemeine Theorie des Individualstils heranzuziehen. Hierzu bedarf es zunächst einer kritischen Überprüfung der psychologischen Erzählanalysen. Anschließend möchten wir unsere Erkenntnisse an einem besonders einschlägigen Beispiel für die Stilproblematik diskutieren und vertiefen: den Briefen des surrealistischen Autors Antonin Artauds, die er 1943 in der Nervenheilanstalt Rodez verfasst hat. Den Theaterschriftsteller, Dichter, Schauspieler und Romancier umgibt eine quasi-mythische Aura, die oft dazu beigetragen hat, dass seine Biographie, Persönlichkeit und psychische Verfassung zur Beurteilung seiner Werke herangezogen wurde. Schon von Kindheit an von Krankheiten gezeichnet, wurde er später drogenabhängig, unternahm 1936 eine Reise nach Mexiko, wo er sich von der Kaktusdroge Peyote Heilung versprach und wurde bei seiner Rückkehr nach Europa in Irland wegen öffentlicher Ruhestörung festgenommen und fortan in verschiedene psychiatrische Kliniken eingewiesen. Als schizophran diagnostiziert, verbrachte er die Jahre 1943–1946 in der Nervenheilanstalt Rodez, wo man ihn mit Elektroschocks behandelte. Artauds Schriften aus Rodez sind deshalb oft als Ausdruck seiner zunehmenden Schizophrenie oder als Reaktion auf die Elektroschocktherapie gelesen wurden. Der ausgefeilten surrealistischen Poetik Artauds werden solche biographischen oder ausschließlich auf Artauds psychische Konstitution gerichteten Ansätze, wie wir zeigen werden, allerdings nicht gerecht. Das zu Anfang skizzierte Problem der Prägung des Stils durch die Autorpsyche ist damit aber noch nicht geklärt, vielmehr stellt sich umso dringlicher die Frage, wie mit dem unwägbareren Einfluss der Autorpsyche in stilistischen Untersuchungen umgegangen werden soll. Unser Artikel versteht sich nicht als Lösungsvorschlag, sondern als Exploration dieser Frage und verschiedener Lösungsansätze. Ziel ist es also nicht, das Problem aus dem Weg zu räumen, sondern seine theoretischen Konsequenzen zu erforschen.

### **Erzählanalytische Verfahren in Psychoanalyse und Psychologie**

Spätestens seit Sigmund Freud beschäftigen sich Psychoanalytiker mit der Idee, dass die Art, wie ein Patient erzählt, Rückschlüsse auf seine Persönlichkeit und seine psychische Verfassung liefern könnte. In einer Anmerkung zum *Bruchstück einer Hysterie-Analyse* berichtet Freud, er habe einmal

aufgrund der stilistischen Klarheit der Erzählung einer Patientin eine psychische Störung ausschließen können.<sup>7</sup> Damit hat er Überlegungen zur diagnostischen Relevanz der sprachlichen Gestalt des Erzählens eingeleitet, auf die mittlerweile zahlreiche Techniken der Erzählanalyse zurückgreifen.

All diesen Verfahren liegt die Theorie zugrunde, dass sich die Erzählerpsyche nachhaltig auf die Erzählgestalt auswirkt. Wie zahlreiche Studien zeigen, beeinflussen die Struktur der psychischen Prozesse des Patienten, die Stärke und Qualität seiner Emotionen,<sup>8</sup> seine Fähigkeit, eigene und fremde Motive, Gefühle und Gedanken zu reflektieren und verschiedene Perspektiven einzunehmen,<sup>9</sup> den Stil der Patientenerzählung. Auch bestimmten Störungsbildern werden in dieser Forschung charakteristische Erzählmerkmale zugewiesen. Analysen der Erzählungen depressiver Patienten zeigen, dass sie sich allumfassend aus der Kommunikation herausziehen, was sich stilistisch durch einen Rückzug auf die erste Person Singular und die signifikant seltenere Verwendung von dramatischen und engagierenden Redefiguren wie der direkten Rede abzeichnet.<sup>10</sup> Lisa Capps und Elinor Ochs wiederum weisen in den Narrativen einer Agoraphobikerin durchgängige Hilflosigkeitsgesten nach, die sie bis in bevorzugte Satzkonstruktionen, Modalverbauswahl und den Einsatz von Verstärkungspartikeln verfolgen.<sup>11</sup> Arnulf Deppermann und Gabriele Lucius-Hoene untersuchen in den Erzählungen traumatisierter Patienten, wie prekär diese an den „Grenzen des Sagbaren“ entlanglaufen; der Schrecken beeinträchtigt die Gedächtnis- und Wahrnehmungsfähigkeit, „sodass die Betroffenen häufig um angemessene Ausdrucksformen ringen müssen“<sup>12</sup>; stilistische Merkmale wie abstrakter Ausdruck, „geringer Detaillierungsgrad“, „achronologische Darstellung“

---

<sup>7</sup> Sigmund Freud, „Bruchstück einer Hysterie-Analyse“, in *GW V*, hrsg. von Anna Freud (London: Imago, 1941), 161–286, hier 174.

<sup>8</sup> Vgl. Tilmann Habermas, Michaela Meier und Barbara Mukhtar, „Are Specific Emotions Narrated Differently?“, *Emotion* 9 (2009): 751–62, hier 755–7.

<sup>9</sup> Vgl. Tilmann Habermas, „Who Speaks? Who Looks? Who Feels? Point of View in Autobiographical Narratives“, *International Journal of Psychoanalysis* 87, Nr. 2 (2006): 498–518, hier 500–1.

<sup>10</sup> Vgl. Wilma Bucci und Norbert Freedman, „The Language of Depression“, *Bulletin of the Menninger Clinic* 45 (1981): 334–58, hier 336–7, 348–50.

<sup>11</sup> Vgl. Lisa Capps und Elinor Ochs, *Constructing Panic: the Discourse of Agoraphobia* (Harvard: Harvard UP, 1995), 66.

<sup>12</sup> Arnulf Deppermann und Gabriele Lucius-Hoene, „Trauma erzählen: kommunikative, sprachliche und stimmliche Verfahren der Darstellung traumatischer Erlebnisse“, *Psychotherapie und Sozialwissenschaft* 7 (2005): 35–73, hier 37.

und „Darstellungslücken“ sind die Folgen.<sup>13</sup> Obwohl Capps und Ochs vorsichtig sind, ihre Ergebnisse zu verallgemeinern, und auch Deppermann und Lucius-Hoene betonen, dass „von *der* Sprache des Traumas nicht die Rede sein“ könne,<sup>14</sup> zeichnet sich umso deutlicher ab, dass Psychoanalyse und Psychologie davon ausgehen, dass die individuelle Patientenpersönlichkeit den Stil einer Erzählung so stark beeinflusst, dass aus einer Stilanalyse umgekehrt wieder Rückschlüsse auf die Persönlichkeit und bestimmte Störungsbilder gezogen werden können. Capps und Ochs versprechen sich von der individuellen Analyse grammatischer und diskursiver Merkmale einen Einblick in die psychische Struktur der Patienten und „the fabric of their emotional world“,<sup>15</sup> sowie vielleicht sogar eine Einsicht darüber, wie Sprache, Grammatik und Stil dazu beitragen, eine bestimmte Identität und Sicht der Welt überhaupt erst zu formen und zu perpetuieren.<sup>16</sup> Hier entspringt Stil nicht nur der Persönlichkeit, sondern er bildet sie auch aus.

Betrachten wir nun noch genauer einige Erzählanalysen zur Schizophrenie, um die Brücke zu Artaud zu schlagen. Zahlreiche Studien haben gezeigt, dass schizophrene Patienten in ihrem Sprachverständnis und Sprachgebrauch deutlich von gesunden Individuen abweichen – durch geringeres Sprachaufkommen, niedrige Ideendichte, Informativität und Komplexität, Inkohärenz, Ambiguität, Ungeordnetheit und Exzentrizität.<sup>17</sup> Weitaus schwieriger ist es dagegen, Einigkeit darüber herzustellen, welches die spezifischen Merkmale des schizophrenen „Stils“ sind. Grund dafür ist eine Reihe von Faktoren. So haben Fineberg und ihre KollegInnen gezeigt, dass Schizophrene auffallend weniger „Ich“ sagen als Patienten mit affektiven Störungen,<sup>18</sup> während die Forschergruppe um Hong nachweist, dass Schizophrene häufiger „Ich“ sagen als Individuen ohne psychiatrische Auffälligkeit.<sup>19</sup> Oft werden in dieser Manier nur einzelne Sprachbesonderheiten gegen wech-

<sup>13</sup> Deppermann und Hoene, „Trauma erzählen“, 63.

<sup>14</sup> Deppermann und Hoene, „Trauma erzählen“, 62.

<sup>15</sup> Capps und Ochs, *Constructing Panic*, 180.

<sup>16</sup> Capps und Ochs, *Constructing Panic*, 54.

<sup>17</sup> Vgl. Aubrey M. Moe u. a., „Idea Density in the Life-Stories of People with Schizophrenia: Associations with Narrative Qualities and Psychiatric Symptoms“, *Schizophrenia Research* 172 (2016): 201–5, hier 201–2; Kai Hong u. a., „Lexical Use in Emotional Autobiographical Narratives of Persons with Schizophrenia and Healthy Controls“, *Psychiatry Research* 225 (2015): 40–9, hier 40.

<sup>18</sup> Vgl. S. K. Fineberg u. a., „Word Use in First-Person Accounts of Schizophrenia“, *The British Journal of Psychiatry* 206 (2015): 32–8, hier 32.

<sup>19</sup> Vgl. Hong u. a., „Lexical Use“, 46.



selnde Kontrollgruppen getestet, sodass es schwierig ist, auf Basis dieser Einzelstudien eine allgemeingültige Aussage über „den“ schizophrenen Stil zu treffen.<sup>20</sup> Saavedra wiederum weist nach, dass Schizophrene in einem frühen Stadium ihrer Behandlung kaum kohäsiv erzählen, ihre Referenzen selten kontextualisieren und häufig mit Neologismen arbeiten.<sup>21</sup> Mit längerer Behandlungsdauer nehmen Kohärenz und Kohäsion aber wieder zu,<sup>22</sup> sodass zur individuellen Ausprägung der Schizophrenie auch ihre Behandlungsgeschichte in die Form des jeweiligen Erzählstils einwirkt. Überdies scheint es eine Rolle zu spielen, ob Schizophrene mündlich erzählen oder ihre Erfahrungen schriftlich ausformulieren.<sup>23</sup>

Abschließend müssen wir festhalten, dass die psychologischen Untersuchungen bislang keine allgemeingültigen Aussagen zum Stil einer bestimmten psychischen Disposition oder eines bestimmten Krankheitsbildes treffen können – zu groß scheinen die individuellen Unterschiede durch andere Persönlichkeitsfaktoren, die Gestaltungsfreiheit in der geschriebenen Sprache ebenso wie die Untersuchungsmethoden in der Forschung. Da sich die Untersuchungen meist auch nicht auf literarisch ausgestaltete Texte, sondern gesprochene oder schriftliche Alltagssprache beziehen, bleibt aus literaturwissenschaftlicher Perspektive auch das Problem der Eigengesetzlichkeit der Literatur unberücksichtigt, sodass eine unmittelbare Übertragung dieser Verfahren auf literarische Analysen ohnehin äußerst problematisch wäre. Nichtsdestotrotz weisen die zahlreichen Studien in all ihrer Widersprüchlichkeit doch auf ein wichtiges Ergebnis hin: Die Persönlichkeit und psychische Konstitution eines Sprechers steht unzweifelhaft in einer Wechselwirkung mit seinem sprachlichen Ausdruck. Eine Theorie des Stils, die das Problem der Autorpsyche löst, indem sie sie gänzlich ausklammert, wäre damit ebenso unbefriedigend und verkürzend wie Lesarten, die alle Textphänomene durch die genaue Diagnose der psychischen Befindlichkeit des Autors zu erklären suchen.

---

<sup>20</sup> Vgl. Lynn E. DeLisi, „Speech Disorder in Schizophrenia: Review of the Literature and Exploration of Its Relation to the Uniquely Human Capacity for Language“, *Schizophrenia Bulletin* 27, Nr. 3 (2001): 481–96, hier 489.

<sup>21</sup> Vgl. Javier Saavedra, „Quantitative Criteria of Narrative Coherence and Complexity in Persons with Paranoid Schizophrenia“, *J Nerv Ment Dis* 198 (2010): 349–55, hier 352.

<sup>22</sup> Vgl. Saavedra, „Quantitative Criteria“, 353.

<sup>23</sup> Vgl. De Lisi, „Speech Disorder“, 492.

## Artaud und der Surrealismus

Vor allem die frühe Phase des Surrealismus kennzeichnet ein Künstlerverständnis, das sich explizit gegen einen Zusammenhang von Individualstil und Autorpersönlichkeit wendet. Mit der anti-künstlerischen Aussage „nous n'avons pas du talent“ beriefen sich die Autoren des ersten Manifests des Surrealismus auf Schreibweisen wie die *écriture automatique*, die das kollektive Unbewusste aktivieren sollten.<sup>24</sup> Das Stilproblem wird im Surrealismus nicht nur auf „Verfahren“, „*procédés*“, „Methoden“ verlagert, so Karlheinz Barck, sondern kulminiert in einer „tendenzielle[n] Überwindung einer kunst- und subjektzentrierten Stilauffassung“.<sup>25</sup> Denn das revolutionäre Autorverständnis der Surrealisten bestand darin, sich gegen ein kompositorisches Prinzip zu sperren, die Zensur der Rationalität zu umgehen und die Subjektivität des Autors, d.h. dessen bewusste Kontrolle über sein künstlerisches Produkt auszuschalten.<sup>26</sup>

Doch ganz ausschalten lässt sich die Autorsubjektivität im Surrealismus doch nicht. Im Zweiten Manifest verteidigt Breton die Methode der *écriture automatique*, gibt aber zu bedenken, dass sie oft missverstanden wurde. Denn es ginge dabei nicht darum, die „Feder einfach über das Papier fliegen zu lassen“ oder „pittoreske Traumbilder aneinanderzureihen“, sondern die unentdeckte Stimme des Unbewussten wachzurufen, um damit andere Denkweisen bzw. Logiken offenzulegen,

de nous livrer des étendues *logiques* particulières, très précisément celles où jusqu'ici la faculté logique restent inexplorées, mais encore on demeure aussi peu renseigné que jamais sur l'origine de cette voix qu'il ne tient qu'à chacun d'entendre, et qui nous entretient le plus singulièrement d'autre chose que ce que nous croyons penser.<sup>27</sup>

<sup>24</sup> André Breton, „Manifeste du surréalisme, 1924“, in *Œuvres complètes I*, hrsg. von Marguerite Bonnet (Paris: Gallimard, 1988), 309–46, hier 330.

