SKETCHNOTE BY ELVIA VASCONCELOS OF BEATRICE GRÜNDLER'S PRESENTATION WTIHIN THE WORKSHOP "WHAT IS THE DIGITAL DOING? A WORKSHOP IN THE INTERFACE", NOVEMBER 2020.

HUMAN INTUITION AND COMPUTATIONAL CLUSTERING:

TACKLING A FLUID TEXTUAL TRADITION

"In a nutshell, such research constitutes an attempt of 'catching a text on the move'."

Beatrice Gruendler and Mahmoud Kozae

Beatrice Gruendler and Mahmoud Kozae, Human Intuition and Computational Clustering: Tackling a Fluid Textual Tradition. *Interface Critique* 5 (2025): 142–172.

DOI: https://doi.org/10.11588/ic.2025.1.113536

Transmission

In Arabic literature, which continues over fifteen centuries until today, texts evolve at different speeds which may be placed at two ends of a spectrum. In the manuscript age, works of elite literature and scholarship were governed by conventions of transmission which alternated between oral and written modes, controlled by experts who would correct a text at each instance. As a result, such texts remained often remarkably stable (and lend themselves to be edited using the stemmatic principle of reconstructing one original). At the other end of the spectrum, in Arabic popular literature and middle literature (i.e., showing narrative sophistication), often of anonymous authorship, story tellers or redactors took liberties to change and adapt texts in retelling or rewriting.

The work discussed here, *Kalīla wa-Dimna*, began its life as elite literature, when it was translated in the mid-eighth century CE from Pahlavi into Arabic as a manual of statecraft. But its twofold ethical and entertaining content, presented in the form of fables, allowed it to become over time a popular work. Moreover, its stories, which pertain to human interaction, such as friendship, trust, betrayal, scheming, questions of justice, and the conduct of life, were prone to provoke responses from the agents of transmission, based on their own life experience and expectations (different from medieval European copyists, who copied Latin, Arabic copyists could understand the texts they worked with). Linguistically speaking, this is reflected in the shift from the high idiom of classical Arabic to the hybrid register of

Middle Arabic with relaxed grammar, pseudo-corrections, and dialectal elements.

Editing

This leads – in the over 160 preserved manuscripts identified by the research team – to almost as many instantiations of the work. To 'read' these is a challenge, since no printed edition can do justice to such mutability. Indeed, its few editors (four since 1816) have each picked a single manuscript, skewing the perception of the textual tradition as a whole. In contrast, the *digital realm* offers the possibility of juxtaposing multiple witnesses and compare their alterations down to the smallest detail and distinguish between the shared changes of given versions and individual articulations. However, as the reader of such a multitext edition is faced with many moving parts, one must find a new way of reading such variety. A quantitative analysis can aid in the grasping of aspects of these changes in a systematic manner.

Agency

The human agency of the (anonymous or little known) copyist-redactors of *Kalīla wa-Dimna*, is perceivable in the different degrees to which individual chapters, or fables, were rewritten – and which parts within them more than others. Two chapters (of the four analysed here) stand out by their particularly broad spectrum of difference, so much so, that one cannot pin down a common plot line. Rather, the plot 'splinters' into different micro-narratives.

The following contribution to *Interface Critique* is a reflection on two ways we have developed for analysing the massively diverging manuscript tradition of *Kalīla wa-Dimna*.[1] To briefly introduce *Kalīla wa-Dimna*: This book of advice in the form of fables featuring human and animal characters was first composed from Sanskrit parts in a (lost) Middle Persian version (c. 550 CE) and translated into Syriac (c. 590 CE) and Arabic (c. 750 CE). All later translations into over forty languages in Europe, North Africa and West, South, and Southeast Asia derive from the Arabic.[2] Nonetheless, no full text in Arabic existed for nearly half a millennium (750–1220 CE). The surviving Arabic manuscripts date from the 13th–19th century and include remediation from manuscript into print and back to manuscript. Moreover, during the 17th–19th centuries, literacy and readership grew incrementally in North Africa and West Asia, broadening the work's audience considerably, including, for instance, access via endowed lending libraries.[3] Rewritten by innumerable intervening copyist-redactors, *Kalīla wa-Dimna* diversified into versions too distant from each other to be approached as descendants

1. The *Kalīla and Dimna – AnyonymClassic* project has received funding from the European Research Council, under the European Union's H2020-EU.1.1. – EXCELLENT SCIENCE program, Advanced Grant no. 742635. The *Arabic Literature Cosmopolitan* project has been funded by the Gottfried Wilhelm Leibniz Prize awarded by the German Research Foundation (DFG) in 2017. See the joint projects' website: https://www.geschkult.fu-berlin.de/en/e/kalila-wa-dimna/. We thank the editors and Khouloud Khalfallah for their valuable comments on an earlier draft of this contribution.

2. The work's structure as a frame story with sub-stories which in turn contain sub-sub-stories will not be discussed in the present analysis. The narrative layers are not *per se* represented in the data, though there is the option to do so in the future.

3. Konrad Hirschler, *The Written Word in the Medieval Arabic Lands: A social and cultural history of reading practice* (Edinburgh 2012).

from an "original"[4]; they can only be observed in their received forms and their incremental or structural alterations from one manuscript to the other.

The collection of fables constituting *Kalīla wa-Dimna* is not a 'book' in the conventional sense but a textual tradition that underwent a six-centuries-long iterative process of rewriting, as documented by innumerable intervening copyist-redactors who reacted to both the text and their own socio-historical context. It is fair to say that *Kalīla wa-Dimna*, though usually recognizable, cannot be grasped as a stable unity. The Berlin-based joint project of AnonymClassic and Arabic Literature Cosmopolitan at Freie Universität Berlin is the first to identify and trace these motions and investigate their text-historical, literary, and transcultural dimensions.[5] For the researchers, the very mutability of the text moves centre stage and becomes itself focus of the investigation. A way needed to be found to simultaneously observe and analyse this fluid tradition's numerous textual witnesses, because in each instance of producing a new written copy, changes were introduced in ways large and small. In a nutshell, such research constitutes an attempt of 'catching a text on the move.'

---

4. The strong rewriting of *Kalīla wa-Dimna* calls for explanation; among the many factors, some are particularly compounding: (a) copyist-redactors who were aware of the textual fluctuation switched between *Vorlagen* [models] or combined several to a greater or lesser extent to create their own new collages; (b) in contrast to medieval copyists of Latin works, Arabic copyists were able to read the language of the texts. By consequence, as they copied the content, ethics and practical philosophy (instruction how to use a book vs. decoding a story; the status of scholarship; true vs. calculated friendship; royal injustice) was of relevance to them and solicited (more or less) intervention by individual redactor.

5. Beatrice Gruendler et al., An Interim Report on the Editorial and Analytical Work of the AnonymClassic Project, *Medieval Worlds: Comparative and Interdisciplinary Studies* 11 (2020), pp. 241–279; https://doi.org/10.1553/medievalworlds_no11_2020s241, with further literature on *Kalīla wa-Dimna* cited there; Matthew L. Keegan and Beatrice Gruendler, Before and after Ibn al-Muqaffaʿ: Introduction to the Special Issue on (Un)declared Co-Authorship in the *Kalīla wa-Dimna* Textual Tradition. *Journal of Abbasid Studies* 9 (2023), pp. 1–7; https://doi.org/10.1163/22142371-00802010, access: July 3, 2025, 4:12pm.

Here, two prefaces and two chapters of *Kalīla wa-Dimna* will be used as examples. Prefaces form an important part of the work, since there exist several, and they were added to the work in its successive translations.[6] Some have a narrative character, such as "The Voyage of Burzoy," which tells the story of the book's import from India and foregrounds the value of knowledge and the figures of the learned king and his traveling scholar. "The preface Ibn al-Muqaffa'" (d. 757 CE), the Arabic translator-adaptor, marks the endpoint of the book's transmission history (viewed from the perspective of Arabic readers). It imparts advice on how to read parables in-depth and decode them and formulates an ethics for scholars. The two chapters derive from the Indian *Mahābhārata* and concern issues of trust. The chapter of "The Rat and the Cat" tells of a case of temporary trust between two natural foes in order for both of them to survive. The chapter of "The Lion and the Jackal" treats the theme of broken trust between a king and his advisor, and the question of whether this can be repaired. Both chapters can be read on the interpersonal as well as on the political level. Particularly the latter chapter veers in its rewritings between emphasising ethical dimensions vs. political pragmatism. Together, the four examples give a sense of *Kalīla wa-Dimna's* thematic spectrum.