<sup>25</sup> Karlheinz Barck, „Stildiskurs und Stilkritik in der Perspektive des französischen Surrealismus“, in *Stil: Geschichten und Funktionen eines kulturwissenschaftlichen Diskurselements*, hrsg. von Hans Ulrich Gumbrecht und K. Ludwig Pfeiffer (Frankfurt: Suhrkamp, 1986), 248–68, hier 258.

<sup>26</sup> Breton definiert diesen Schreibprozess als „Automatisme psychique pur par lequel on se propose d'exprimer, soit verbalement, soit par écrit, soit de toute autre manière, le fonctionnement réel de la pensée, en absence de tout contrôle exercé par la raison“, Breton, „Manifeste du surréalisme“, 328.

<sup>27</sup> André Breton, „Second Manifeste du surréalisme, 1930“, in *Œuvres complètes I*, hrsg. von Marguerite Bonnet (Paris: Gallimard, 1988), 775–828, hier 807.

In diesem Sinne etabliert die *écriture automatique* nicht nur einen „anderen Begriff von künstlerischer Praxis“,<sup>28</sup> sie generiert auch ein anderes Verständnis von Autorsubjektivität. Denn das obige Zitat zeigt, dass die surrealistische Methode keineswegs einen von der Autorpsyche abgekoppelten Schreibprozess impliziert. Sie zielt vielmehr darauf ab, das Potential der Stimme auszuloten, die der bewussten Persönlichkeit nicht unmittelbar zugänglich, als Ausdruck des Unbewussten jedoch ein integraler Bestandteil der Autorpsyche ist.

Dass gerade Antonin Artaud die Stimme des Unbewussten über alle künstlerischen Grenzen hinaus ausgeschöpft hat und damit weiter als irgendjemand anderer in die Richtung der surrealistischen Forderung vorgedrungen sei, das menschliche Verstehen neu auszurichten, gesteht Breton 1946 in einer „Hommage à Antonin Artaud“ ein: „Chaque fois, qu’il m’arrive d’évoquer – avec nostalgie – ce qu’a été la revendication surréaliste s’exprimant dans la pureté et dans son intransigeance originelles, c’est la personnalité d’Antonin Artaud, magnifique et noir, qui s’impose à moi“.<sup>29</sup> Indem Breton den Surrealismus in seiner „reinsten“ Form der Persönlichkeit Artauds und nicht seinem Werk zuweist, koppelt er die Frage nach Artauds Zugehörigkeit zum Surrealismus an seine Biographie und seinen ‚Lebensstil‘.

## Schizophrenes Erzählen?

Breton begründet damit eine Lesart des Artaud’schen Werks, in der eine enge Verbindung zwischen Autorpsyche und Erzählstil hergestellt wird. Artauds wechselvolle Biographie lud immer wieder dazu ein, das Verhältnis von Schizophrenie und literarischem Schreiben zu untersuchen und so hat die Artaud- und die Schizophrenie-Forschung von den 1960er Jahren<sup>30</sup> bis heute<sup>31</sup> zahlreiche psychologisch und psychoanalytisch orientierte Studien produziert, die in Artauds Schizophrenie den Motor seines kreativen

<sup>28</sup> Barck, „Stildiskurs und Stilkritik“, 258.

<sup>29</sup> André Breton, „Hommage à Antonin Artaud“, in *Œuvres complètes III*, hrsg. von Marguerite Bonnet (Paris: Gallimard, 1999), 736–9, hier 738.

<sup>30</sup> Vgl. Bernd Mattheus, *jede wahre sprache ist unverständlich: über antonin artaud und andere texte zur sprache veränderten bewusstseins* (Berlin: Matthes & Seitz, 1977), 64.

<sup>31</sup> Vgl. Steven D. Brown, „Collective Emotions: Artaud’s Nerves“, *Culture and Organization* 11 (2005): 235–47; Louis Sass, „Self-disturbance and Schizophrenia: Structure, Specificity, Pathogenesis (Current Issues, New Directions)“, *Schizophrenia Research* 152 (2014): 5–11.

Schaffens oder im Schreiben eine autotherapeutische Maßnahme sehen.<sup>32</sup> Darüber hinaus hat Szymkowiak eine ausführliche, linguistische Studie vorgelegt, in der sie Artauds Bildung von Neologismen mit den Datierungen der Elektroschockbehandlungen in Zusammenhang bringt.<sup>33</sup> Doch gelingt es diesen Herangehensweisen überhaupt, Artauds surrealistische Poetik zu berücksichtigen?

Unter Bezugnahme auf und in Abgrenzung zu den erzählanalytischen Verfahren der Psychologie möchten wir am Beispiel von zwei exemplarischen Textausschnitten Artauds Stil im Spannungsfeld zwischen schizophrenen Merkmalen und einer surrealistischen Schreibweise analysieren. So schreibt er in einem Brief an Parisot vom 6. Oktober 1943:

J'ai vu devant moi de beaucoup de cercueils tomber je ne sais quelle matière noire, quelle immortelle urine de ces muets de vivre, qui miettes de matière, en miettes, goutte par goutte, s'abolissaient. Le nom de cette matière est caca, et caca est la matière de l'âme, dont j'ai vu tellement des cercueils répandre leurs flaques devant moi. Le souffle des ossements a un centre et ce centre est le gouffre Kah-Kah, Kah le souffle corporel de la merde, qui est l'opium d'éternelle survie. Toute la merde sortie de l'entassement de tant de cercueils est un opium arraché qui ne s'était pas encore assez forée dans le gouffre de la fécalité, le focal de sa fécalité. [...]. L'odeur du cu éternel de la morte est l'énergétique opprimée d'une âme à qui l'homme a refusé la vie.

**pho ti ti ananti phatiame**

**fa ti tiame ta fatridi.**<sup>34</sup>

Stilistisch lassen sich hier Anzeichen für eine Sprache des Deliriums und des Wahns finden,<sup>35</sup> die noch dazu durch die zweifache Bezugnahme auf das Rauschmittel Opium verstärkt wird. Auffällig sind die Isotopien des Fäkalen, „urine“, „caca“, „merde“, „cu“, und des Todes, „matière noire“, „cercueils“, „la morte“. Die durch zahlreiche Wortwiederholungen und chiasmatische Figuren bis ins Exaltierte gesteigerte Körpersprache entwickelt dabei eine Eigendynamik, die zwischen der Konkretion der Fäkalsprache

<sup>32</sup> Vgl. Mattheus, *jede wahre sprache*, 66; Laure Westphal, „L'écrivain schizophrène et la construction du délire“, *L'évolution psychiatrique* 82 (2017): 339–50; Gilles Deleuze und Félix Guattari, *L'Anti-Œdipe* (Paris: Minuit, 1972).

<sup>33</sup> Vgl. Céline Szymkowiak, *Langage et schizophrénie: une approche linguistique des Cahiers de Rodez d'Antonin Artaud* (Dijon: A.B.E.L.L., 2002).

<sup>34</sup> Antonin Artaud, „Lettres de Rodez“, in *Œuvres complètes IX* (Paris: Gallimard, 2006), 165–211, hier 174.

<sup>35</sup> Vgl. Westphal, „L'écrivain schizophrène“.

und semantischen Abstraktionen oszilliert. Das sprechende Ich ist in dieser Passage schwer zu lokalisieren. Es erscheint als gebrochene Instanz, die zwar als Beobachter auftritt, gleichzeitig jedoch als Handlungsträger zurückgenommen und der „schwarzen Materie“ ausgesetzt ist. In diesem Zusammenhang lässt sich eine Verbindung zum Leitmotiv des „envoûtement“ herstellen – also der Verzauberung durch die Macht dunkler Kräfte –, das die Briefe aus Rodez thematisch durchzieht. Für eine Gebrochenheit des Subjekts spricht auch das zweifach auftretende Motiv des Abgrunds, „le gouffre“, sowie das der Selbstauflösung, wie in „ces muets de vivre [...] qui s’abolissaient“. Auffällig für Artauds Texte aus Rodez ist dabei die Abstrahierung vom personalisierten „Ich“ zu kollektiven, die Menschheit anklagenden Aussagen, wie hier, dass die Menschen die Fäkalität des Todes und deren energetisches Potential ablehnen. Der vom Autor geprägte Neologismus „fécalité“ gehört zu den sprachlichen Besonderheiten der Passage, ebenso wie die spezifisch Artaud’sche Schreibweise von „cu“ (statt „cul“),<sup>36</sup> onomatopoeischen Ausdrücken wie „Kah-Kah Kah“ und schließlich die rhythmisch basierte Nonsens-Sprache am Schluss der Passage, die auch seinen in Rodez verfassten Essay *Artaud le Môme* kennzeichnet.

Würde man diese Analyse mit den Befunden der psychologischen Erzählanalysen vergleichen, so ließen sich einige Übereinstimmungen feststellen, etwa der Niederschlag von Inkohärenz, Ungeordnetheit und Exzentrizität in dieser Textpassage. Besonders augenfällig ist die De-Kontextualisierung von Referenzen. Dadurch wirkt der Text auf seine Leser verworren, hermetisch und irrational. Wie in vielen von Artauds Schriften aus Rodez entzieht sich die Passage einer „diskursiven Logik“.<sup>37</sup> Was man seinen Texten allerdings nicht bescheinigen kann, ist eine geringe Komplexität und Ideendichte. Im Gegenteil, die meisten seiner Briefe zeichnen sich durch einen hohen Grad an Poetizität aus. Allein die Alliteration von „matière“, „muets“ und „miettes“ spricht für eine poetische Ausgestaltung, die eine figurative Dichte und semantische Komplexität herstellt und dabei die Rhythmik der Sprache berücksichtigt. Schon auf der Textoberfläche greift eine psychologische Deutung also zu kurz.

Dies gilt gleichermaßen für eine biographische Lesart, die im Hinblick auf die Metaphernsprache psychologisierende Deutungen vornimmt. Die blinden Flecken solcher Interpretationen lassen sich besonders an Passagen, in

<sup>36</sup> Anmerkung der Herausgeber in Artaud, „Lettres“, 274.

<sup>37</sup> Vgl. Westphal, „L’écrivain schizophrène“, 345.

denen Artaud aus dem Alltag in der Nervenheilanstalt erzählt, deutlich machen:

J'avais cette nuit un pain sur ma table de nuit, les envoûteuses et envoûteurs du Bd de la Madeleine m'ont envoyé un rat qui s'est introduit dans ce pain, y a fait un trou et l'a de l'intérieur rongé pendant toute la nuit, laissant sur mes livres au matin je ne sais combien de crottes de rats. Et ce n'est pas le premier rat que cette terre envoie sur ma nourriture et sur mes écrits. Car ces rats s'appellent néant et vide [...]. Car Kraum-dam est un corps plein et toujours plein, qui a toujours souffert du souffle brahm, brahma, ama, brahma [...].<sup>38</sup>

Diese Passage lädt auf den ersten Blick zu einer biographisch-pathologisierenden Lesart ein. Denn das Rattenkot, das Artaud morgens in seinen Büchern und seiner Nahrung vorzufinden meint, erscheint als Produkt seiner Wahnvorstellung, von dunklen Kräften verfolgt und verzaubert zu werden. Diesbezüglich diagnostiziert z.B. Westphal dem Autor eine „angoisse dissociative“.<sup>39</sup> Gleichzeitig könnten die Rattenfäkalien symbolisch als Ausdruck des zunehmenden physischen und psychischen Zerfalls des Autors in der Nervenheilanstalt Rodez gedeutet werden. Dass diese Ratten dann auch noch „Nichts und Leere heißen“, könnte auf Artauds Agonie und Todesangst hinweisen. Mit einer solchen Lesart lässt sich die Textpassage jedoch auch nicht gänzlich erfassen. Denn Artaud reiht zwar auf inkohärent-dekontextualisierende Weise Rattenkot, Leere und rhythmische Anspielungen auf den hinduistischen Schöpfungsgott Brahma aneinander. Bezeichnenderweise suchen die Ratten dabei aber nicht nur die körperliche Nahrung heim, sondern auch die geistige, „mes livres“ und „mes écrits“: Repetitionen, Alliterationen und Assonanzen weisen diese Heimsuchung als eine kleinteilige Arbeit am Sprachmaterial aus, die Ausdrucksweisen des Irrationalen und Unbewussten mit sich durch Artauds Werk hindurchziehenden Motiven der Körperlichkeit und des Fäkalen in Verbindung bringt und diese durch Rhythmik und die sonore Qualität des Textes neu zu materialisieren sucht. Dies schlägt sich nicht nur in der rhythmischen Nonsense-Sprache nieder, sondern auch in der Vielzahl der metapoetischen und existenziellen Passagen, die einen ähnlichen Stil in Bezug auf die Motive und die rhythmisch-repetitive Klanglichkeit aufweisen: „Vivre, c'est éternellement de survivre en remâchant son moi d'excrément, sans nul peur de son âme fécale, force affamante d'enterrement. Car toute humanité veut vivre mais elle

<sup>38</sup> Artaud, *Œuvres IX*, 205–6.

<sup>39</sup> Westphal, „L'écrivain schizophrène“, 343.

ne veut pas payer le prix et ce prix est le prix de la peur.“<sup>40</sup> Diese stilistische Arbeit am Sprachmaterial soll laut Artaud zur Gründung einer „poésie fécale“ beitragen, die auf einer körpergesteuerten Triebhaftigkeit basiert und sich auf das Fäkale, das Sexuelle, das Blasphemische und den Tod stützt.<sup>41</sup> „Le corps“, „le sang“, „le souffle“, „le cri“, sind außerdem Leit motive seiner metatheatralischen Überlegungen.<sup>42</sup>

Aus der Untersuchung der Textausschnitte geht hervor, dass Artauds Suche nach einer neuen Sprache des Somatischen sein poetisches Programm des „théâtre de la cruauté“ konsequent weiterführt, mit dem er die Darstellungsformen in der dramatischen Kunst revolutioniert hat.<sup>43</sup> Weder eine psychologisierend biographische oder eine pathologisierend diagnostizierende Lesart können dieser Poetik gerecht werden. Denn das irrationale und inkohärente Erzählen ist Teil einer Schreibweise, die das Unbewusste ausschöpft, indem sie den unergründlichen, sich selbst stets fremden Teil der Persönlichkeit auslotet und die dem Bewusstsein unzugänglichen Stimmen zum Sprechen bringt.<sup>44</sup> Damit verweist Artauds Erzählstil auf eine Poetik, die das surrealistische Programm ganz im Sinne Bretons bis ins Äußerste verwirklicht, indem sie das Potenzial der anderen, unentdeckten Bereiche des Logischen und des menschlichen Verstehens freisetzt. Auch wenn Deutungen, die von Erzählgestalt oder Metaphernsprache direkt auf Autorbiographie und -psyche kurzschließen, an dieser komplizierten Poetik scheitern müssen, bedeutet das aber noch lange nicht, dass diese Faktoren für Artauds surrealistischen Stil unerheblich wären. Denn Artaud führt seine Schreibweise durch Rhythmik, Stilfiguren und Bildsprache immer wieder durch Körper und Psyche hindurch, um zu seiner entgrenzten, das Bewusste übersteigenden Textgestalt zu gelangen. Seine *écriture* läuft also gezielt durch den Barthes'schen *style*, der sich vom Psyche-Soma des Autors nicht ablösen lässt.

<sup>40</sup> Artaud, „Lettres“, 175. Hervorhebung A.G. und M.S.

<sup>41</sup> Artaud, „Lettres“, 174.

<sup>42</sup> Diese Motive tauchen exemplarisch in seinem Brief an den französischen Theaterregisseur und Schauspieler Roger Blin vom 25. März 1946 auf. Vgl. Antonin Artaud, „Lettre à Roger Blin“, in *Ceuvres complètes XI* (Paris: Gallimard, 1974), 215–8.

<sup>43</sup> Vgl. Antonin Artaud, *Le théâtre et son double* (Paris: Gallimard, 2003).

<sup>44</sup> Unter Bezugnahme auf Deleuze und Guattaris *L'Anti-Œdipe* bescheinigt Grossman Artaud eine „Poétique de l'aliénation“, Evelyne Grossman, „Poétique d'aliénation chez Antonin Artaud“, *L'Esprit Créateur* 38 (1998): 36–142.

## Fazit

Die lange literaturtheoretische Tradition, die Autor und Stil miteinander verbindet, lässt sich – trotz der poststrukturalistischen Wende, die auch Artaud für sich veranschlagt hat<sup>45</sup> – nach unseren Überlegungen nicht auflösen. Dies hat schon die Untersuchung der psychologischen Erzählanalysen gezeigt. Einerseits haben wir festgestellt, dass die vorliegenden Modelle für eine literarische Analyse unzureichend bleiben, denn es liegen weder genug Forschung noch einheitliche Ergebnisse vor. Darüber hinaus sind die bisherigen Methoden nicht komplex genug, um literarischem Stil gerecht zu werden. Andererseits lässt sich auf Basis der vorliegenden Ergebnisse nicht von der Hand weisen, dass Persönlichkeit oder Autorpsyche und Stil in enger Wechselwirkung zueinander stehen.

In diesem Zusammenhang ist der surrealistische Stil ein besonders aufschlussreiches Beispiel, weil er mit Schreibweisen wie der *écriture automatique* zwar die *bewusste* Persönlichkeit des Autors aus der literarischen Produktion ausschließen will, dafür aber einen anderen Teil der Psyche – das Unbewusste – zum Sprechen bringt. Deshalb lässt sich gerade die Poetik des Surrealismus nicht vom Einfluss der Autorpsyche lösen. Dies hat sich auch bei Artaud gezeigt. Wir stoßen in seinem Fall auf eine fundamentale Unentscheidbarkeit in der Stilfrage. Welche Aspekte seines Stils auf biographisch-persönliche Ursachen – Delirium, Schizophrenie, Drogenkonsum, auf die surrealistische Programmatik – das Ausloten des Unbewussten, oder auf seinen individuellen poetischen Gestaltungswillen zurückzuführen sind, lässt sich letztlich nicht klären. Die Stilfrage bei Artaud ist gerade deswegen besonders spannungsreich, weil Psycho-Somatik und Poetik eng ineinander verweben sind.

Damit ist auch die Frage nach einer allgemeinen Theorie des Personalstils von Unentscheidbarkeit geprägt. Wir können die Autorpsyche nicht ausklammern, aber wir besitzen auch kein Analysewerkzeug, um ihren Einfluss von genuin literarischen Interessen abzutrennen. Greifen wir zu Methoden, die sich auf einen bestimmten Aspekt des Stils konzentrieren, wie stilometrische oder psychologische Analysen, erhalten wir also nur bedingt aussagekräftige Ergebnisse, es sei denn, wir führen sie wieder in ein

---

<sup>45</sup> Vgl. hierzu vor allem auch die Überlegungen von Jacques Derrida zu Artauds „théâtre de la cruauté“, in *L'écriture et la différence* (Paris: Seuil, 1967), 341–68; und seinen legendären Vortrag zur Ausstellungseröffnung *Antonin Artaud: Works on Paper*, die 1996 im MOMA in New York stattfand; Jacques Derrida, *Artaud le Moma* (Paris: Galilée, 2002).



kontextgetragenes, genaues *close reading* zurück. So können wir immerhin auf die Stellen hinweisen, an denen Körper, Psyche und Poetik in ein enges Verschränkungsverhältnis treten. Diese Unentscheidbarkeit muss aber nicht notwendigerweise das Problem darstellen, als das es so lange gehandelt wurde. Die Faszination, die der Individualstil auf uns ausübt, speist sich schließlich auch daraus, dass er überdeterminiert ist und damit immer ein Stück weit unmessbar und unkontrollierbar bleibt.



# Interkulturalität und Autorschaft

## Tamara Jadrejić und Alida Bremer

Marijana Erstić (Siegen)

**ZUSAMMENFASSUNG:** Schriftstellerinnen und Schriftsteller wie Alida Bremer (geb. 1959) und Saša Stanišić (geb. 1978) katapultierten sich in den vergangenen Jahren durch ihre literarischen Auseinandersetzungen mit dem Zerfall Jugoslawiens oder das Verhandeln des ‚Gastarbeiter‘-Themas an die Spitze der deutschsprachigen Kulturlandschaft. Neben einer Neudefinition der ‚Interkulturalität‘ fordern sie auch eine der ‚Autorschaft‘ ein. Die beiden Begriffe verweisen nämlich nicht notwendig auf Menschen mit nur einer Muttersprache. Doch Deutschland ist kein Einzelfall. So schreibt und veröffentlicht die aus Kroatien stammende, zwischen 1992 und 2004 in Italien und derzeit in den USA lebende Autorin Tamara Jadrejić (geb. 1964) erfolgreich seit Jahren vorwiegend auf Italienisch, aber auch auf Englisch und auf Kroatisch. Im folgenden Beitrag wird anhand der Texte von Jadrejić und Bremer das Problem der Autorschaft, der Sprache und der Identität thematisiert.

**SCHLAGWÖRTER:** Interkulturalität; Autorschaft; Jadrejić, Tamara; Bremer, Alida

**ABSTRACT:** In recent years, literary authors such as Alida Bremer (born in 1959) and Saša Stanišić (born in 1978) have catapulted themselves into the forefront of the German-speaking cultural landscape through their literary examinations of the collapse of Yugoslavia or their treatment of the ‚guest worker‘ topic. Besides a new understanding of ‚interculturality‘, they also demand a reexamination of ‚authorship‘; after all, these terms don't necessarily refer to people with only one mother tongue. But Germany is not an isolated case. The Croatian author Tamara Jadrejić (born in 1964), who lived in Italy between 1992 and 2004 and currently lives in the USA, writes and publishes successfully in Italian, but also in English and Croatian. The article discusses the problem of authorship, language and identity on the basis of texts by Jadrejić and by Bremer.

**KEYWORDS:** interculturality; authorship; Jadrejić, Tamara; Bremer, Alida

## Von der Migrantenliteratur zur Postnationalität

Die Welle an deutschsprachigen, aus den Ländern Ex-Jugoslawiens stammenden Autorinnen und Autoren scheint derzeit nicht nachzulassen. Literaten wie Alida Bremer (geb. 1959), Marica Bodrožić (geb. 1973), Saša Stanišić (geb. 1978), Nicol Ljubić (geb. 1971), Adriana Altaras (geb. 1960) oder Filmemacher wie Damir Lukačević (geb. 1966) gehören aktuell zu den renommierten

zeitgenössischen deutschsprachigen Schriftstellerinnen und Schriftstellern und bilden einen Großteil des z. Z. herrschenden ‚eastern turn‘. Neben einer Neudefinition der ‚Interkulturalität‘ erwarten sie auch eine Neudefinition der ‚Autorschaft‘, schreiben sie doch nicht ausschließlich nur in ihrer Muttersprache. So sprach man im Jahr 2017 auf einer komparatistischen Tagung in Zadar von Postnationaler Literatur.<sup>1</sup>

Einen Beweis für die Bedeutung dieser zeitgenössischen, postnationalen Werke liefern aber auch die internationalen wissenschaftlichen Publikationen, die sich den besagten Narrativen zumeist derart nähern, dass sie eine Neudefinition der ‚Interkulturalität‘ einfordern.<sup>2</sup> Denn die Erzähler aus der zweiten oder gar dritten ‚Ausländer‘-Generation, die ehemaligen Kriegsflüchtlinge, die Kinder aus deutsch-südosteuropäischen Ehen mussten das Fremde und das Eigene, die Sprachen und die Länder offener, ‚interkultureller‘ denken. Was früher als ‚Gastarbeiter-‘ und ‚Migrantenliteratur‘ bezeichnet werden konnte, entzieht sich nunmehr den meisten Klassifizierungsversuchen. Dies zeigt beispielsweise der Roman *Vor dem Fest* (2014) von Saša Stanišić, der die Stimmung vor einem Fest in der Uckermark auffängt.<sup>3</sup> Während die ersten Werke der angeführten Literaten größtenteils noch den Krieg zum Thema hatten (bei Stanišić z. B. der Roman *Wie der Soldat das Grammophon repariert* aus dem Jahr 2007),<sup>4</sup> kommen in den späteren Werken die Interkulturalität und der Kosmopolitismus zum Vorschein.

Doch Deutschland ist kein Einzelfall. So schreibt und veröffentlicht die aus Kroatien stammende, zwischen 1992 und 2004 in Italien und derzeit in den USA lebende Autorin Tamara Jadrejić (geb. 1964) erfolgreich seit Jahren auf Italienisch, Englisch und Kroatisch. Im Beitrag wird anhand eines italienischsprachigen Textes von Tamara Jadrejić und eines deutschsprachigen Textes von Alida Bremer analysiert, wie sich die diagnostizierte Interkultura-

<sup>1</sup> Vgl. demnächst: Marijana Jeleč, Goran Lovrić und Slavija Kabić, *Die Darstellung Südosteuropas in der Gegenwartsliteratur* (Frankfurt am Main: Peter Lang, 2018).

<sup>2</sup> Vgl. z. B. Sara Michel, „Identitätskonstruktionen und Essensdarstellungen in der Migrationsliteratur am Beispiel von Aglaja Veteranyis Roman ‚Warum das Kind in der Polenta kocht‘ und Saša Stanišićs ‚Wie der Soldat das Grammophon repariert‘“, *Germanica* 57 (2015): 139–57; Goran Lovrić, „Kontinuitäten und Diskontinuitäten in deutschsprachigen Familiennarrativen kroatischstämmiger Autorinnen“, in *Familie und Identität in der Gegenwartsliteratur*, hrsg. von Goran Lovrić und Marijana Jeleč (Frankfurt am Main: Peter Lang, 2016), 85–101 und Marijana Erstić, „Vielfalt in einem deutschsprachigen ‚Dalmatien‘-Narrativ: der Roman ‚Olivas Garten‘ (2013) von Alida Bremer“, *Diagonal: Zeitschrift der Universität Siegen* 37 (2016): 45–52.

<sup>3</sup> Saša Stanišić, *Vor dem Fest* (München: btb, 2014).

<sup>4</sup> Saša Stanišić, *Wie der Soldat das Grammophon repariert* (München: btb, 2006).

lität bzw. hybride Nationalität/Postnationalität auf die besprochenen Werke und die Autorschaft auswirkt.

So werden im ersten Teil des Aufsatzes der Autor und die Autorschaft hinterfragt. Es folgt ein Kapitel über Tamara Jadrejčićs *raccolta I prigionieri di guerra* und Jürgen Habermas' (geb. 1929) Konzept der ‚postnationalen Identität‘. Danach wird der Roman Alida Bremers *Olivas Garten* und Homi Bhabhas (geb. 1949) Konzept des ‚Dritten Raumes‘ (‚third space‘) erörtert. Zum Schluss werden einige Fragen an die Stilometrie gestellt.

## Die Figur des Autors

Der Autor, von lat. *auctor* – d. h. Gewährsmann, Vorbild etc. –, ist eine der wichtigsten Größen innerhalb der Literaturwissenschaft. Hinter dem Namen eines Autors verbergen sich Informationen, wie die historische Einordnung bei Toten bzw. das Alter bei den Lebenden und natürlich auch die Werke, aber auch das Geschlecht und die Herkunft bzw. die Nationalität. Gleichwohl: Autoren, die in einer ‚Fremdsprache‘ schreiben, hat es immer schon gegeben. Das bekannteste Beispiel ist sicherlich der aus Polen stammende und auf Englisch schreibende Joseph Conrad (1857–1924). Die Fülle der Autoren aus Osteuropa jedoch, die in den westlichen Ländern in der Sprache des Gastlandes schreiben, ist so groß, dass sie heutzutage ein bemerkenswertes Phänomen darstellt.