The present analysis concerns not directly the 'raw data' of the manuscripts, their handwritten witnesses, but rather a second level of their digital processing and the way one can 'read' the data on this level of abstraction, which offers new insights but also poses its own problems. In summation, we preliminarily conclude that both

6. See: Johannes Stephan, A Sense of No Ending: *Kalīla and Dimna* between Fluidity and Completeness. *How to End Things in Arabic Literature,* ed. Lale Behzadi und Bilal Orfali, Special Issue, *Al-Abhath*, 71 [1–2] (2023), pp. 53-81, https://brill.com/view/journals/alab/71/1-2/article-p1_2.xml, access: January 25 2024, 10:07am.

the approach of the intuitive eye and the mathematical reasoning of data analysis complement and correct each other and bring us a step closer to grasping the written *mouvance* of *Kalīla wa-Dimna*. The analysis is made possible by prior work: the 'raw data' of the manuscripts has been collected and analysed, the process of which shall be briefly summarized here. The text versions change profoundly, and thus a research process that begins with them runs the risk of disorienting the researcher completely. The consonantal skeleton of the Arabic script is ambiguous (short vowels are not written), and certain letters look alike in the different manuscript hands. The copyist-redactors' rereading leads to lexical changes, which in turn triggers the alteration of entire phrases to contextualize the new meanings of words. Furthermore, redactors actualise details of material life and social as well as legal customs. They interfere with the linguistic register, simplifying the classical Arabic grammar and introducing pseudo-corrections or colloquial expressions (those might be inadvertent or intentional, for dramatic effect). Finally, they themselves may be aware of differing versions and combine several models, or *Vorlagen*, into one. Our narratological analysis has therefore proceeded by plot structure, dividing select prefaces and chapters into segments that represent each step of the plot. Manuscripts were typed, segmented and transformed into an annotated data format, marking segments by unique identifying labels and allowing these segments to be aligned by an algorithm.[7] Not only does our resulting digital annotation interface permit us to show the

7. See Thomas Bremer, Paul Molitor, Marcus Pöckelmann, Jörg Ritter, and Suzanne Schütz, Zum Einsatz digitaler Methoden bei der Erstellung und Nutzung genetischer Editionen gedruckter Texte mit verschiedenen Fassungen – Das Fallbeispiel der *Histoire philosphique des deux Indes* von Guillaume Thomas Raynal, *Editio* 29[1] (2015), pp. 29–51; and Susanne Schütz and Marcus Pöckelmann, LERA – Explorative Analyse komplexer Textvarianten in Editionsphilologie und Diskursanalyse, *Book of Abstracts of the Third Annual Conference of Digital Humanities for German-Speaking Regions* (2016), pp. 239–243.

full text of each digitized manuscript within a multitext edition and align matching segments of the manuscripts across them – but it also displays a graphic overview of the presence, absence, or altered position of each segment across all digitized manuscripts in the form of small boxes. Such a graphic view is, so to speak, an abstract 'map' of the varying narrative as the manuscripts preserve it. This 'mapped narrative' is our focus here.

This essay shifts the reading from the manuscript texts to their abstracted structures, represented as maps. It becomes, in the sense of Franco Moretti, a simultaneous 'distant reading' of multiple texts. In fact, we employ two different ways of distant reading, relying on traditional philological 'reading by eye' and on automated algorithmic analysis, to see how both approaches inform each other. To return to the 'raw data,' over 160 manuscripts of *Kalīla wa-Dimna* survived, and of the c. 119 manuscript witnesses surveyed in the project, those representing the widest spectrum of divergence have been selected for the synoptic editions of selected prefaces and chapters.[8] This translates into 6 – 16 manuscript witnesses per multitext edition. The manuscript sample is not always the same; between the chapters, the interrelations of the manuscripts differ. Manuscripts that do not exhibit much rewriting in one chapter may do so in another and vice versa, so that each chapter must be considered first on its own. Those manuscripts that exhibit the broadest range of changes in a given chapter are selected for its respective multitext edition. In some chapters, manuscripts form recognisable groups.In others, individual versions dominate, and the selection must be tailored to show this. The actual texts, represented in columns for each manuscript, do not concern us here. Rather, our analysis draws from the abstract digital

8. Hereafter summarily referred to as chapters.

maps rendered from the segments into which they were subdivided. This allows a bird's eye view over the narrative's structural changes (see figures 2, 4, 6, and 8, discussed below).

First approach: Data showing a degree of diversity are arranged by eye according to apparent similarity

Over the last seven years (2018–2024), the precision of the editions' structural representation as digital maps has been gradually improved.[9] The represention of text segments and their display within our digital maps has been refined and made more precise in two subsequent new versions of the project's specialised editing software. With the improved options of correction, the researchers were able to further subdivide the segments, therefore rendering the changing textual affiliations in a more accurate manner. The working process of the edition has led to the segmenting being done for each preface or chapter in two iterations: The first is based on narrative structure (when reading and segmenting the manuscripts one by one), the second on comparison between the segmented manuscripts (once they are collated online).

In the first iteration, which is based on the subsequent reading of 6 -16 versions of a chapter in the process of establishing an edition, one segments the manuscripts in order to isolate *similar* passages with the purpose of aligning them in the edition. The segmentation is guided by structural elements, such as stories within stories, or framing and enframed narratives,[10] as well as dialogue and the change

9. See Mahmoud Kozae and Marwa M. Ahmed, Toward Usable and FAIR Software for Arabic Textual Scholarship. An Interim Report on the Editorial and Analytical Work of the AnonymClassic Project. *Medieval Worlds* 11 (2020), pp. 272-76.

10. For an introduction to framing narratives in Arabic literature and the *Kalīla wa-Dimna*, see: Stephan, Johannes and Beatrice Gruendler, From the Frame Tale to Framing Intertextuality: An Introduction, *Journal of Arabic and Islamic Studies* 24[1] (2024), pp. 5–24.

of speakers within it, and inserted wise sayings and analogies that illustrate and bolster the arguments made. However, the larger ones among these identified elements (such as long substories or lengthy statements of speakers in a dialogue) need to be further subdivided based on content. Furthermore, segmenting itself has its limits. On the one hand, to create segments, one needs to carefully weigh content vs. syntax. On the other hand, some changes cannot be shown by the process of segmenting but need commentary. Examples of this include cases where the wording of a phrase is moved and reapplied to a different plot step, or where part of a passage is relocated to another segment. Segmenting thus remains a rough approximation that both produces and works with a perspective that can be likened to a narrative "bird's eye view" – a basic sense of the overall layout of things, made from afar. This approach necessitates a follow-up, close reading analysis. Nonetheless, the first round of segmentation is sufficient in that it is able to align the versions side by side within a provisional multitext edition.

In the second iteration, the segments are adjusted by returning to each of the edited versions, now easily comparable in parallel columns. Those passages with similar content are then analysed for their *differences*, and added content present in only few witnesses is separated out into new segments to show their interrelation. Similar parts of passages that were previously undetected can now be easily analysed by further subdivision, and undetected passages aligning with others can be shifted between segments. Passages present in only a few versions (or one version) can be separated as segments of their own. Between the first and the second phase, the number of segments may double, reaching up to 374 within a given chapter.[11]

11. For example, within the chapter of "Dimna's Trial."

Figure 1 shows the result as a table of units, generated from the shortest chapter (Mc) as an example (based on the same data as the chapter's map). In the map, however, the direction of the narrative is shown from left to right. Each line represents the segments present in one manuscript and the shelf marks of all manuscripts used in the chapter's edition are provided as column on the left side.

seq.	unit title	I344	R2536	P400	P3465	P5881	P3475	P3473	R2407	BWII672	A4095	P3466	CCCP578	L8751	L4044	P3471
1	Mouse and Cat															
2	Chapter Title	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	Abstract	2							2	2					2	2
4	The King Asks about Escaping Enemies	3	2	2	2	2	2	2	3	3	2	2	2	2	3	3
5	The Philosopher Responds that Emotions Change	4	3	3	3	3	3	3	4	4	3	3	3	3	4	4
6	The Intelligent Adapts to Situations	5	4	4	4		4		5	5	4	4		4	5	5
7	Enmity Must Not Stop One from Requesting Help				5	4	5	4	6	6	5	5	4	5	6	6
8	One's Disposition toward Friends and Foes Changes															7
9	The Intelligent Man Could Make Friends or Foes															8
10	The Resolute Attains His Need				6	5		5	7	7	6	6	5	6	7	9
11	The Parable of the Cat and Rat Is Invoked	6	5	5	7	6	6	6	8	8	7	7	6	7	8	10
12	The King Requests the Full Parable	7	6	6	8	7	7	7	9	9	8	8	7	8	9	11
13	The Rat and the Cat Live Next to a Tree	8	7	7	9	8	8	8	10	10	9	9	8	9	10	12
14	The Cat Is Caught	9	8	8	10	9	9	9	11	11	10	10	9	10	11	13
15	The Rat Seeks Food	10	9	9	11	10	10	10	12	12	11	11	10	11	12	14
16	Foes Surround the Rat	11	10	10	12	11	11	11	13	13	12	12	11	12	13	15
17	The Rat Weighs His Predicament	12	11	11	13	12	12	12	14	14	13	13	12	13	14	16
18	Intelligence Is Revealed in Crises									15						
19	The Intelligent Always Keeps Alert			12	14	13	13	13	15		14	14	13	14	15	17
20	Intelligence Is Like the Sea	13	12	13	15	14	14	14	16	16	15	15	14	15		18
21	Situations Do Not Blind the Intelligent	14	13	14	16		15		17	17	16	16	15	16	16	19
22	The Rat Decides on a Pact with the Cat	15	14	15	17	15	16	15	18	18	17	17	16	17	17	20
23	The Rat Comes Up to the Cat	16	15	16	18	16	17	16	19	19	18		17	18	18	21
24	The Cat Acknowledges Being Trapped	17	16	17	19	17	18	17	20	20	19	18	18	19	19	22
25	The Rat Resolves to Be Sincere					18										
26	The Rat Points Out their Common Problem	18	17	18	20	19	19	18	21	21	20	19	19	20	20	23
27	The Rat Proves His Sincerity	19	18	19	21	20	20	19	22	22	21	20	20	21	21	24
28	People Fail without Mutual Trust					21	21		23	23	22	21	21	22	22	25
29	The Rat Promises Faithfulness					22	22	20	24	24	23	22	22	23	23	26
30	The Intelligent Does Not Postpone Action					23	23	21	25	25	24	23	23	24	24	27
31	The Cat Should Desire Common Rescue		19	20		24	24	22	26	26	25	24	24	25	25	28
32	Sailors Depend on Ships and Vice Versa	20	20	21	22	25	25	23	27	27	26	25	25	26	26	29
33	Cooperation Ensures Delivery for Both	21	21	22								26				30
34	Everything Happens According to Fate														27	
35	The Cat Puts His Trust in the Rat	22	22	23	23	26	26	24	28	28	27	27	26	27	28	31
36	The Rat Explains the Strategy	23	23	24		27	27	25	29	29	28	28	27	28	29	32
37	The Rat Explains the Detail of the Last Rope				24											
38	The Cat Accepts the Pact														30	
39	The Cat Fulfills His Part	24	24	25		28	28		30	30	29	29	28	29	31	33
40	The Foes Leave Frustrated	25	25	26	25		29	26	31	31	30	30	29	30	32	34
41	The Rat Approaches to Cut the Rope	26	26			29	30	27	32	32	31		30			
42	The Rat Begins to Cut the Rope			27	26			28	33	33		31		31	33	35
43	The Cat Deems the Rat's Efforts Slow	27	27	28	27	30	31	29	34	34	32	32	31	32	34	36
44	The Noble Must Help	28	28	29	28	31	32	30	35	35	33	33	32	33	35	37
45	The Cat States the Rat's Duty of Reciprocation				29	32	33	31	36	36	34	34	33	34	36	38
46	The Rat Should Forget Enmity	29	29	30	30	33	34	32	37	37	35	35	34	35	37	39
47	Good Deeds Erase Bad Deeds	30	30	31		34		33	38	38	36	36	35	36	38	40
48	The Punishment of Betrayal Comes Swiftly				31	35	35	34	39	39	37	37	36	37	39	41
49	The Rat Must Fulfill the Pact	31	31	32								38				42
50	The Rat Begins His Defence															
51	The Two Kinds of Friends	32	32	33	32	36	36	35	40	40	38	39	37	38	40	43
52	Trusting a Foe Is Like Eating Bad Food	33	33	34												44
53	The Intelligent Trades Favors				33	37	37	36		41	39	40	38	39	41	45
54	The Intelligent Protects Himself	34	34	35												46
55	Forced Friendship Does Not Last	35	35	36	34	38	38	37	41	42	40	41	39	40	42	47
56	Friendship Is Pursued For a Need					39	39			43	41	42	40	41	43	48
57	The Rat Acknowledges His Remaining Obligation	36	36													49
58	The Rat Explains the Reasons for Self-Protection					40	40	38	42	44	42	43	41	42	44	50