Der Autor hat also auch eine ordnende Funktion, denn sein Name verrät in der Regel nicht nur seine Herkunft, sondern dient auch der Katalogisierung. Er ist zudem zentraler Bezugspunkt bei der Textinterpretation, und er ist der Hauptakteur in den Literaturgeschichten. Er ist der Hauptakteur auch eines sehr beliebten Mediums: Der Autorenlesung. Und natürlich: das Konzept des Autors und der Autorschaft ist relativ neu, es existiert in Deutschland seit dem 18. Jahrhundert, seit der sog. Genieästhetik, also auch seit dem Selbstverständnis der Dichter des Sturm und Drang (1767–1785). In Folge dieses Umbruchs entsteht der Autorenschutz bzw. der Schutz des geistigen Eigentums, mit allen seinen Konsequenzen, wie z.B. den Plagiatsaffären, die wir heute in ganz Europa kennen. Heinrich Bosse (geb. 1937) sprach z.B. von der Autorschaft als einer Werkherrschaft.<sup>5</sup> Das heißt, das Thema der Autorschaft hat noch lange nicht ausgedient, auch wenn der französische Philosoph und Soziologe Roland Barthes (1915–1980) bereits 1967 zum Umdenken

<sup>5</sup> Heinrich Bosse, *Autorschaft ist Werkherrschaft: über die Entstehung des Urheberrechts aus dem Geist der Goethezeit* (München: Fink, 2014).

hinsichtlich des Autors aufgefordert hat und bekanntlich den Tod des Autors ausrief.<sup>6</sup> Barthes' Aufsatz ist in dem amerikanischen Pop-Magazin *Aspen* erstmals erschienen („The Death of the Author“).<sup>7</sup> Die Überlegungen von Roland Barthes lassen sich zwar sehr gut auf die Literatur der sog. Postmoderne übertragen und auf ihr Spiel mit den Zitaten, vielleicht überhaupt auf einen Großteil der europäischen Literatur im 20. Jahrhundert. Doch der Text von Barthes ist und bleibt eine Provokation. Nach einer Irritationsphase wird heute niemand ernsthaft aus dem Unterricht oder der Lehre gänzlich den Autor ausschließen. Wir brauchen vielmehr den Autor, um die Texte, so Fotis Jannidis (geb. 1961) in der Einleitung zu seinen im Jahr 2000 beim Reclam-Verlag erschienenen Aufsätzen zur Theorie der Autorschaft, historisch so zu kontextualisieren und sie in literaturhistorische Zusammenhänge so einzubetten, dass sie verstehbar seien.<sup>8</sup> Mehr noch, zum Mythos einzelner Bestseller und zum Mythos einzelner Autoren hat die Biographie nicht wenig beigetragen.

Doch der sog. *eastern turn* erweitert die Problematik der Autorschaft. Eine Zäsur im Leben vieler im Ausland lebender, in Fremdsprachen schreibender Autoren aus Südosteuropa oder aus Kroatien war der Jugoslawische Nachfolgekrieg 1991–1995. Dass diese Zäsur in den Werken thematisiert wird, zeigt die Verflechtung von Autorbiographie und Narrativ.

## Tamara Jadrejić und die Postnationale Identität

Die Autorin und examinierte Literaturwissenschaftlerin Tamara Jadrejić ist 1992 aus Kroatien nach Italien ausgewandert. Mittlerweile lebt und schreibt sie in den USA und veröffentlicht auch Texte in den kroatischen Tageszeitungen – viel und gerne über Donald Trump. Ihre Erzählensammlung *I prigionieri di guerra* ist im Jahr 2007 auf Italienisch erschienen.<sup>9</sup> Am Beginn der Erzähl-

<sup>6</sup> Roland Barthes, „Der Tod des Autors“, *Texte zur Theorie der Autorschaft*, hrsg. von Fotis Jannidis u. a. (Stuttgart: Reclam, 2000), 185–93.

<sup>7</sup> Roland Barthes, „The Death of the Author“, *Aspen* 5–6 (1967), item 3, auch unter: [www.ubu.com/aspens/aspens5and6/threeEssays.html#barthes](http://www.ubu.com/aspens/aspens5and6/threeEssays.html#barthes) (06.02.2018).

<sup>8</sup> Fotis Jannidis, „Autor und Interpretation“, in *Texte zur Theorie der Autorschaft*, 7–29, hier 7.

<sup>9</sup> Tamara Jadrejić, *I prigionieri di guerra* (San Giovanni in Persiceto (Bo): Eks & Tra, 2007). Übersetzung der Vf.in: „Migrantenliteratur ist ein wachsendes Phänomen, und es muss ernst genommen werden. Aber nicht zu ernst. Migrantenliteratur wird immer mehr zum Teil der italienischen Literaturszene, kein separates Ding. Nach dem ‚Eks & Tra‘-Preis für die Geschichte ‚Der Junge, der sich nicht wusch‘ wollte ich mich mit italienischen Schriftstellern vergleichen. Ich habe meine Texte an den Italo Calvino Wettbewerb geschickt, weil ich nicht nur eine ‚fremde‘ Schriftstellerin bleiben wollte“.

sammlung steht die Erzählung „Il bambino che non si lavava“ = „Ein Junge, der sich nicht wusch“, für die Jadrejić 2001 den ersten Preis beim Wettbewerb des Verlages Eks & Tra bekam. Danach entstanden weitere Geschichten, die den Kroatienkrieg 1991–1995 zum Thema hatten. 2003 wurden die Erzählungen mit dem Italo Calvino-Preis ausgezeichnet. Die Autorin dazu in einem Interview:

La scrittura migrante è un fenomeno che comunque sta crescendo e che bisogna prendere sul serio. Però non troppo. Scrittura emigrante sta diventando parte della scena letteraria italiana, non una cosa a parte. Dopo il premio Eks & Tra per il racconto „Il Bambino che non si lavava“ volevo mettermi a confronto con scrittori italiani. Mandai i miei testi al concorso Italo Calvino, perché non volevo rimanere solo una scrittrice „straniera“.<sup>10</sup>

Die Sammlung (oder *la raccolta*) *I prigionieri di guerra* folgt gut vier Jahre nach dem Calvinopreis beim Verlag Eks & Tra und umfasst sieben Erzählungen. Sie basiert dem Klappentext zufolge auf persönlichen Erfahrungen. Das Personal des Sammelbandes/Episoden-Buches wechselt. Die eigentliche Protagonistin der Geschichten bleibt jedoch immer präsent – es ist die Angst. Die eines Jungen, der um die Rückkehr des Vaters von der Front bangt. Oder die der Soldaten, die im Krieg verlassene und leerstehende Häuser ausrauben und Angst vor den möglichen Kosequenzen haben. Oder die eines Ehepaares, das um seine Ersparnisse für die ‚alten Tage‘ bangt, wobei die alten Tage ja bereits eingetreten sind. Und schließlich geht es auch um eine Frau, die Angst vor ihrem aus dem Krieg zurückgekehrten, gewalttätigen Mann hat. All diese Menschen hat im Krieg niemand körperlich gefangengenommen. Und dennoch sind sie alle vom Krieg und von seinen Konsequenzen in Geiselhaft genommen worden. Sie sind damit *i prigionieri di guerra*: Für sie alle gibt es keine sicheren Zufluchtsorte, sie bleiben Gefangene ihrer Angst und somit auch des Krieges.

Der Text von Jadrejić wurde von der Italianistin Tatjana Peruško (geb. 1966) ins Kroatische übersetzt und erschien im Jahr 2008 bei dem Verlag Naklada Ocean more.<sup>11</sup> In der Tagespresse wurden 2007 das Original und die Übersetzung gefeiert. Eine deutsche Übersetzung ist nie erschienen.

Wie ungefähr gleichzeitig die auf Deutsch schreibenden und aus Südosteuropa stammenden Autoren Marica Bodrožić, Saša Stanišić, Nicol Ljubić

<sup>10</sup> [www.barbararomagnoli.info/ex-jugoslavia-la-guerra-vista-dal-tinello-di-casa-intervista-a-tamara-jadrecic/](http://www.barbararomagnoli.info/ex-jugoslavia-la-guerra-vista-dal-tinello-di-casa-intervista-a-tamara-jadrecic/), aufger. am 06.02.2018.

<sup>11</sup> Tamara Jadrejić, *Zarobljenici rata* (Zagreb: Ocean more, 2008).

etc. hat auch Tamara Jadrečić als den Ausgangspunkt ihrer Geschichten den Jugoslawischen Nachfolgekrieg gewählt. Das Thema scheint sowohl in Italien als auch in Deutschland in den 2000er Jahren ein gewisser Erfolgsgarant gewesen zu sein. Die Autoren präsentieren sich damit eben nicht als *prigioniera di guerra*, vielmehr widersetzen sie sich der ‚Gefangennahme‘ durch den Krieg, indem sie in einer anderen Sprache die Erfahrungen anbieten, international und postnational agieren, wachrütteln, intervenieren.

Der Begriff ‚postnationale Identität‘ geht u.a. auch auf Jürgen Habermas zurück. Habermas hat seinen Vortrag „Die postnationale Konstellation und die Zukunft der Demokratie“ bereits am 5. Juni 1998 vor dem „Kulturforum der Sozialdemokratie“ in Berlin gehalten.<sup>12</sup> Im Vortrag beleuchtet Habermas die wirtschaftlichen und sozialen Folgen, Chancen und Risiken der Globalisierung im Rahmen der Europäischen Union. Es gehe ihm dabei u.a. auch

um die Frage, ob die Europäische Union den Verlust an nationalstaatlichen Kompetenzen überhaupt wettmachen kann. Als Testfall gilt die Dimension einer umverteilungswirksamen Sozialpolitik. Diese Frage der Handlungsfähigkeit hängt mit der weiteren, allerdings analytisch zu unterscheidenden Frage zusammen, ob [...] politische Gemeinschaften eine kollektive Identität jenseits der Grenzen einer Nation ausbilden und damit Legitimitätsbedingungen für eine postnationale Demokratie erfüllen können.<sup>13</sup>

Denn, so Habermas, „Wenn diese beiden Fragen keine affirmative Antwort finden, ist ein europäischer Bundesstaat nicht möglich“. <sup>14</sup> Auch wenn es Habermas vor allem um die politischen, wirtschaftlichen und sozialen Kategorien der Post-Nationalität geht, übertragen kann man dies auch auf die Ebene der Kultur. Mehr noch, es scheint, dass die Kultur wieder einmal jene Avantgarde- und Vorreiterstellung als Vorbild für die wirtschaftliche, soziale und politische Ebene einnimmt und einnehmen soll. Gleichwohl: die Tatsache, dass die Verhandlung des Jugoslawienkrieges als Publikationsgarant gedient hat, zeigt nicht nur die Erwartungen Westeuropas, sondern ebenso, wie sehr die Herkunft und die Nationalität auch jenseits der Sprache in das Narrativ hineinbrechen. Ganz so, als zeige sich die Autorschaft nicht nur als Werkherrschaft oder als Nationalsprache, sondern v.a. als ein Rückgriff auf das Themenrepertoire des ehemaligen Heimat- oder Herkunftslandes. Und

<sup>12</sup> Jürgen Habermas, „Die postnationale Konstellation und die Zukunft der Demokratie“, <http://library.fes.de/pdf-files/akademie/online/50332.pdf>, aufger. am 06.02.2018. Vgl. auch Habermas, *Postnationale Konstellation* (Frankfurt am Main: Suhrkamp, 1998).

<sup>13</sup> Habermas, „Die postnationale Konstellation“, 136.

<sup>14</sup> Habermas, „Die postnationale Konstellation“, 136.



doch agieren die mehrsprachlichen und mehrstaatlichen Autoren international.

Tamara Jadrečić hat nach *Prigionieri die guerra* kein Buch mehr auf Italienisch veröffentlicht. Sie lebt heute als Journalistin und Übersetzerin in den USA.

Ihre bereits erwähnten Schriftstellerkollegen aus Deutschland, Marica Bodrožić, Saša Stanišić, Nicol Ljubić haben längst weitere Themen für ihre Romane und Erzählungen gefunden. Dabei muss angemerkt werden, dass die meisten dieser (noch) jungen deutschen Kolleginnen und Kollegen im Rahmen der Vorbereitung ihrer Bücher das Grenzgänger-Stipendium der Robert-Bosch Stiftung erhalten haben. Die Thematisierung des Jugoslawischen Nachfolgekrieges wurde in den 2010er Jahren von der Bosch-Stiftung mehrmals prämiert und bezuschusst.

### Alida Bremer und der Dritte Raum

Eine Ausnahme bildet die Autorin Alida Bremer. Auch sie war Bosch-Stipendiatin. Auch ihr Text behandelt den Krieg. Aber ihr Werk *Olivas Garten*<sup>15</sup> (erschienen beim Ullstein-Verlag 2013 als Hardcover und 2016 als TB) thematisiert einen anderen Krieg – den Zweiten Weltkrieg. Das Werk ist der erste Roman dieser kroatischstämmigen und auf Deutsch schreibenden Autorin, die in Deutschland nicht nur als Schriftstellerin, sondern auch als Spiegelkolumnistin, promovierte Literaturdozentin, Literaturscout und Übersetzerin bekannt ist, und die jahrelang an der kroatischen Präsentation während der Buchmessen in Leipzig und in Frankfurt am Main mitgewirkt hat. Im Jahr 2008 hat sie den Länderschwerpunkt *Leipzig liest kroatisch* geleitet. Ihr erster Roman ist weitestgehend autobiographisch geprägt.

So heißt die Hauptfigur und Ich-Erzählerin des Romans ebenfalls Alida, auch sie stammt aus Dalmatien, auch sie ist seit Jahrzehnten mit einem deutschen Mann verheiratet und lebt, wie die Autorin, in Münster, die Eltern wohnen jeweils in Split.

Anders freilich als die Autorin (der Leser/die Leserin erfährt das aus dem Nachwort) erbt ihr Alter Ego zu Beginn des Romans einen Olivenhain von der Großmutter Oliva, Olivas (Oliven-)Garten eben. So entspinnen sich um das Erbe, die Herkunft, das Heimatland Geschichten der Gegenwart und Erinnerungen an das Vergangene, erzählt als die Geschichte(n) einer Familie im Zwanzigsten und zu Beginn des Einundzwanzigsten Jahrhunderts.

<sup>15</sup> Alida Bremer, *Olivas Garten* (Köln: Eichborn, 2013).