59	The Rat Resumes Cutting the Rope				35											
60	Ill-Timed Harvest							39	43	45						
61	Everything Has Its Time	37	37				41	40	44	46	43	44	42	43	45	51
62	Bad Timing Brings Failure	38	38					41	45	47	44	45		44		52
63	The Rat Explains Not Cutting the Last Rope	39	39	37		41	42	42	46	48	45	46	43	45	46	53
64	The Rat Cuts All but One Rope						43	43	47	49	46	47		46	47	54
65	The Huntsman Approaches	40	40	38	36	42	44	44	48	50	47	48	44	47	48	55
66	The Cat Expects to Die	41	41	39								49			49	56
67	The Cat Requests Fulfillment				37										50	
68	The Rat Resolves to Cut the Last Rope					43	45	45	49	51	48	50	45	48		57
69	The Rat Begins to Cut the Last Rope															
70	The Rat Fulfills the Pact				38	44	46	46	50	52	49	51	46	49	51	58
71	The Cat Breaks from the Last Rope	42	42	40												
72	The Cat and the Rat Escape	43	43	41	39	45	47	47	51	53	50	52	47	50	52	59
73	The Cat Calls Back the Rat	44	44	42	40	46	48	48	52	54	51	53	48	51	53	60
74	The Rat Refuses the Call					47										
75	The Cat Reminds the Rat of their Friendship	45	45	43	41	48	49	49	53	55	52	54	49	52	54	61
76	If One Gives Up a Friend One Loses Out				42	49	50	50	54	56	53	55	50	53	55	62
77	The Cat Suggests a Reward	46	46	44	43	50	51	51	55	57	54	56	51	54	56	63
78	The Cat Reassures the Rat	47	47	45		51	52	52	56	58	55	57	52	55	57	64
79	The Treatment of Friends or Foes	48	48	46			53								58	65
80	The Rat Distinguishes Hidden from Overt Enmity				44	52	54	53	57	59	56	58	53	56	59	66
81	Hypocricy Is the Worst Danger						55								60	
82	Riding an Elephant's Tusk				45	53	56	54	58	60	57	59	54	57	61	67
83	An Enemy Is Befriended in Times of Need				46	54	57	55	59	61	58	60	55	58	62	68
84	One Should Not Befriend an Enemy Indefinitely	49	49	47												69
85	Young Cattle Only Nurse for a While	50	50	48	47	55	58	56	60	62	59	61	56	59	63	70
86	Clouds and Rain Come and Go					56		57	61	63	60	62	57	60	64	71
87	The Intelligent Adapts					57	59	58	62	64	61	63	58	61	65	72
88	Enmity Turns into Friendship in Times of Need	51	51	49	48	58	60	59	63	65	62	64	59	62	66	73
89	Heated Water Cools Down	52	52	50		59	61	60	64	66	63	65	60	63	67	74
90	Eternal Enmity															
91	The Cat Has No Excuse															
92	The Cat Is the Rat's Worst Enemy	53	53	51		60	62		65	67	64	66	61	64	68	75
93	The Rat's Need for Friendship Has Gone			52		61	63	61	66	68	65	67	62	65	69	76
94	The Weak Must Avoid the Strong			53		62	64	62	67	69	66	68	63	66	70	77
95	The Cat Desires the Rat for Food	54	54	54		63	65	63	68	70	67	69	64	67	71	78
96	The Prudent Rat					64	66	64	69	71	68	70	65	68		79
97	The Cautious Weak Is Better Than Deluded Strong						67		70	72	69	71	66	69	72	80
98	The Intelligent Escapes Forced Friendship					65	68	65	71	73	70	72	67	70	73	81
99	Trusting Quickly Is Risky				49		69	66		74	71	73	68		74	82
100	The Intelligent Distrusts Forced Friendship				50		70	67	72	75	72	74	69	71	75	83
101	The Intelligent Must Leave Forced Friendship														76	
102	The Best Thing Is Distance From the Enemy				51		71	68	73	76	73	75	70	72	77	84
103	Staying Away from the Enemy is the Best Judgment						72	69	74	77	74	76	71	73	78	85
104	The Rat Will Love the Cat from Afar				52	66	73	70	75	78	75	77	72	74		86
105	The Rat Does Not Need a Reward				53		74	71	76	79	76	78	73	75	79	87
106	The Last Verdict Is No Reunion		55		54	67	75	72		80	77	79	74	76	80	88
107	The Rat Departs to Safety	55	56	55												89
108	The Philosopher Comments on the Rat's Foresight					68										90
109	The Philosopher Comments on Human Foresight					69										91
110	The Gist of the Chapter			56										77		
111	The Ending of the Chapter	56			55	70	76	73	77	81						92

Fig. 1 Table of units of the chapter of "The Rat and the Cat"

Through the twofold process of segmentation, the map of changing narratives becomes more fine-grained. The increased precision in the representation of diversity brings a new challenge: the overview constituted by parallel rows of the small green boxes, representing the plots steps, fluctuates among the versions. This strains the human eye (see figures 2, 4, 6, and 8). It was definitely possible to place

those versions most similar to each other by trial and error, until we had arrived at what we considered an optimal arrangement in which greater mutual distance of text columns in the edition represented greater difference. Yet, there were questions as to whether this could not be improved.

One might think, for instance, a chronological arrangement of the manuscripts would be the most logical. But manuscripts' dates are not a useful parameter, since a number of younger manuscripts contain text versions that are close to much older manuscripts, so that a physical witness's date only gives an *ante quem* of the text version contained in it.[12]

An added complexity is that manuscripts do not change independently from each other: some do use a single model (which they copy or rewrite), but others combine elements from several versions, and this we have made visible by segmenting the text accordingly. For instance, in the chapter of "The Lion and the Jackal," the necessity of investigating a legal case is illustrated by the analogy of testing wine through three senses: seeing color, smelling scent, and tasting flavor (segment 95). In some versions, the wine is replaced by meat (segment 96). Two manuscripts then combine both ways of sensorial testing.[13] The increasing precision of the display with the growing number of ever smaller segments offers a nearly dizzying fluctuation to the human eye.

---

12. For instance, Paris 5881, dated 1092/1681, copies Ayasofya 4214, datable before 761/1360; Paris 3475, dated 1175/1761, copies Rabat 3655, dated to c. 1265–80 CE; and Paris 3473, dated 1110/1699, copies Princeton 169H, dated 997/1588. Arch. Museum EY 344, falsely dated 410/1019 and probably belonging to 17-18th century) and New York 1981.373, datable before 1026/1617, both late manuscripts, contain text versions that are very similar to dated 14th-century manuscripts. Dates are given according to both the Islamic and Gregorian calendars separated by a slash.

13. Riyadh 2536, dated 747/1346 and Paris 3471, dated 1053/1643.

In the continued quest to make sense of the movements and trends of this proliferation, and in the attempt to reach an automated workflow to verify our observations, we combined two approaches. We both arranged segmented versions in an edition by eye, and applied machine learning methods, specifically clustering techniques, to decipher the complex relationships between the diverse manuscripts. When new manuscripts are added to the same chapter, and when working with new chapters, such combined techniques contributed to automating some steps of the analysis. Our ambition to unravel the intricate tapestry of text evolution may appear overwhelmingly challenging, but methodical and computational exploration offered an exciting pathway. Clustering techniques particularly appeared to provide a valid option for exploring the affiliations among manuscripts. Clustering is not a monolithic process; it involves a multi-layered approach with an array of techniques at each step that invite experimentation.

To discuss our methods of computational clustering, we will focus on illuminating one particular technique as a basis by which other approaches in future scholarship might be assessed and disseminated. This computational journey occurs in three steps. The first involves transforming a manuscript's narrative structure into a numerical problem. In this way, each manuscript, for every chapter, is represented as a numerical vector. Each position in this vector aligns with a narrative unit within the text. The presence or absence of a unit in a manuscript is thus denoted numerically: zero for absence, and for presence the numerical order of the unit's occurrence within the manuscript.

Following this numerical transformation, the second step is to select an appropriate algorithm by which manuscripts might be

clustered by similarity. The algorithm of choice for this work is Agglomerative Clustering. Its method begins with calculating pairwise similarity between data points and identifying the most analogous pair. Following this, the algorithm computes the centroid of this most similar pair. The centroid is a point that represents the average position of two or more data points, lying at an equal average distance from each of these points in all dimensions. The centroid is needed as a representative of a pair for the next level of calculation. In subsequent iterations, this centroid replaces the original pair, and this continues until all data points are paired. A threshold of minimum Euclidian distance determines which datapoint falls under which cluster. The later examples will demonstrate this visually.

The third step involves selecting the similarity metric for the algorithm. A ubiquitous method for calculating the distance between two vectors of equivalent size is the Euclidean Distance, expressed as:

$$\sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

where q and p denote two points in Euclidean n-space and i ranges from 1 to n, the number of dimensions. In the manuscript structure data, each q and p represent the order of a unit in the manuscript, while n is the total number of units.

In terms of operationalizing these three steps, the Python package SciPy offers an accessible and visually engaging platform.[14] Not only does it equip the researcher with the necessary tools to perform the operations, but it also generates dendrograms (tree diagrams). These hierarchical diagrams allow scholars to visually discern how the manuscripts relate to each other in terms of Euclidean distance.

14. See: https://scipy.org/, access: August 14, 2024, 1:50pm.

The algorithm requires the specification of a number that is manually chosen based on the nature of data and investigative questions. The number is used by the algorithm to determine the threshold of minimum Euclidian distance that divides the data points into the desired number of clusters. From our preliminary experience with the manuscripts, we have chosen to look for three clusters. This choice harmonises with our observation of the existence of two major *continua*[15] – i.e., two large internally cohesive groups – distinct from a number of early manuscripts as well as versions cutting across them, representing so-called "cross-copied" manuscripts made by copyist-redactors who combined various *Vorlagen* into their own new versions. Such is the initial outcome of this computational venture, laying the groundwork for subsequent, more nuanced analyses.

As the SciPy visualisation of the algorithm shows the relative distances between the manuscript versions in the form of a dendrogram, it links (along the vertical y-axis) those versions closest to each other in an angular bracket (or arch), then less similar ones again by bracketing those brackets and so forth until all manuscripts are placed within superposed brackets. Their horizontal extension (on the x-axis) marks incrementally the increased contrast between lower and higher brackets, and the colours are given by the algorithm to distinguish between parallel and superposed brackets of similarity. Besides the grouping into links, the horizontal length of bracket also shows the degree of closeness, that is, a bracket with short horizontal extension (or low arch) indicates a high degree of similarity, whereas

15. For a definition, see Beatrice Gruendler, Continuum: The Interrelation of some Arabic Versions of *Kalīla wa-Dimna*, in: *An Unruly Classic*: Kalīla and Dimna *and Its Syriac, Arabic, and Early Persian Versions*, ed. Beatrice Gruendler and Isabel Toral (Leiden, 2024). As discussed below, the major *continua* are the Paris continuum and the London continuum. A third continuum, the Queen continuum, switches between them but aligns with either in a given chapter. A fourth, the Iberian continuum, is close to the early Arabic manuscripts and moves across languages into Castilian, Hebrew, and Latin.

a bracket extending far to the right (or high arch) indicates low similarity.[16] This representation introduced a new mathematical metric, namely, the degree of difference between the individual manuscripts, since within an established group similarity, some members might be more or less similar to others. The algorithmic visualisation also displays the respective scope of difference for each chapter, that is, how much one chapter fluctuates in comparison to another. Finally, the different configurations of the arches (as more cascading and fitting hierarchically beneath each other vs. more parallel and next to each other) also visualises the qualitative *difference of the differences* among the chapters. The dendrograms of the four selected prefaces and chapters introduced above are hereafter juxtaposed to their maps and lead to several observations.

16. The order of the manuscripts on the y-axis is not relevant *per se*, what matters is the clusters they form.

The Preface of Ibn al-Muqaffaʿ

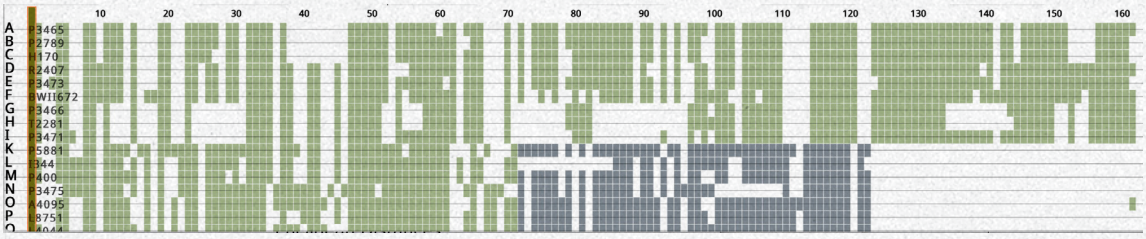


Fig. 2 Map of the preface of Ibn al-Muqaffaʿ

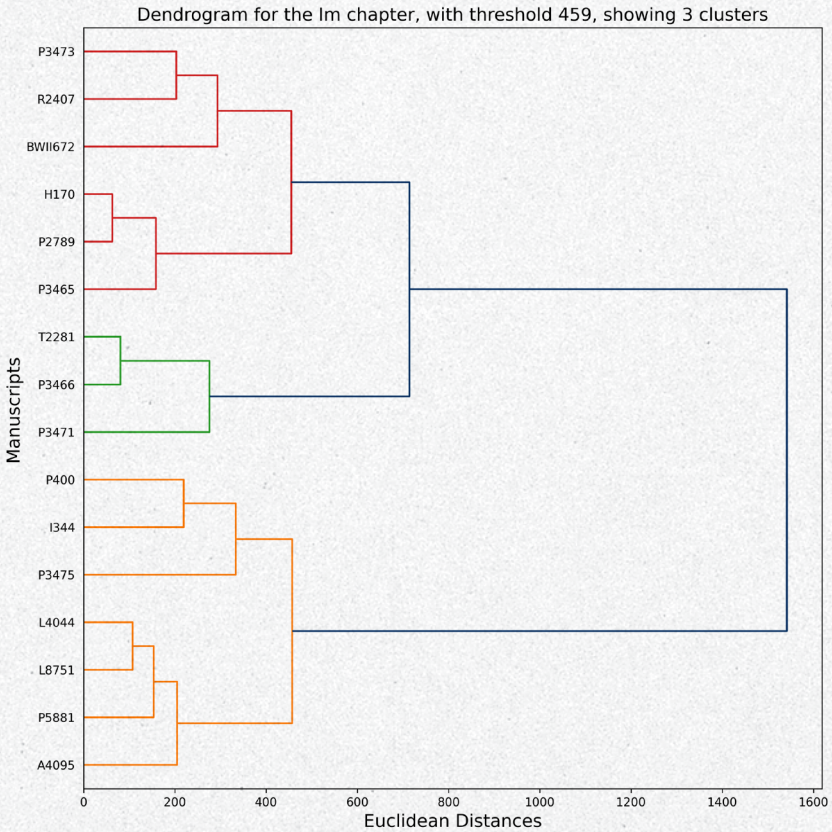


Fig. 3 Dendrogram, or branching diagram, showing relationships of similarity among manuscript versions of the preface of Ibn al-Muqaffaʿ

The preface of the Arabic translator-redactor Ibn al-Muqaffaʿ (abbreviated as Im) shows two large red and orange clusters, showing manuscripts that are relatively close to each other, as represented by their low brackets. This maps onto the two *continua* visible in the map, namely the Paris continuum (corresponding to the dendrograms's red cluster) and the London continuum (corresponding to the orange cluster). Indeed, the London continuum structurally differs from Paris continuum in that the enframed tale of "The Merchant and His Partner" falls into the preface's middle in the London continuum, whereas it is placed later in the Paris continuum and is followed by further enframed tales which appear only in the latter continuum. The changing positions of the segments are marked by their grey colour in the map. The visualisation shows that the two *continua* differ strongly from each other (the last bracket joining them is very high), whereas their internal resemblance within the continua is close (low brackets within the red and orange clusters). A third smaller green cluster corresponds to the Queen continuum (here including the cross-copied manuscript Paris 3471), which aligns with the Paris continuum in this preface but is partly abridged.[17] The result of this preface's clustering is that three fairly cohesive manuscript groups could be identified.

17. The Queen continuum is cross-copied at a large scale: it initially follows the Paris continuum and then shifts to the London-continuum in the chapter of "The Owls and the Crows" (abbreviated Oc).

The Chapter of “The Rat and the Cat”

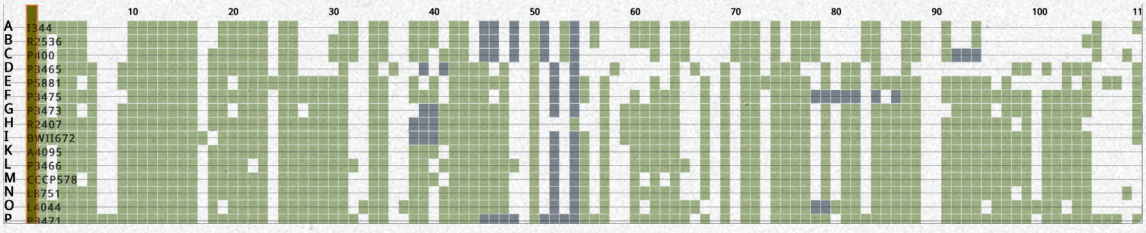


Fig. 4 Map of the chapter of “The Rat and the Cat”