Bezüge zu Italien sind dabei sehr stark vorhanden – in der erzählten Geschichte, aber auch als Inspirationsquelle. Die Familie er- und überlebt innerhalb der erzählten Zeit sieben Staatsformen: die K. u. K.-Monarchie, den Staat der Serben, Kroaten und Slowenen, das Königreich Jugoslawien, die italienische/deutsche Besatzung bzw. den (faschistischen) Unabhängigen Staat Kroatien, die Sozialistische Föderative Republik Jugoslawien und die heutige Republik Kroatien. Ferner fünf politische Systeme: österreichisch-ungarische Monarchie, absolutistische Monarchie im Königreich Jugoslawien, faschistische Diktatur, kommunistische Diktatur und schließlich die Demokratie im heutigen Kroatien.

Schließlich auch drei Kriege: den Ersten und den Zweiten Weltkrieg sowie den Kroatienkrieg (der im Roman nur sporadisch erwähnt wird). Und sie erfährt eine schier unzählige Reihe an Schicksalsschlägen wie auch an Triumphen.

Mit diesem historischen Aufriss und mit der geradezu mythischen Komponente des Textes, der der geschilderten Gegend Dalmatiens inmitten aller Veränderungen eine hohe Vitalität bescheinigt, nähert sich die Autorin Bremer den großen sizilianischen Erzählern Giovanni Verga (1840–1922) und Giuseppe Tomasi di Lampedusa (1896–1957) an, auf die sie auch in ihren weiteren Werken immer wieder verweist. Buchstäblich kommt dies in der Geschichte „Der Gattopardo“<sup>16</sup> (2015) zum Vorschein, die denselben (deutschen) Titel trägt wie die letzte Übersetzung des Bestsellers von Tomasi di Lampedusa (2011).<sup>17</sup> Und fast würde in Bremers Texten, so auch in *Olivas Garten*, der Eindruck des Anachronistischen und Epigonenhaften entstehen, wenn es nicht den Witz und die Leichtigkeit der aktuellen Blickweise gäbe. Das zeigt sich in den Gesprächen mit dem deutschen Ehemann, aber auch in den Auseinandersetzungen mit der Bürokratie: Denn die Besitzverhältnisse des Olivenhains und der weiteren Liegenschaften der Familie sind zu Beginn schier unentwirrbar.

Auf der zweiten Erzählebene wird in Rückblicken, die wie Erinnerungsstücke scheinbar zufällig angeordnet sind, die Geschichte der dalmatinischen

<sup>16</sup> Alida Bremer: „Der Gattopardo“, *Lichtungen: Zeitschrift für Literatur, Kunst und Zeitkritik* XXXVI, Bd. 144 (2015): 5–7.

<sup>17</sup> Giuseppe Tomasi di Lampedusa, *Der Gattopardo*, hrsg. von Gioacchino Lanza Tomasi, aus dem Italienischen neu übersetzt und mit einem Glossar von Giò Waeckerlin Induni (Zürich: Piper, 2011). Vgl. auch Marijana Erstić, „Luchino Visconti: ‚Il Gattopardo‘ (1963)“, *Italienischer Film des 20. Jahrhunderts in Einzeldarstellungen*, hrsg. von Andrea Grewe und Giovanni di Stefano (Berlin: Erich Schmidt Verlag, 2015), 153–67.

Familie vor, während und nach der italienischen Besetzung Dalmatiens im Zweiten Weltkrieg (1941–1943) erzählt.

Als bruchstückhafte Erinnerung, in mal längere, mal kürzere Kapitel unterteilt, sprunghaft und ungewiss, korrespondiert der Text hier mit den historischen und aktuellen Gedächtnistheorien, namentlich mit jenen von Henri Bergson (1859–1941) und Aleida Assmann (geb. 1947). Hier wie dort ein Erinnern, das sich als virtuell und nie ganz zuverlässig entpuppt und manchmal an Träume gekoppelt ist.<sup>18</sup> Hier wie dort eine ‚rumorende Erinnerung‘, die das im Zweiten Weltkrieg Geschehene nicht zu einer leeren Formel erstarren lässt.<sup>19</sup>

Dabei wird ein Tabuthema der Geschichtsschreibung geschildert: Während des Zweiten Weltkrieges und nach dem Zusammenfall des Königreichs Jugoslawien wird in Kroatien im Jahr 1941 das von den Nationalsozialisten, v.a. aber von den italienischen Faschisten geförderte kroatische Ustascha-Regime installiert. Die kroatischen Faschisten überlassen dem faschistischen Italien im Gegenzug für dessen Unterstützung einen Großteil der Küste Dalmatiens und die Inseln.<sup>20</sup> Wieder einmal ändern sich die Machtstrukturen auf dem Gebiet dramatisch. In Dalmatien und auf den Inseln entsteht eine starke Widerstandsbewegung, an der, dem autobiographischen Roman zufolge, auch Olivas Familie beteiligt ist. Die Frauen, u.a. auch Oliva, geraten im Zuge des Krieges in italienische bzw. deutsche Lager, Teile der Familie landen in El Shatt in Nordafrika. Bremer sagt hierzu: „Erinnerung mischt sich mit den Erzählungen von anderen. Das alles wurde so oft erzählt, dass es gar nicht mehr wahr ist“, auch deshalb habe sie die Romanform gewählt.<sup>21</sup>

Die titelgebende Oliva hat sich nach der Rückkehr aus dem deutschen Lager auf ihre Ottomane gelegt, die sie bis zu ihrem Tode kaum mehr verlässt.<sup>22</sup>

---

<sup>18</sup> Vgl. Bremer, *Olivas Garten*, 142–7 und Henri Bergson, *Materie und Gedächtnis: eine Abhandlung über die Beziehung zwischen Körper und Geist*, Philosophische Bibliothek Bd. 441 (Hamburg: Felix Meiner, 1991), 66ff.

<sup>19</sup> Vgl. Aleida Assmann, *Erinnerungsräume: Formen und Wandlungen des kulturellen Gedächtnisses* (München: Beck, 2009), 329–37.

<sup>20</sup> Vgl. Ludwig Steindorff, *Kroatien: vom Mittelalter bis zur Gegenwart* (Regensburg: Pustet Verlag, 2001), 173–89.

<sup>21</sup> Nadine Lange, „Olivas Garten‘ von Alida Bremer: der Traum von Jugoslawien“, *Der Tagesspiegel* (19.11.2013), [www.tagesspiegel.de/kultur/olivas-garten-von-alida-bremer-der-traum-von-jugoslawien/9099192.html](http://www.tagesspiegel.de/kultur/olivas-garten-von-alida-bremer-der-traum-von-jugoslawien/9099192.html).

<sup>22</sup> Bremer, *Olivas Garten*, 9.

Sie ist, anders als andere Familienmitglieder, während des Zweiten Weltkrieges unpolitisch und überfordert, nach dem Krieg traumatisiert und depressiv. Ihr Oliven-Garten ist die dalmatinische Landschaft schlechthin, ihr Schicksal eine an den Mythos des Ortes gebundene Metapher, die für den ewigen Schlaf der mediterranen Gegend steht. Das Olivenhain-Erbe, jenes reale der Erzählerin, wie auch jenes fiktionale der Autorin, ist die Kulturgeschichte Dalmatiens.

Das wird bis in die Schrift hinein verdeutlicht: Die Kapitel, die Oliva gewidmet sind, erscheinen in olivfarbener Schrift, sie sind deutlich abgesetzt, vielleicht in Anlehnung an die rote Schrift Hélène Cixous (Alida Bremer ist promovierte Literaturwissenschaftlerin und Romanistin). Gedanken, Gefühle, Erinnerungen, mediterrane Kochrezepte wechseln sich ab.

All das repräsentiert eine Zeit, die stehen bleibt, weil das Ewiggleiche in immer neue Uniformen verkleidet wiederkehrt. Es verwundert kaum, dass Oliva „1991 – als ein neuer Krieg ausbricht – verwirrt [ist], da sie glaubt, in die Vergangenheit versetzt worden zu sein“.<sup>23</sup> Im Text heißt es später:

Oliva erinnert sich, wie die italienischen Soldaten mit Paradeschritt durch Vodice marschierten, sie trugen Federn an ihren Mützen und rasselten mit ihren Gewehren. Und sie erinnert sich an die Frankovci, daran, wie sie H-L-A-P riefen: ‚Hoch lebe Ante Pavelić‘ [Ustascha-Führer, Anm. d. Verf.]. Sie erinnert sich, dass dieser Ante Pavelić schwarz trug, wie die Italiener, die das Haus ihres Bruders in Brand gesetzt haben. Es riecht nach Brand. Und dann erinnert sie sich an die deutschen Soldaten, die die Frauen auf das Schiff trieben. Und sie erinnert sich an die Tschetniks, mit ihren langen Haaren und hämischen Gesichtern. Und dann erinnert sie sich an die Partisanen, wie sie nach der Befreiung tanzten, und die Handgranaten, die an ihren Ledergürteln hingen und gefährlich hüpften. Oliva will sich an all das nicht erinnern. Sie will schlafen. Ich bin Dalmatien, träumt Oliva. Ich bin das kroatische Dalmatien, das in der alten romanischen Sprache Dalmato Gebete spricht, die ich nie gehört habe.<sup>24</sup>

Es folgt ein Vaterunser in der ausgestorbenen romanischen Sprache – dem erwähnten Dalmatischen/dem Dalmato:

Tuota nuester, che te sante intel sil, siat santificout el naun to. Vigna el raigno to. Sait fuot la voluntuot toa, coisa in sil, coisa in tiara. Duote costa dai el pun nuester cotidiun. E remetiaj le nuestre debete, coisa nojiltri remetiaime a i nostri debetuar. E naun ne menur in tentatiaun, mia deleberiajne del mal.<sup>25</sup>

<sup>23</sup> Bremer, *Olivas Garten*, 9.

<sup>24</sup> Bremer, *Olivas Garten*, 282.

<sup>25</sup> Bremer, *Olivas Garten*, 282.

Dalmatien, Olivas Sprache, sind, Alida Bremer zufolge, ohnehin nie einsprachige, rein nationale Räume gewesen. *Olivas Garten* bietet aber auch einen rezenten Vergleich zwischen der mediterranen Lebensweise Dalmatiens/Kroatiens und der Lebensweise des Westens/Deutschlands. Als eine Brücke zwischen den beiden Welten fungiert die Ich-Erzählerin.<sup>26</sup> Zum Schluss erbt Alida Olivas Garten und entschließt sich dazu, ihn zusammen mit ihrem deutschen Ehemann zu behalten und zu kultivieren, auch weitere Familien-Grundstücke werden nicht verscherbelt. Wie in dem Roman *Il Gattopardo/Der Leopard/Der Gattopardo* muss auch hier sich alles ändern, damit alles beim Alten bleibt.<sup>27</sup> Die thematisierten Schicksale beweisen bei Jadrečić und mehr noch bei Bremer die Bedeutung der Autobiographie für das Narrativ.

Doch die Texte dieser Autorinnen werfen nicht nur Fragen nach den autobiographischen Elementen auf, sondern vielmehr, was es heißt, nicht in der Muttersprache zu schreiben. Sie fragen danach, was überhaupt die Muttersprache sei. Wird sie, die Muttersprache, in eine Fremdsprache übersetzt, oder die Erstsprache in eine Zweitsprache? Ist nicht, wie Jacques Derrida (1930–2004) sagte, jede Übersetzung zugleich eine Dekonstruktion,<sup>28</sup> die die Autoren des *eastern turn* als postnationale Dekonstruktivisten auszeichnet? Oder bedeutet die Zweisprachigkeit die Ausformung eines dritten Raumes, eines *third space*? Vielleicht weisen diese Autorinnen auf einen fundamentalen Wandel im Umgang mit den Sprachen hin, hin zu einer verstärkten Globalisierung... Denn die Wahl der Zweit-Sprache bedeutet einen Schritt in die Richtung einer postnationalen Identität, die es ermöglicht, das thematisierte Schicksal weiter zu reichen, lesbar zu machen. Erst kürzlich haben in diesem Sinne in der Zeitschrift *Die Horen* der deutsche Schriftsteller Jan Wagner und sein mazedonischer Kollege Nikola Madžirov gemeinsam über den Kosmopolitismus und das Überbrücken der neuen Mauern, über die Macht der Sprache und die Möglichkeit der Übersetzung nachgedacht und über das Literarische in einem Europa, dass ihnen als „eine Konstante der literarischen Sicherheit“ erscheint, fanden doch in Europa auch in den letzten Jahr-

---

<sup>26</sup> Bremer, *Olivas Garten*, 315–6.

<sup>27</sup> Vgl. Tomasi di Lampedusa, *Der Gattopardo*, 41.

<sup>28</sup> Jacques Derrida, „Babylonische Türme: Wege, Umwege, Abwege“, *Übersetzung und Dekonstruktion*, hrsg. von Alfred Hirsch (Frankfurt am Main: Suhrkamp, 1997), 119–65.

zehnten „viele Autoren ihr Zuhause, die vor dem Schatten ihrer Kindheit flohen“. <sup>29</sup> Das Europa des *eastern turn* ist ein Ort der Vielstimmigkeit.

Es war Homi Bhabha, der davon ausging, dass jede Bezeichnung eine Übersetzung bedeutet – selbst innerhalb einer Kultur. Bhabhas *dritter Raum* oder *third space* entsteht beim Zusammentreffen von Kulturen und meint eine Zone, in der Kulturen sich verändern und wandeln. Durch die gegenseitige Beeinflussung werden bestimmte kulturelle Aspekte angenommen, andere wiederum verändert, manche auch ignoriert... Das ursprüngliche Machtgefälle wird in einem ‚Dazwischen‘, in einem *in-between* verändert und aufgelöst. Die Ambivalenzen der verschiedenen Weltansichten werden somit manifestiert, denn in dem Raum *in-between* bedeutet die Stimme der schwächeren Kultur nicht weniger viel als die der stärkeren. Hybridität bei Bhabha meint jedoch nicht, dass aus zwei Punkten und Kulturen eine dritte entsteht, sondern sie bezeichnet vielmehr einen dritten Raum, einen *third space*. Dieser ermöglicht es auch anderen Positionen zur Geltung zu kommen. In einem Interview definiert er den Begriff wie folgt:

Ich denke, selbst in dieser Position der Underdogs gibt es Möglichkeiten, die auferlegten kulturellen Autoritäten umzudrehen, einiges davon anzunehmen, anderes abzulehnen. Hybridisierung heißt für mich nicht einfach Vermischen, sondern strategische und selektive Aneignung von Bedeutungen, Raum schaffen für Handelnde, deren Freiheit und Gleichheit gefährdet sind. Dadurch werden die Symbole der Autorität hybridisiert und etwas eigenes daraus gemacht <sup>30</sup>

Somit stellt *third space* bei Homi Bhabha einen Aushandlungsort dar:

Bei ‚Third Space‘ geht es um Transformationen in gesellschaftliche Verhandlungen. D. h. Menschen kommen mit unterschiedlichen Einstellungen zusammen und streiten miteinander um Bedeutungen. Dabei entstehen neue Freiräume. <sup>31</sup>

Diese neuen Freiräume werden in *Olivas Garten* literarisch erprobt und weitergeführt. Denn dass der Garten Olivas immer schon vielschichtig und mehrsprachig war, zeigt auch Bremers Roman, der auf Deutsch die Ge-

<sup>29</sup> Jan Wagner und Nikola Madžirov, „Die Himmelsrichtungen machen den Menschen Angst“, *Die Horen: Zeitschrift für Literatur, Kunst und Kritik* 62/265 (2017): 4–30, hier 23.