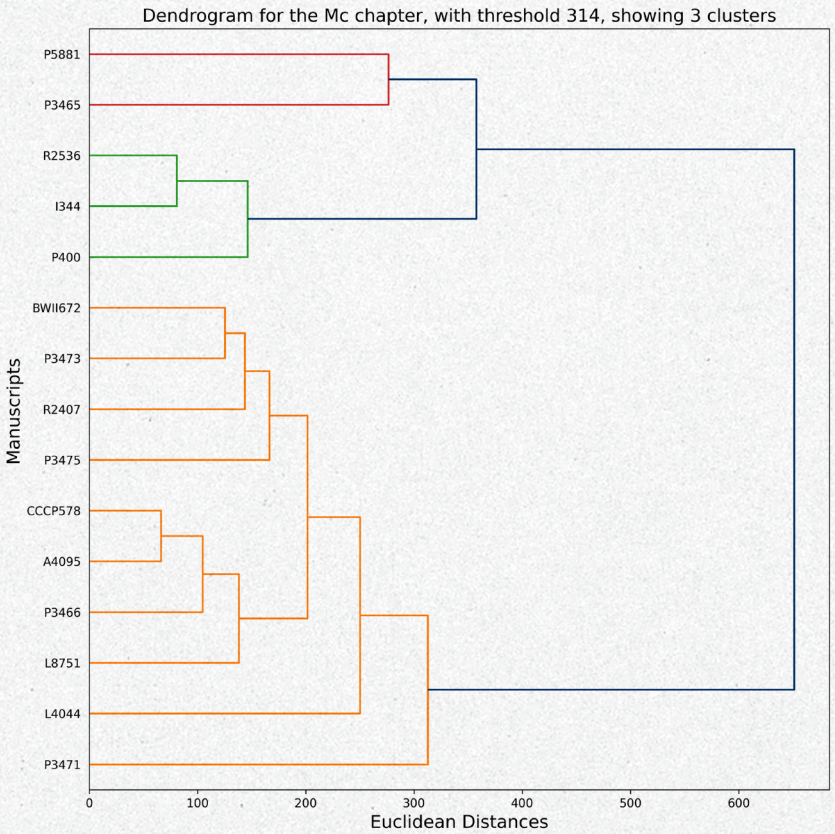


Fig. 5 Dendrogram of the chapter of “The Rat and the Cat”

In the relatively short chapter of "The Rat and the Cat" (abbreviated Mc), the clustering result is different. The two largest *continua* are structurally close to each other, so as to form a single, rather large orange cluster, of which the Paris continuum forms the upper branch and the Queen continuum, aligned with the London continuum, the lower branch. One manuscript (London 4044, attributed to the 15th century CE) receives a higher arch within the cluster due to a number of unique segments. Among these, the manuscript with the highest number of unique segments is the above-mentioned cross-copied manuscript (Paris 3471) which combines segments of various other manuscripts (92 segments of maximally 111, the highest number of any manuscript (see fig. 1), visible in the map's nearly uninterrupted bottom row of boxes. This dendrogram shows that the larger continua (Paris and London) share a structure, while their mutual difference is not articulated on that level but can only be ascertained by close comparative reading of their segments' texts.

Distinct from these is a small green cluster consisting of an early manuscript (Pococke 400, dated 755/1354), which belongs to those that fall outside any continuum, and two manuscripts belonging to the Iberian continuum, which extends from Arabic to Castilian, Hebrew and Latin, and from Latin to numerous modern languages.[18] Both form a sort of backdrop against which the larger Paris and London *continua* take shape. The great difference from the *continua* and the closeness to the earliest layer of manuscripts (here represented by Pococke 400), which the dendrogram displays, suggest that the Iberian continuum historically formed earlier than the other *continua*, a hypothesis that remains to be investigated.

18. Beatrice Gruendler and Khouloud Khalfallah, Anthologizing *Kalīla wa-Dimna*: The Incipit of the *First Risāla, Journal of Abbasid Studies* 9 (2022), 105–160, esp. p. 120 note 27 and p. 132 note 52. https://doi.org/10.1163/22142371-00802009, access: August 14, 2024, 1:50pm..

The red pair is constituted by two manuscripts that both abridge the chapter, yet in different ways, visible by their relatively high bracket.

The Preface of “The Voyage of Burzoy” (Long Version)

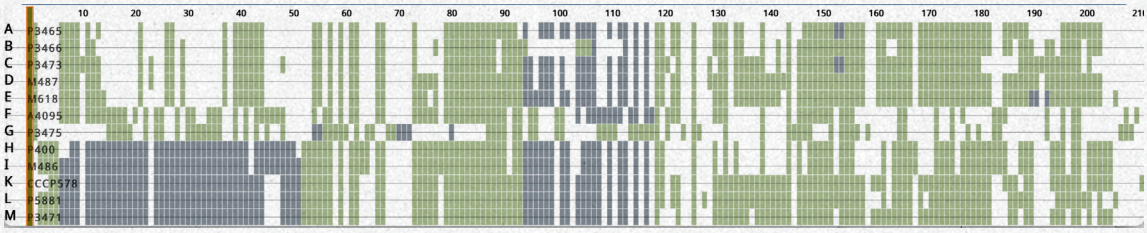


Fig. 6 Map of the preface of “The Voyage of Burzoy” (long version)

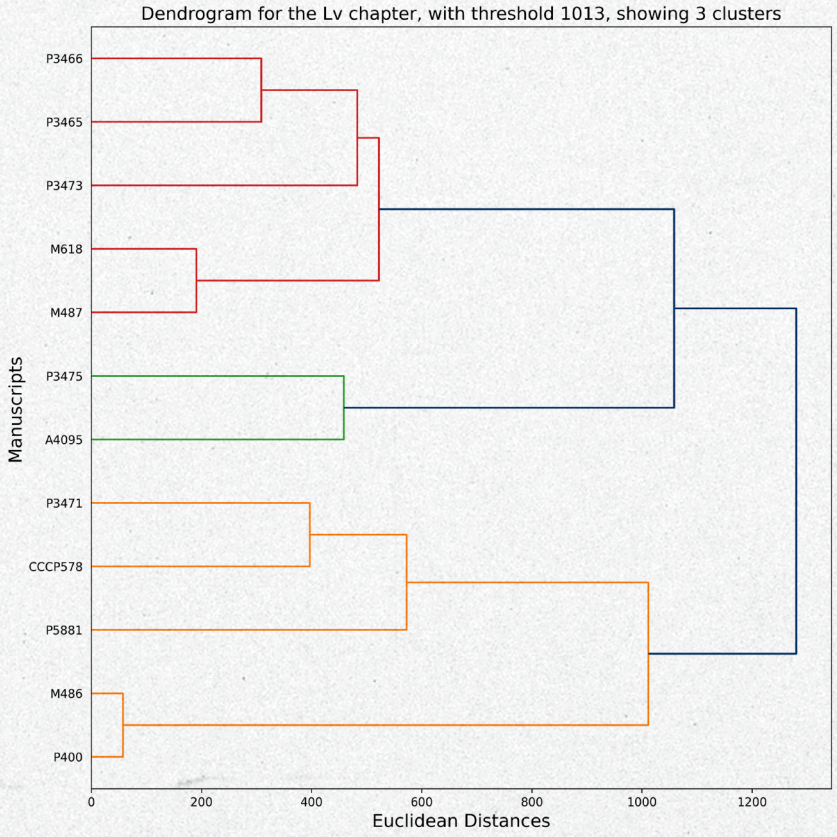


Fig. 7 Dendrogram of the preface of “The Voyage of Burzoy” (long version)

The preface of "The Voyage of Burzoy" (long version, abbreviated Lv) showcases a particular structural innovation. The red top cluster does not take part in this; its manuscripts all belong to the Paris continuum as well as the Queen continuum, aligning with it here.

The orange cluster at the bottom represents the innovation, it assembles a number of manuscripts that share the relocation and addition of a part of the plot, namely a public speech of the traveling scholar Burzoy to announce his mission to bring *Kalila wa-Dimna* from India to Persia on the behest of the Sasanian king Kisrā Anūshirwān (Husrav, r. 531–579). The addition of this speech gives the figure of the scholar a particular weight and strengthens the chapter's focus on the value of knowledge and learning. The preface's beginning is relocated and placed into this scholar's mouth, and a new introduction added. The manuscripts sharing this innovation have since been identified as belonging to a structural group, the First *Risāla* Group, further characterized by an added incipit which makes *Kalīla wa-Dimna* usable as an anthology.[19]

A green pair of idiosyncratic manuscripts shows an interim moment, since both resemble part of the restructuring of the orange cluster, but still lack the scholar's speech. Both are strongly abridged and one of them (Paris 3475) changes the plot in the middle of the narrative (visible by the grey boxes of all other manuscripts in the map).

19. Gruendler and Khalfallah, Anthologizing *Kalīla wa-Dimna*.

The Chapter of “The Lion and the Jackal”

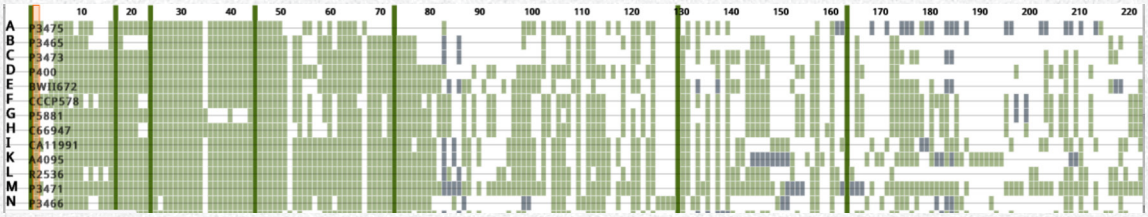


Fig. 8 Map of the chapter of “The Lion and the Jackal”

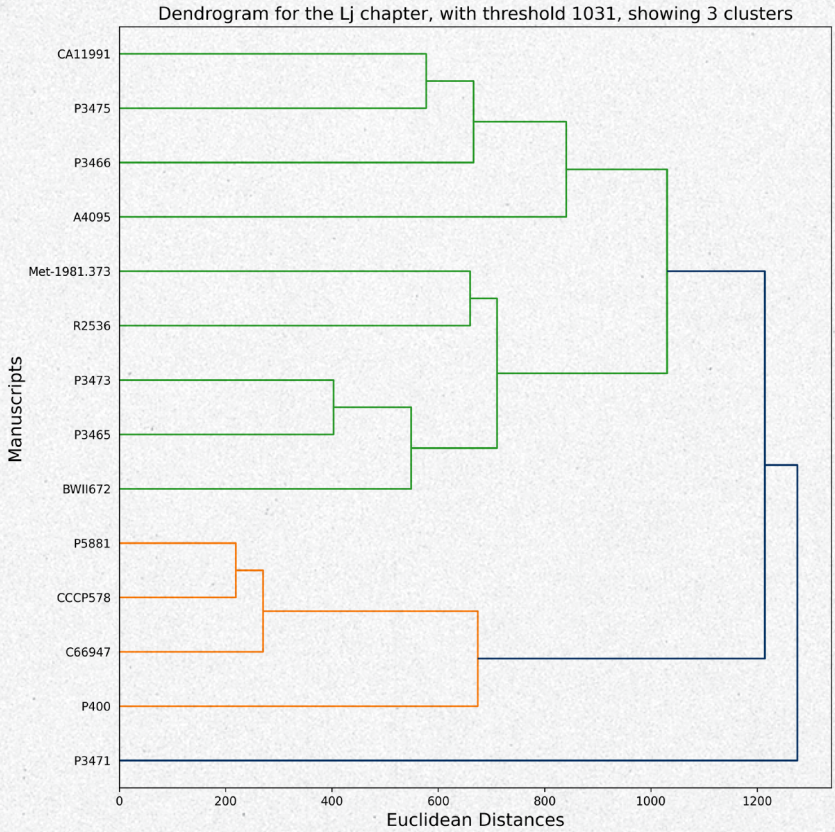


Fig. 9 Dendrogram of the chapter of “The Lion and the Jackal”

In the chapter of the "Lion and the Jackal" (abbreviated Lj), yet another configuration is obtained. This chapter shows an unusually large variety of *individual* versions which overshadows any coherence via *continua*. This occurs particularly in the chapter's second half, where the characters discuss a case of royal misjudgement and injustice. In the last three scenes, separated in the map by vertical lines are the two dialogues of the lion king with his mother, and a dialogue of lion king with the courtier he mistreated, who is a jackal. The rewriting that has been seen to be incremental in the chapters previously discussed increases in these three scenes, producing a large number of rare and unique segments in individual manuscripts, resulting in generally high brackets within the green cluster.

A smaller orange cluster results from two early manuscripts, the above-mentioned Pococke 400 (see Mc chapter) and Parker 578, dated to the 14th century CE. That latter version was reused in a cross-copied manuscript Ayasofya 4214, datable before 761/1360 (for the editions, its illustrated copy Paris 5881, dated 1092/1681, is used). Paris 5881 is in turn slightly rewritten in another manuscript (Azhariyya 66947, dated 1048/1639). The cross-copied manuscript Paris 3471 stands apart by the high number of its segments combined from different *Vorlagen* (third row from the bottom in the map).

The orange cluster's closer affiliations constitute, however, a minority in the Lj chapter, which shows an extreme case of interference by copyist-redactors down to recognisable individual profiles. This extends even to the chapter's outcome, which ranges in the manuscripts from reconciliation between the two protagonists of lion king and jackal and their irrevocable break up. This can be motivated

by the chapter's provocative ethical and political dimensions, which have been investigated elsewhere.[20]

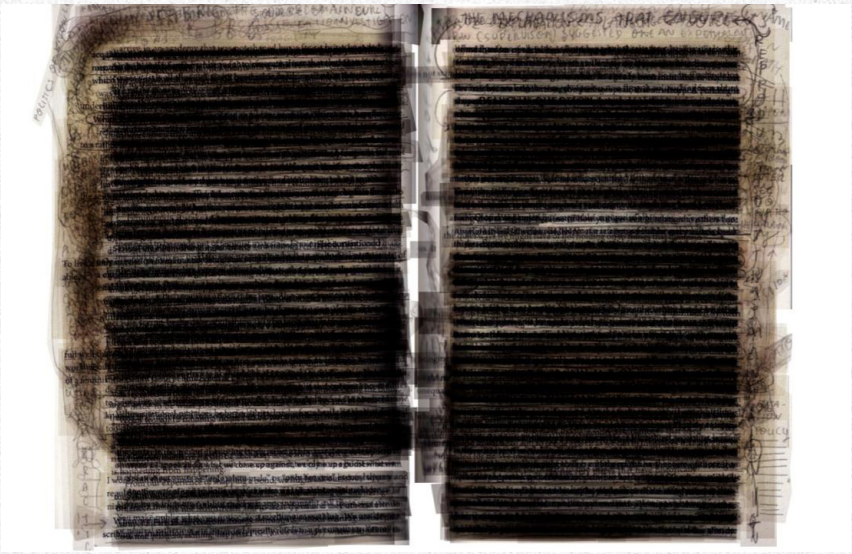
In summation, the clustering experiment has been beneficial in reducing the complexity of a massive textual variation, by giving a metric to gage the versions' difference in type and degree, chapter by chapter. Clusters of *continua* evolving along gradual rewriting in one chapter (Im) alternate with individual interventions in another (Lj). An overall cohesion of continua in a further chapter (Mc) contrasts with shared structural innovation in yet another (Lv). Our mathematical investigation has provided a yardstick by which countless barely-known or anonymous copyist-redactors between the thirteenth and the nineteenth century can be assessed and compared for their degrees of intervention, and range of creativity.

This, in turn, opens new questions about the themes and issues triggered by different types of writerly responses. The dendrograms show how the algorithm assigns widely divergent threshold values,[21] which divide the datapoints of each chapter into the three desired clusters, indicating the degree of internal difference of each chapter's versions. This threshold value is low in Mc (314) and Im (459) and high in Lv (1013) and Lj (1031). This leads to the question of which factors might be at work here, potentially a chapter's structure and style. Regarding structure, the presence of many subtales in a chapter, such as in Im, provide indeed fixed points that give the text some stability (Mc is very short and not relevant in this respect). Both Lv

20. For a close analysis of the chapter, see Beatrice Gruendler with Oualid El Khattabi and Mahmoud Kozae (contributing authors), Miscarriage of Justice and Dissenting Re(d)actions in *Kalīla and Dimna*, in: *Variants in Classical Textual Traditions: Errors, Innovations, Proliferation, Reception?* ed. Glenn W. Most (Berlin 2024), pp. 337–376.