<sup>30</sup> Lukas Wieselberg und Homi Bhabha, „Migration führt zu ‚hybrider‘ Gesellschaft“, *Science ORFAt* (09.11.2017), <http://sciencev1.orf.at/news/149988.html> (06.02.2018); vgl. auch Isolde Charim und Homi Bhabha, „Die Leute wollen teilnehmen“, *taz* (19.11.2007), [www.taz.de/!5191403/](http://www.taz.de/!5191403/), 06.02.2018.

<sup>31</sup> Lukas Wieselberg und Homi Bhabha, „Migration führt zu“.

schichte der italienischen Besetzung des ohnehin multikulturellen Dalmatiens während des Zweiten Weltkrieges thematisiert.

### Schlussbetrachtung

Auch wenn die Tendenzen zum Autobiographischen unverkennbar sind, die Autorinnen und Autoren, die das Thema Heimat in einer Fremdsprache verhandeln, erweitern nicht nur ihr Lesepublikum, sie werfen gerade durch das Autobiographische häufig auch einen ironisch-kritischen Blick sowohl auf ihr Herkunfts-, als auch auf ihr neues Heimatland. Sie zeigen somit Chancen einer postnationalen Literatur, indem sie „um Bedeutungen streiten“ (wie Bhabha es sagen würde).

Die Globalisierung des Sprechens, die vor allem in den neuen Medien spürbar ist, hat längst auch die Literatur und den Literaturbetrieb erfasst. Die Autorschaft ist kein Herkunfts- und Nationalitätsbeweis mehr – das zeigen nicht nur die von mir vorgestellten Autorinnen, sondern eine ganze Schar an Literaten, die derzeit den Literaturbetrieb dominiert.

Eine Frage an eine digitale Literatur- und Sprachwissenschaft, an die Stilometrie wäre, ob es Mechanismen gibt, mit denen die Textstrukturen dieser neuen Generationen der ‚Gastarbeiterliteratur‘ gemessen werden könnten. Mit welchem Sprachspiel wäre ihres vergleichbar, Thomas Manns (1875–1955), Tomasi di Lampedusas? Auf der anderen Seite nutzen Autorinnen wie Jadrečić und Bremer Möglichkeiten digitaler Medien, um ihre Texte zu publizieren und bekannter zu machen, ob literarisch oder essayistisch. Auch hier könnte gefragt werden, ob mit den Möglichkeiten der Stilometrie Unterschiede oder Gemeinsamkeiten zu den traditionellen Medien ausgemacht werden können.





# Intención autoral y mercado

## Una propuesta

Shelley Godsland (Amsterdam)

**RESUMEN:** Este artículo desarrolla un replanteamiento de las ideas sobre la intención autoral articuladas en el siglo xx por académicos tales como Stoll, Hirsch y, en particular, Wimsatt y Beardsley. Lo innovador de mi proposición es la formulación del concepto de la intención autoral derivada de factores extra-textuales, particularmente del mercado. A través de un estudio de caso basado en una novela española (*El tiempo entre costuras* de María Dueñas, 2009), demuestro cómo y por qué la intención autoral la pueden impulsar el deseo de ver un escrito adaptado a un formato audiovisual, o consideraciones económicas, y cuáles son las consecuencias de esta nueva conceptualización de la noción de la intención autoral.

**ZUSAMMENFASSUNG:** Der Beitrag schlägt ein Überdenken der Ideen von Stoll, Hirsch und v. a. Wimsatt und Beardsley hinsichtlich der Autorintention vor, wie sie im 20. Jahrhundert verstanden wurde. Das Innovative dieses Ansatzes ist die Betrachtung der Autorintention als abhängig von Faktoren, die außerhalb des Textes liegen: vor allem des Wirtschaftsmarkts. Dies wird exemplarisch an einem spanischen Roman (*El tiempo entre costuras* von María Dueñas, 2009) entwickelt. Es kann gezeigt werden, warum die Autorintention den Wunsch wecken kann, zusätzlich ein Drehbuch daraus zu machen oder wirtschaftliche Aspekte einzubeziehen. Was sind dann die Folgen einer solchen neuen Konzeption von Autorintention?

**PALABRAS CLAVE:** intención autoral; adaptación literaria; mercado literario; Dueñas, María; *El tiempo entre costuras*.

**SCHLAGWÖRTER:** Autorintention; Literaturverfilmung; Buchmarkt; Dueñas, María; *El tiempo entre costuras*

El debate sobre la intención autoral marcó importantes campos de la crítica y la filosofía en el siglo xx, y la discusión se caracterizó por un fuerte enfrentamiento conceptual e ideológico entre pensadores y académicos. Por un lado estudiosos como E. D. Hirsch en su *Validity in Interpretation* [La validez en la interpretación] (1967) proponían “the sensible belief that a text means what its author meant”,<sup>1</sup> mientras por otro, numerosos estudiosos mantenían que el escritor no puede ejercer intencionalidad autoral sobre su creación porque,

<sup>1</sup> Eric Donald Hirsch, *Validity in Interpretation* (New Haven y Londres: Yale UP, 1967), 1.

una vez se publique, pierde sobre su texto todo control porque son el crítico (a veces) o el público (en otras ocasiones) quienes otorgan significado e interpretación a su texto. Principales entre los exponentes de este último punto de vista son los reconocidísimos Wimsatt y Beardsley, cuya obra referenciaré más adelante. La polémica giró siempre en torno a la intencionalidad del autor en cuanto al *significado* del texto (en los múltiples sentidos de la noción de ‘significado’) – un enfoque muy acorde con la época en la que se discutía el tema (años 30 a 70 del siglo xx). Sin embargo, un elemento de intencionalidad autorral que los eruditos no investigaron fue lo que llamo yo la intencionalidad extra-textual, término que formule acá para referir a aquellas intenciones de quien escribe pero que van mucho más allá del simple ‘significado’ de un texto. Entre tales consideraciones enumeraría la *intención* de vender cierta cantidad de ejemplares de una novela, por ejemplo; la *intencionalidad* de que la obra logre ediciones en varios formatos; la *intención* de que apele a un público internacional y que se traduzca; la *intención* de que se convierta o se adapte en otro formato, por ejemplo, film, cómic o programa de televisión y que se difunda extensamente. Según mi propuesta, entonces, la intencionalidad del autor va más allá de lo que pueda significar el texto en sí, para ahora incluir una contemplación por parte de quien produce el escrito de numerosos elementos relacionados con lo comercial, con lo económico y con la dinámica del mercado en el objeto físico que es el libro (y, en el consumista siglo XXI, el *ebook*, el *audiobook*, el *merchandising* y las adaptaciones a géneros varios), situación bien entendible, ya que podemos suponer que quien en el nuevo milenio produce un texto no ignora las posibilidades para su comercialización y necesite venderlo. Por ende, postulo que nos encontramos frente a dos formas de acercamiento autorral a lo que es *intención* (aunque huelga decir que un solo escritor puede incurrir en los dos tipos de intención – la tradicional y la que sugiero yo –, procesos que son producto y reflejo de diferentes momentos en el mercado literario).

Lo que quiero hacer en este estudio, por lo tanto, es ofrecer una primera exploración y una puesta en práctica inicial de mi propuesta sobre esta intención extra-textual del autor. Propongo como objeto de mi breve estudio de caso una muy reciente novela española que, más allá de sugerir una valiosísima interrogación de la intención de la autora en cuanto a los significados que le quiso dar al contenido de su opera prima, sirve también para elaborar mi discusión sobre la intención extra-textual o comercial de quien escribe: *El tiempo entre costuras* (2009) de la académica y novelista española María Due-

ñas.<sup>2</sup> De fácil y amena lectura, *El tiempo entre costuras* supone una especie de cruce entre la novela de espías, la histórica, la rosa y la de aventuras. Cuenta las peripecias de una humilde modista madrileña, hija ilegítima de una madre también costurera, que justo antes de la Guerra Civil se escapa al norte de África con un amante. Embarazada y abandonada, nuestra protagonista – Sira – recurre al diseño y confección de ropas para ganarse la vida. Eventualmente, una vez terminada la contienda, vuelve a la capital para ejercer de espía para los británicos, ya que su profesión la permite contacto con las esposas de los altos mandos Nazis, de quien recolecta información y datos. Desde su publicación se han vendido en exceso de 20 millones de ejemplares internacionalmente; la obra ha sido traducida a más de una veintena de idiomas; y se le ha adaptado para una exitosísima serie televisiva cuyos derechos se han comercializado por el mundo. Es en esta adaptación para la pequeña pantalla y su conjunción con la novela en sí lo que será el enfoque de este estudio. Sin embargo, no es en la dinámica, proceso o puesta en práctica de la adaptación que me quiero enfocar, sino que, partiendo del concepto de la intención autoral y, en particular, de la premisa que esbozo arriba, quiero indagar en la idea de que esta novela de Dueñas nos sirve como ejemplo que sustenta mi tesis acerca de esa otra manifestación de la intención del autor – la extra-textual o comercial.

Son numerosos los críticos y reseñadores los que aseveran que Dueñas escribió *El tiempo entre costuras* expresamente con vistas a que se convirtiera en algún producto audiovisual: o bien un film o bien un programa de televisión (es decir, que escribió un texto que ideó y percibió como un guion pero que se publicó primero como novela en papel). No propongo ofrecer un detallado análisis de su ficción para comprobar que éste sea el caso, sino que parto del postulado (a sostenerse científicamente a base de un detallado análisis tanto de la novela como de la serie televisiva en otro apartado) de que los siguientes elementos de *El tiempo entre costuras* comprueban sin lugar a dudas que tal fue el propósito de la novelista al idear y producir su *bestseller*:

1. Conversaciones cuya estructura y teatralidad se asemejan a los diálogos contenidos en el guión para algún producto audiovisual (film, programa de televisión, etc.);
2. Detalladas indicaciones sobre los movimientos y desplazamientos físicos de los personajes que sugieren acotaciones teatrales/cinematográficas;

---

<sup>2</sup> María Dueñas, *El tiempo entre costuras* (Barcelona: Planeta, 2009).

3. Lujosas descripciones de atuendos, complementos, peinados y maquillaje que bien pueden interpretarse como instrucciones a vestuario en una compañía de producción filmica o televisiva;
4. Precisos retratos de interiores, fondos y paisajes que también sugieren la delineación de unos telones de fondo para una película o serie en la pequeña pantalla;
5. Un llamativo personaje central (Sira/Arish, la costurera-espía) a cuyo alrededor surge, desaparece y se mueve un elenco de personajes que correspondería a aquello de un film, incluyendo cientos de extras.

Se reconoce, claro está, que estos son ‘ingredientes’ que bien pueden figurar en cualquier obra de ficción escrita. Sin embargo, en el caso de *El tiempo entre costuras* podemos identificar dos tendencias principales que confirman la hipótesis de que, al redactar su ficción, Dueñas tenía la clara intención (extra-textual o comercial) de que posteriormente la obra se transformara genéricamente: por un lado es la *conjunción* de todos aquellos componentes que enumero arriba; y por otro la ‘fidelidad’ de la serie televisiva con la narrativa original.

Cabe mencionar que Dueñas no innova al escribir un texto publicable en papel para que luego se haga film o teleserie, puesto que van siendo ya varios autores (sobre todo en EEUU) cuya *intención* al escribir ha sido producir una novela con el fin de que se convierta rápida y fácilmente en un producto filmico. En su estudio titulado *Bestsellers*, John Sutherland nombra a King, Crichton y Grisham en este contexto, y nos dice:

These three chart-topping novelists all multiplied their sales from the synergies of film-and-book, book-and-film reciprocity. So much so that it was suspected that there was what the trade called ‘wag the dog’ at work – novels, that is, were conceived first as scenarios and screenplays.<sup>3</sup>

(También interesa recordar que ya en el siglo XIX en España eran numerosos los conocidos escritores que producían novelas para ser convertidas en obra dramática, o viceversa.) En este sentido, a María Dueñas no la incluiría *en términos generales* en esa categoría que Foucault, en su ya célebre ensayo, “¿Qué es un autor?”, identificaría como “fundadores de discursividad”.<sup>4</sup> Sin embargo, planteo que una nueva forma de “discursividad” para finales del siglo XX y el nuevo milenio la componen la conceptualización y puesta en práctica de

<sup>3</sup> John Sutherland, *Bestsellers* (Oxford: OUP, 2007), 76.

<sup>4</sup> Michel Foucault, “¿Qué es un autor?” [1969], traducido por Corina Yturbe, *Littoral* 9 (1983): 51–82, aquí 67.

una ideología textual-comercial según la que se concibe la ficción visual a la vez que – o inclusive antes de que – se idee la narración escrita (con el objetivo de que ésta última se transmuta en la primera). Sin embargo, aunque en esos *términos generales* que menciono a Dueñas le preceden los norteamericanos aludidos por Sutherland, como *española* y como *mujer* a la autora la podemos postular como “fundadora de discursividad” porque es de las primeras – si no la primera – en escribir expresamente para que lo textual se convierta en lo visual, evidenciando, de esta forma, esa *intención* extra-textual o de mercado, tal como analizaré a continuación.