21. The threshold value is algorithmically calculated based on the distances between the points and the specified number of desired clusters.

and Lj inversely lack subtales, which translates into greater structural flexibility. As for style, conversation and debate permit change more easily than a narrative, even more so if the topic is of relevance to redactors and prone to provoke intervention, such as royal mistreatment and injustice. Indeed, both Lv and Lj contain much conversation, and in the latter chapter, this is mostly adversarial debate in which the protagonists try to persuade or contradict each other. To wrap up, the experiment has shown how eye and algorithms can complement each other, and it may serve as argument or provide encouragement to continue along this path.



Artwork by Elvia Vasconcelos. From the Artistic Research Project “Sketches as a Conversation Interface” conducted while Dorothea Schlegel Artist in Residence at the Cluster of Excellence “Temporal Communities: Doing Literature in a Global Perspective”, 2021.

*****

<END>

List of Manuscripts

(The short abbreviations in parentheses are used in the figures.)

Arch. Museum EY 344 (I344) = Istanbul, Archaeology Museum (Arkeoloji Müzesi), EY 344 (dated 410/1019, and tentatively attributed to the 18th century)

Ayasofya 4095 (A4095) = Istanbul, Süleymaniye Kütüphanesi, Ayasofya, 4095 (dated 618/1221)

Ayasofya 4214 = Istanbul, Süleymaniye Kütüphanesi, Ayasofya, 4214 (datable before 761/1360)

Azhariyya 66947 (C66947) = Cairo, al-Azhariyya, Ms. 66947 (1048/1639)

Chicago 11991 (C11991) = Chicago, Oriental Institute, A 11991 (attributed to the 17th century CE)

Hamburg 170 (H170) = Hamburg, Staats- und Universitätsbibliothek, cod. orient. 170 (dated before 1321/1904)

London 4044 (L4044) = London, British Library, Or. 4044 (attributed to the 15th century CE)

London 8751 (L8751) = London, British Library, Or. 8571 (dated 799/1396)

Manchester 2 (M486) = Manchester, John Rylands Library, Ms. 2 [formerly 486] (dated 1040/1631)

Manchester 537 (M487) = Manchester, John Rylands Library, Ms. 537 [formerly 487] (dated 1083/1672)

München 618 (M618) = München, Bayerische Staatsbibliothek, Cod. arab. 618 (dated 1044/1634)

NewYork 1981 (Met-181-373) = New York, Metropolitan Museum of Art, Ms. 1981.373 (datable before 1026/1617)

Paris 2789 (P2789) Paris, Bibliothèque nationale de France, arabe 2789 (dated 1041/1631-1632)

Paris 3465 (P3465) = Paris, Bibliothèque nationale de France, arabe 3465 (dated to c. 1220 CE)

Paris 3466 (P3466) = Paris, Bibliothèque nationale de France, arabe 3466 (datable before 854/1450)

Paris 3471 (P3471) = Paris, Bibliothèque nationale de France, arabe 3471 (dated 1053/1643)

Paris 3473 (P3473) = Paris, Bibliothèque nationale de France, arabe 3473 (dated 1110/1699)

Paris 3475 (P3475) = Paris, Bibliothèque nationale de France, arabe 3475 (dated 1175/1761)

Paris 5881 (P5881) = Paris, Bibliothèque nationale de France, arabe 5881 (dated 1092/1681)

Parker 578 (CCCP578) = University of Cambridge, Corpus Christi College, Parker Library, 578 (dated 8th/14th century CE)

Pococke 400 (P400) = University of Oxford, Bodleian Library, Pococke 400 (dated 755/1354)

Princeton 169H = Princeton University Garrett Library 169H [Hitti no. 176] (dated 997/1588)

Rabat 3655 = Rabat, Bibliothèque royale de Rabat, Ms. 3655 (dated to ca. 1265–1280 CE)

Riyadh 2536 (R2536) = Riyadh, King Faisal Library, Ms. 2536 (dated 747/1346)

Riyadh 2407 (R2407) = Riyadh, King Faisal Library, Ms. 2407 (dated 1103/1691)

Tunis 2281 (T2281) = Tunis, Bibliothèque nationale de Tunisie, Ms. 2281 (dated 1076/1666)

Wetzstein II 672 (BWII672) = Berlin, Staatsbibliothek zu Berlin, Wetzstein II 672 (dated 1246/1830)

References

Bremer, Thomas, Paul Molitor, Marcus Pöckelmann, Jörg Ritter, and Suzanne Schütz. Zum Einsatz digitaler Methoden bei der Erstellung und Nutzung genetischer Editionen gedruckter Texte mit verschiedenen Fassungen – Das Fallbeispiel der *Histoire philosphique des deux Indes* von Guillaume Thomas Raynal. *Editio* 29[1] (2015), pp. 29–51.

Gruendler, Beatrice, The Interrelation of some Arabic Versions of *Kalīla wa-Dimna* as Continua, in: *An Unruly Classic:* Kalīla *and Dimna and Its Syriac, Arabic, and Early Persian Versions*, ed. Beatrice Gruendler and Isabel Toral (Leiden: Brill, 2024), pp. 19–68.

Gruendler, Beatrice et al., An Interim Report on the Editorial and Analytical Work of the AnonymClassic Project. *Medieval Worlds: Comparative and Interdisciplinary Studies* 11 (2020), pp. 241–279. https.://doi.org/10.1553/medievalworlds_no11_2020s241, access: July 3, 2025, 4:15pm

Gruendler, Beatrice and Khouloud Khalfallah, Anthologizing *Kalīla wa-Dimna*: The Incipit of the *First Risāla*. *Journal of Abbasid Studies* 9 (2022), pp. 105–160. https://doi.org/10.1163/22142371-00802009, access: July 3, 2025, 4:12pm.

Gruendler, Beatrice with Oualid El Khattabi and Mahmoud Kozae (contributing authors), Miscarriage of Justice and Dissenting Re(d)actions in *Kalīla and Dimna*, in: *Variants in Classical Textual Traditions: Errors, Innovations, Proliferation, Reception?* ed. Glenn W. Most (Berlin: De Gruyter, 2024), pp. 337–376.

Hirschler, Konrad, *The Written Word in the Medieval Arabic Lands: A social and cultural history of reading practice* (Edinburgh: Edinburgh University Press, 2012).

Keegan, Matthew L. and Beatrice Gruendler, Before and after Ibn al-Muqaffa': Introduction to the Special Issue (Un)declared Co-Authorship in the *Kalīla wa-Dimna* Textual Tradition. *Journal of Abbasid Studies* 9 (2023), pp. 1–7; https://doi.org/10.1163/22142371-00802010, access: July 3, 2025, 4:12pm.

Stephan, Johannes, A Sense of No Ending: *Kalīla and Dimna* between Fluidity and Completeness. *How to End Things in Arabic Literature,* ed. Lale Behzadi und Bilal Orfali, Special Issue *Al-Abhath* 71 [1–2] (2023), pp. 53-81, https://brill.com/view/journals/alab/71/1-2/article-p1_2.xml, access: January 25 2024, 10:07am.

Kozae, Mahmoud and Marwa M. Ahmed, Toward Usable and FAIR Software for Arabic Textual Scholarship. An Interim Report on the Editorial and Analytical Work of the AnonymClassic Project. *Medieval Worlds* 11 (2020), pp. 272–276.

Schütz, Susanne and Marcus Pöckelmann. LERA – Explorative Analyse komplexer Textvarianten in Editionsphilologie und Diskursanalyse, in: *Book of Abstracts of the Third Annual Conference of Digital Humanities for German-Speaking Regions* (2016), pp. 239–43.

Stephan, Johannes and Beatrice Gruendler. From the Frame Tale to Framing Intertextuality: An Introduction. *Journal of Arabic and Islamic Studies*, 24[1] (2024), pp. 5–24.