Con referencia a lo que identifica como “fundadores de discursividad,” Foucault prosigue: “Lo particular de estos autores es que no son solamente los autores de sus obras, de sus libros. Produjeron algo más: la posibilidad y la regla de formación de otros textos”.<sup>5</sup> El filósofo explica que puede existir un caso como el de Ann Radcliffe quien, además de ser la autora de varias novelas, también asentó las bases para lo que sería el gótico literario que caracterizó los siglos XVIII y XIX. Sin embargo, según los postulados del pensador francés, ella no sería “fundadora de discursividad,” situación que se debe, primeramente, a que estos “fundadores de discursividad” “son muy distintos, por ejemplo, de un autor de novelas, que en el fondo no es nunca, sino el autor de su texto”.<sup>6</sup> Además, Foucault nos asegura que, dentro de su esquema, la apelación “fundadores de discursividad” debe atribuirse a aquellos individuos que ‘fundan’ movimientos, filosofías o sistemas de pensamiento que yo extrapolaría como los “grandes relatos” lyotardianos. De hecho, al respecto Foucault menciona a Marx y Freud. No creo que sea la primera en fijarme en la diferenciación de sexo entre el primer ejemplo foucaultiano en este respecto (Radcliffe) y los otros dos (Marx y Freud), y postulo que nuestro pensador bien pudo elegir una figura masculina como ejemplo de un “autor de novelas” cuyo quehacer creativo funcionara para fundar una corriente o tendencia literaria, así como a una mujer – feminista o proto-feminista, por ejemplo – que le sirviera de muestra como “fundadora de discursividad” (menciona a Melanie Klein, pero solo para luego demoler su sugerencia, comparando la psicoanalista negativamente con Freud). Sin embargo, si de intención autoral hablásemos, la aseveración foucaultiana según la que “un autor de novelas [...] en el fondo nunca es más que el autor

<sup>5</sup> Foucault, “¿Qué es un autor?”, 67.

<sup>6</sup> Foucault, “¿Qué es un autor?”, 67.

de su propio texto”<sup>7</sup> no es sostenible – particularmente cuando de mi propio postulado hablamos. En primer lugar, a nivel material, a un nivel simple, en el caso de María Dueñas y su *El tiempo entre costuras* vemos claramente que Dueñas sí es “más que [la] autor[a] de su propio texto,” ya que, según mi planteamiento, la española concibió un producto audiovisual a la vez que ideó la novela escrita, con la clara *intención* de que este último se convirtiera en el primero (manteniendo y promoviendo *ambos* productos, claro está). Mucho más allá de esta consideración del autor como creador del dual o doble artefacto literario-visual plástico, sin embargo, está la cuestión de los “fundadores de discursividad” foucaultianos. De nuevo debo expresar mi desacuerdo con el filósofo, ya que, sobre todo en el caso que acá me concierne – el de los “autores de novelas” que escriben con el objetivo específico de que su obra se convierta en obra audiovisual, para así comercializar mejor sus esfuerzos – está claro que formulan una especie de nuevo “récit” que, además, es uno que se fundamenta en el “grand récit” del capitalismo-consumismo masivo, ya que producen una obra literaria destinada a ser comprada y “engullida” como tal de forma compulsiva y, además, como artefacto fílmico. Por lo tanto, a aquellos autores que menciona Sutherland sí los podemos considerar como “fundadores de discursividad” – aunque en este caso la “discursividad” se deslice de lo meramente ideológico hacia lo comercial o lo extra-textual, como lo denominó yo. Y, claro está, es de esta tendencia que participa Dueñas. El producir el texto escrito de *El tiempo entre costuras* con el objetivo de convertir la obra en film o programa de televisión la sitúa a la española como *fundadora* de discursividad – sobre todo, como ya se resumió, en términos nacionales (es española) y de sexo (es mujer que escribe una novedosísima ficción sobre otras mujeres) – y el estatus de un escritor como “fundador de discursividad” se vincula estrechamente con nociones de intención del autor.

Inevitablemente, observaciones de este tipo sirven para sugerir un acercamiento analítico a *El tiempo entre costuras* y, posiblemente, otras novelas españolas, a partir del concepto de la *intención* de un autor – o autora – en el momento de creación textual. Por *intención* entendamos simplemente “*what he intended*” (“lo que [él] pretendía”) – donde el “he”/“él” es quien produce un texto, según William Wimsatt y Monroe Beardsley en su artículo fechado 1946

---

<sup>7</sup> Foucault, “¿Qué es un autor?”, 67.

sobre la *intencionalidad* de quien escribe,<sup>8</sup> que sigue siendo un estudio de referencia sobre el tema de la intención del autor. Huelga decir que estos dos académicos norteamericanos jamás entraron en ninguna discusión del factor mercado como motor de la intención del autor, por lo que aquí tomo prestados varios de sus conceptos como base de mis ideas, para luego reformularlos para que sean aplicables a casos de ‘literatura de mercado,’ o *best seller*, o como queramos llamar a lo que algunos observadores denominan ‘literatura light’. Y esa ‘literatura light,’ esa literatura que se vende como pan caliente, esa literatura que a veces convierte al autor de la misma en un personaje público, en una auténtica estrella mediática – además de generarle cuantiosos ingresos –, ¿una característica de esa literatura es el hecho de que la intención de su autor era venderla como producto de consumo y adjudicarle una futura función audiovisual más allá del libro impreso? La teoría existente sobre la intención del autor suele haberse publicado muchas décadas ha y, en general, se aplica a poesía canónica, no al texto narrativo y mucho menos a cualquier producto literario del *mass market*. Yo formulo un nuevo enfoque analítico partiendo de las premisas existentes en el campo y, aunque pretendo enfocarme específicamente en el ejemplo de *El tiempo entre costuras*, mi remodelada teoría de la intención autoral es aplicable a otros autores en otros contextos lingüísticos y nacionales. En el caso que acá nos concierne, tanto los autores norteamericanos mencionados por Sutherland como Dueñas *intencionalmente* producen una obra escrita con vistas a que se haga film. El por qué lo especulamos: intereses económicos y comerciales; necesidad de transmisión de algún mensaje socio-político; búsqueda de la fama ... (todo lo cual encajaría con mi postulado de que lo que llamo la intención extra-textual del autor en cuanto a consideraciones comerciales emerge a finales de un siglo xx y en un nuevo milenio caracterizados por el consumo frenético, contexto en el que lo meramente escrito, muchas veces, deja de ser un fin en sí mismo).

En un artículo aparecido en 1932, el académico norteamericano Elmer Edgar Stoll plantea la pregunta: “is not a critic ... a judge, who does not explore his own consciousness, but determines the author’s meaning or intention, as if the poem were a will, a contract, or the constitution? The poem is not the critic’s own”.<sup>9</sup> Echo mano en primera instancia a esta temprana cita dentro del campo anglosajón de la contemplación de la intención del autor ya que in-

<sup>8</sup> William Wimsatt y Monroe Beardsley, “The Intentional Fallacy”, *The Sewanee Review* 54, número 3 (1946): 468–88, aquí 468.

<sup>9</sup> Elmer Edgar Stoll, “The Tempest”, *PMLA* 47, número 3 (1932): 699–726, aquí 703.

cluye varios términos clave que deseo destacar y valorar como fundamento tanto para ilustrar y comprender el trayecto de los estudios dentro del campo hasta la fecha, así como para ir considerando y postulando su reformulación y actualización para su posterior aplicación al caso de *El tiempo entre costuras*. Stoll propone que el crítico literario (es decir, el académico como él mismo) cumple una función como “juez,” denominación que, ya desde un comienzo, sugiere que quien opera críticamente sobre un texto tiene en su haber las herramientas necesarias como para disertar y alcanzar conclusiones sobre la ‘corrección’ o su estatus como ‘bueno’ o ‘malo’/‘correcto’ o ‘incorrecto’ del escrito que escudriña. Estos últimos términos – binarios, opuestos – a su vez evocan consideraciones de ‘calidad’, idea que, de igual manera, implica la capacidad de ese “juez” para formular pronunciamientos sobre la excelencia, importancia, relevancia, etc. (o no) del producto literario que estudia. Esta extrapolación lo posiciona al crítico-“juez” casi como una deidad poseedor de unos conocimientos y unas capacidades que lo (porque en la época en la que escribía Stoll los críticos eran mayormente hombres) cualifica para juzgar una obra literaria en varios de sus aspectos. Según Stoll, el sabio crítico-“juez” también es capaz de desentrañar el “sentido” (“meaning”) de un escrito (él menciona específicamente un poema) y la “intención” (“intention”) de su autor. Implícita a mi propuesta sobre la relación triangular que hoy existe entre autor/intención/mercado es una reformulación de muchas de las suposiciones de Stoll. Evidentemente el crítico-“juez” hoy en día bien puede ser una crítica-“jueza,” situación que puede llevar a nuevas prácticas interpretativas. El escrito sobre el que opera puede pertenecer a cualquier género (ya rara vez es el poema que menciona Stoll, pero puede ser literatura de consumo que antes el crítico no veía que mereciera su atención intelectual). Como quien escribe ahora también la mitad de las veces es una mujer, el “significado” (“meaning”) de lo que produce puede ir por derroteros antes desconocidos e insospechados (en épocas en las que la mujer de letras era una excepción a una regla masculina). Y ese “significado” tal vez lo vea diferente esa crítica-“jueza” a como lo vería el crítico-“juez” de Stoll. Igualmente, esa crítica-“jueza” tal vez dedique atención no solo a la intención de quien escribe cuando de “significado” hablamos, sino también capte la intención de una autora del siglo XXI en cuanto a la proyección de su escrito más allá del mero producto impreso. De hecho, esta doble intencionalidad autoral – la tradicional y la que yo propongo – parece ayudar a sostener una de las aseveraciones de Stoll: que el poema no es del crítico (“The poem is not the critic’s own”). De



hecho, no lo es; la intencionalidad del autor o de la autora demuestra que la obra sigue siendo suya, aún cuando lo haya ‘soltado’ y el crítico se haya apoderado material e intelectualmente de ella (postulado que contradicen Wimsatt y Beardsley, tal como veremos más adelante).

En su trabajo ya citado, Wimsatt y Beardsley responden a Stoll para demostrar su desacuerdo con alguna de sus postulaciones, y aseveran:

The poem is not the critic's own and not the author's (it is detached from the author at birth and goes about the world beyond his power to intend about it or control it). The poem belongs to the public. It is embodied in language, the peculiar possession of the public, and it is about the human being, an object of public knowledge.<sup>10</sup>

Tal vez no nos debería sorprender que Wimsatt y Beardsley señalen que la obra a ser examinada por el crítico (el “juez” de Stoll) sea un poema (al igual que Stoll), ya que en el contexto histórico en el que trabajaban la atención crítica se enfocaba en ‘el canon’. Foucault nos recuerda que la atribución de la pertenencia de un texto se deviene sistema a comienzos del s. XIX. Si el texto tiene un autor de carne y hueso, reconocido, reconocible – premisa fundamental de mi tesis – entonces sigue que la aseveración de Wimsatt y Beardsley en cuanto a que el poema pertenece al público (“The poem belongs to the public”) no puede menos que ser errónea. Tampoco aciertan Wimsatt y Beardsley cuando alegan que el producto literario (ellos mencionan el poema, claro está, pero entendamos por ello cualquier obra, en este caso el objeto de nuestro interés crítico-académico) se desprende del autor en el momento en el que nace y “va por el mundo” más allá del poder de su creador para seguir ejerciendo sobre ella su *intencionalidad* como quien la diera vida, ni de controlarla (y por el uso del adjetivo posesivo “his” en el inglés original nos damos cuenta de que el hipotético autor aquí aludido es irresolutamente masculino, tal como entendimos del trabajo de Stoll). Mantengo que no aciertan porque, si bien un crítico o el público pueden atribuirle a una obra un “significado” que no fue la “intención” del quien la concibiera y la desarrollara, para, de este modo, ‘arrancarle’ del autor cierto control sobre lo que escribió, un caso como el de *El tiempo entre costuras* pone patas arriba el postulado de los dos norteamericanos (siempre dentro de mi renovada visión de lo que implica “intención” autoral, huelga decir). Esto es claramente el caso porque de ninguna manera ese producto de María Dueñas “va por el mundo,” como digo yo, “más allá del poder de su creador[a] para seguir ejerciendo sobre ella su

<sup>10</sup> Wimsatt y Beardsley, “Intentional Fallacy”, 470.

*intencionalidad* ... ni de controlarla” porque, en términos materiales, Dueñas sigue teniendo dominio y autoridad sobre su obra, principalmente porque desde el inicio también la conceptualizó como producto audiovisual.

Algunas de las proposiciones de estos estudiosos especializados en la teorización de la intención del autor merecen más interrogación y un replanteamiento acorde a mi tesis central sobre el *contexto* en el que se manifiesta la intención extra-textual o de mercado del autor, sobre todo en lo referente a dos puntos que tienen que ver con género literario y el género/sexo de quien escribe. No cabe duda de que a *El tiempo entre costuras* la podemos clasificar como una literatura ‘de consumo’, ya que, a pesar de que inicialmente la editorial no la lanzara a bombo y platillo, llegó a ocupar un espacio altamente visible y muy rentable dentro del comercio del libro en España – y otros mercados afines, como el televisivo, por supuesto. En cuanto a esta literatura no necesariamente ‘canónica’ del siglo veintiuno, debemos plantear una serie de preguntas, cuya respuesta nos esclarece la intención autoral que puede yacer detrás (principalmente cuando de mi noción de intención extra-textual o de consumo hablamos). Las interrogaciones serían las siguientes. ¿Hasta qué punto es legítimo asegurar que es justamente en el caso de la literatura de consumo en donde *más* control tiene el autor sobre su escrito y en donde *mayor* intención puede ejercer? ¿El autor de literatura de mercado – en conjunción con su agente – decide sobre ediciones, traducciones y ventas? ¿Ejerce la intención de que lo que redacta se haga otro producto – audiovisual, musical, cómic, etc.? Según podemos entender, la respuesta a estas dudas es afirmativa, lo cual, por un lado sirve para revelar como falacia los postulados de Stoll y Wimsatt y Beardsley en cuanto a la intención del autor y su posterior *falta de control* sobre un escrito.

El segundo punto que hace falta que escudriñemos es la contundente aseveración por parte de los tres académicos citados según la que una obra de literatura no le pertenece al autor (ya que en la visión de Wimsatt y Beardsley va solita por el mundo una vez el autor la haya soltado y que sobre ella pueden operar tanto el crítico como el público), ni tampoco al *crítico*. A lo largo de este ensayo ya me aludí varias veces a cuestiones referentes al género o sexo de los académicos cuya obra cito acá; de los ejemplos de “fundadores de discursividad” que menciona Freud; y de la obra literaria que se escribe (evidentemente con referencia principal a Dueñas). Para concluir este estudio mío, por lo tanto, se propone un análisis de los conceptos centrales a la discusión académica en torno a intención autoral y control sobre el texto, desde la óptica de géne-

ro/sexo. En cuanto a las aseveraciones de Stoll y Wimsatt y Beardsley según las que un “poema” tampoco es del crítico, desde mi perspectiva como crítica, francamente no tengo objeción ninguna en no hacerme ‘propietaria’ de una obra sobre la que ejerza una labor analítica; no le encuentro ningún valor ni importancia en que la obra se haga ‘mía’, por lo que habría que cuestionar si la fascinación que parecen evidenciar los profesores ya citados con la imposibilidad – como críticos – de ‘adueñarse’ de un producto literario, tiene un fuerte fundamento y una firme raíz en identidades de género. El interés demostrado por parte de los tres académicos – “jueces” – hasta ahora citados por la *pertenencia* y *propiedad* de una obra literaria (término que sugiere además esa noción de control que también mencionan Wimsatt y Beardsley) ¿lo podemos entender como elemento de la expresión de una identidad tradicionalmente asociada con el hombre, con lo masculino? Huelga decir que ellos – y otros – ejercían de analistas textuales y críticos en un momento histórico en el que las mujeres eran aún escasísimas contribuyentes al campo (a cualquier campo que no fuera el doméstico, de hecho), por lo que tal vez este enigma de la fascinación con el haber, tener y controlar un texto escrito sea producto de esta histórica configuración de la profesión en términos de género/sexo. Si bien estoy de acuerdo con la aseveración de Stoll, replicada textualmente por Wimsatt y Beardsley, en cuanto a que “[t]he poem is not the critic’s own,” debo señalar que no encuentro imperativo alguno para que esta particular realidad inspire preocupación en ningún crítico/crítica ni académico/académica, ya que a los del gremio no nos incumbe hacernos ‘dueños’ del texto que analizamos. Por otro lado, sin embargo, es necesario examinar y argüir el postulado de Wimsatt y Beardsley en donde plantean que el texto tampoco le pertenece al autor, sino que luego de ‘nacer’ va por el mundo más allá del poder del autor de influir más en su *intención* (entendamos por ello la *intención* tanto de quien escribe [es decir, mi idea de la intención extratextual, de mercado o con fines de adaptación del escrito de un género cultural a otro] como de lo que escribe [es decir, las ideas que aparecen en el texto mismo]). La *intención* de María Dueñas al componer *El tiempo entre costuras* fue, como bien se entiende, producir una novela que luego se adaptaría para el cine o para la televisión. En este caso específico, por lo tanto, el “poema” – que ya establecimos que puede ser cualquier producto literario (aunque preferiblemente uno que encaje con los patrones más tradicionales, por lo que estaríamos hablando principalmente de una novela, un cuento corto o una obra dramática) – quedaría bajo el ‘control’ de quien le ‘dé vida’. Además –

y lo que nos es de particular interés con el ejemplo de Dueñas – la *intención* del escritor (en este caso, una escritora) se mantiene intacta, ya que la obra inicial cumple con la *intención* con la que es investida por su creadora (o las *diferentes* intenciones hasta aquí mencionadas – la tradicional y la mía). Por lo tanto, para reformular el postulado de Wimsatt y Beardsley y reconfigurarlo en lo referente a género/sexo, “[t]he poem ... [*does not go*] about the world beyond [*her*] power to intend about it” – conclusión que cuestiona y reformula las tradicionales aserciones críticas que plantean que el autor no entiende su *intención* al conceptualizar, escribir y publicar una obra de ‘literatura’, ya que tanto Dueñas como otros escritores acá referenciados *bien* saben cuál es su intención al escribir – en términos comerciales (acá no asevero nada sobre un posible ‘mensaje’ a ser transmitido por el texto; esto queda para trabajos futuros). Lo que es más, esta aserción del control y conocimiento por parte del escritor de su *intención* extra-textual al crear una obra puede tener importantes consecuencias para la mujer que escribe, ya que le devuelve el derecho a ejercer esa *intención* y, por ende, le otorga una cierta agencia sobre su quehacer literario y las consecuencias de éste, sean éstas una adaptación fílmica o al género que sea.

# *Anhang*



## Tabellen und Code

“Dice Bernardo a Cristo que un Portero al Papa” . . . . .	35
1 Comparación de versiones con base en el ensayo de Fucilla . . . . .	46
Estilometría y la Edad Media castellana . . . . .	49
1 Formas y ocurrencia de la conjunción copulativa en el corpus alfonsí . . . . .	56
2 Número de palabras-token y palabras-tipo de cada uno de los textos tenidos en cuenta . . . . .	57
3 Estadísticas básicas del <i>Libro de la caza de las aves</i> . . . . .	66
In Search of a New Language . . . . .	117
1 xml-document, result of POS-Tagging via FreeLing (detail for the first word “la”). . . . .	124
2 xml-document, result of POS-Tagging via FreeLing (detail for the 27 <sup>th</sup> word “son”) and transcription of the tag “VSIP3PO VSI” . . . . .	126
3 xml-document, result of POS-Tagging via FreeLing (detail for the 44 <sup>th</sup> and 45 <sup>th</sup> words “buzo” and “ciego”) and transcription of the tags “NCMSoo NC”, “AQoMSoo AQ” . . . . .	127
4 Table of used corpus . . . . .	131
5 Examples of Close Reading for Góngora, Picasso, Ramón Gómez de la Serna . . . . .	141
Delta Inside Valle-Inclán . . . . .	151
1 Works of Valle arranged in two competing chronological groups . . . . .	155
2 Predicted period of disputed text . . . . .	156
3 Works of Valle arranged in several competing groups of works . . . . .	158
4 Predicted groups of disputed text . . . . .	159





## Abbildungsverzeichnis

Estilometría y la Edad Media castellana . . . . .	49
1 Obras de don Juan Manuel, según la edición normalizada de Alvar (100 MFW). . . . .	52
2 Obras de don Juan Manuel, según la edición normalizada de Alvar (1000 MFW). . . . .	53
3 Obras de don Juan Manuel, según la edición semipaleográfica de Ayerbe-Chaux (100 MFW) . . . . .	54
4 Obras de don Juan Manuel, según la edición semipaleográfica de Ayerbe-Chaux (1000 MFW). . . . .	54
5 Obras del Scriptorium alfonsí (100 MFW). . . . .	58
6 Obras del <i>Scriptorium alfonsí</i> (3000 MFW) . . . . .	59
7 El corpus alfonsí con las <i>Siete Partidas</i> de 1491 (SP-INC) (100 MFW) . . . . .	61
8 El corpus alfonsí con las <i>Siete Partidas</i> de 1491 (SP-INC) y eliminadas las conjunciones copulativas <i>e</i> y <i>et</i> (100 MFW) . . . . .	62
9 Los capítulos del <i>Libro de la caza de las aves</i> (100 MFW). . . . .	68
10 Los capítulos del <i>Libro de la caza de las aves</i> (500 MFW). . . . .	69
11 Los capítulos del <i>Libro de la caza de las aves</i> (1000 MFW) . . . . .	70
12 El <i>Libro de la caza de las aves</i> dividido en fragmentos de 1000 palabras . . . . .	70
13 El <i>Libro de la caza de las aves</i> dividido en fragmentos de 800 palabras . . . . .	71
14 Dendrograma del <i>Libro de la montería</i> dividido en fragmentos de 1000 palabras. . . . .	73
15 Detalle de la figura 14. . . . .	73
Fernando de Herrera y la autoría de <i>Versos</i> . . . . .	75
1 <i>Amor es más laberinto</i> . SVM. . . . .	85
2 P. SVM. Samples of 5000 words. 100 MFW . . . . .	86
3 P2. SVM. Samples of 5000 words. 100 MFW . . . . .	86
4 P. SVM. Samples of 3000 words. 100 MFW . . . . .	86
5 P2. SVM. Samples of 3000 words. 100 MFW . . . . .	87
6 P. NSC. Samples of 5000 words . . . . .	87
7 P2. NSC. Samples of 5000 words . . . . .	87
8 P. Delta. Samples of 5000 words . . . . .	88
9 P2. Samples of 5000 words . . . . .	88
10 SVM. Samples of 5000 words. 100 MFW. Aleatory reduced version P. . . . .	89
11 NSC. Samples of 5000 words. Aleatory reduced version P . . . . .	89
12 Delta. Samples of 5000 words. Aleatory reduced version P . . . . .	89
Góngora hors norme? Étude stylométrique d'un motif gongorin . . . . .	91

1	Profils de fréquences comparés Góngora/Garcilaso/étalon avec barres d'erreur . . . . .	97
2	Écarts relatifs de Góngora à la moyenne de l'étalon pour chaque SGR. . . . .	98
3	La séquence la plus surreprésentée chez Góngora : NC ADJ NC, employée 154 % de fois plus que dans l'étalon . . . . .	100
In Search of a New Language . . . . .		117
1	Screenshot of FreeLing-Results ( <a href="http://nlp.lsi.upc.edu/freeling/demo/demo.php">http://nlp.lsi.upc.edu/freeling/demo/demo.php</a> ) and a transcription of line 1-5 . . . . .	124
2	Dendrogram of the Corpus of Spanish Golden Age Sonnets (1000 MFW, cosine Delta). . . . .	130
3	Dendrogram for an enlarged corpus for 32 authors of Golden Age sonnets, same parameters, arrow added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets . . . . .	133
4	Dendrogram for sonnet-corpus split in at least two texts, same parameters; arrows added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets . . . . .	134
5	Dendrogram of the split sonnet-corpus (cf. fig. 4), only 300 MFW, arrow added to point out clustering of Góngora, Lope and Litala, colored box added to point out the arm of older sonnets. . . . .	136
6	Dendrogram of the split sonnet-corpus, without 16 <sup>th</sup> century-authors and Tassis y Peralta (less than 1000 words), arrows added to point out clustering of Góngora, Lope and Litala. . . . .	137
7	Dendrogram for the Picasso-corpus, generated with <i>stylo</i> . . . . .	140
8	Screenshot: Result of DARIAH Topics Explorer: corpus and document size. . . . .	144
9	Screenshot: Result of DARIAH Topics Explorer: Topics (here limited to 10) in the corpus of CLIGS-textbox, Hispanic novels of the 19 <sup>th</sup> century <sup>12</sup> . . . . .	145
10	Screenshot: Result of DARIAH Topics Explorer: Topics for Picasso-corpus as cited in fig. 8 – here cut by topic number 14 (of 30). Highlighted color words by N. R.-P. . . . .	146
11	Screenshot: Result of DARIAH Topics Explorer: Topics for Picasso-corpus (quadrupled), Highlighted boxes added by N. R.-P. . . . .	147
12	Screenshot: Result of DARIAH Topics Explorer: Topics for corpus of Spanish poetry 20 <sup>th</sup> century (without Picasso), words in the corpus doubled, 10 Topics, 100 iterations. Highlighted boxes added by N. R.-P. . . . .	148
13	Screenshot: Result of DARIAH Topics Explorer: Topics for corpus of Spanish poetry 20 <sup>th</sup> century (with Picasso), words in the corpus doubled, 10 Topics, 100 iterations. Highlighted boxes added by N. R.-P. . . . .	148
Delta Inside Valle-Inclán . . . . .		151
1	Dendrogram of Valle's novels . . . . .	153
2	Accuracy scores for hypothesis of periodization in Valle's novels . . . . .	157
3	Accuracy scores in cross validation (KNN) of different hypotheses . . . . .	161

<sup>12</sup> See *Collection of 19th Century Spanish-American Novels (1880–1916)*, edited by Ulrike Henny-Krahmer. Würzburg: CLiGS, 2017. <https://github.com/cligs/textbox/master/spanish/novela-hispanoamericana/>

---

4	Dendrogram of Valle's Novels with groups from the classification . . . . .	162
---	--	-----



## Verfasser- und Schlagwortindex

### A

Altspanisch, 49  
 Artaud, Antonin, 165  
 Autorfunktion, 35  
 Autorintention, 195  
 Autorschaft, 5, 49, 117, 181

### B

Bremer, Alida, 181  
 Buchmarkt, 195  
 Burchiello, 35

### C

Calvo Tello, José, 151  
 classification, 151

### D

Digital Humanities, 75  
 Dueñas, María, 195

### E

El tiempo entre costuras, 195  
 Erstić, Marijana, 181

### F

Fradejas Rueda, José Manuel, 49  
 Fucilla, Joseph, 35

### G

Geschichte der Stilistik, 21  
 Godsland, Shelley, 195  
 Góngora, Luis de, 117  
 Gremels, Andrea, 165

### H

Hernández Lorenzo, Laura, 75  
 Herrera, Fernando de, 75

### I

Idiolekt, 91

Interkulturalität, 181  
 Italienische Komik, 35

### J

Jadrejčić, Tamara, 181  
 Jannidis, Fotis, 35

### K

Klassifizierung, 151

### L

Lescasse, Marie-Églantine, 91  
 Lexomik, 49  
 Literaturverfilmung, 195

### M

Machine Learning, 151  
 Magliabecchiano VII, 35  
 Mittelalter, 49  
 Motiv, 91

### N

numerische Hermeneutik, 91

### P

Pérez Medrano, Alan Joaquín, 35  
 Picasso, Pablo, 117  
 Poetik, 165  
 Psychoanalyse, 165

### Q

Quattrocento, 35  
 Quellen, 49

### R

Rißler-Pipka, Nanette, 5, 117  
 Roman, 151

### S

Scheurer, Maren, 165

Schuchardt, Hugo, 21

Siglo de Oro, 75

Spitzer, Leo, 21

Stefanelli, Diego, 21

Stil, 91, 117, 165

Stilistik, 5

Stilometrie, 5, 49, 75, 91, 117, 151

## **T**

Topic Modeling, 117

## **U**

überwachtes Machine Learning, 151

## **V**

Valle-Inclán, 151

Vossler, Karl, 21