READ for Solving Manuscript Riddles: A Preliminary Study of the Manuscripts of the 3rd ṣaṭka of the Jayadrathayāmala

Serbaeva, Olga ; White, Stephen

DOI: https://doi.org/10.1007/978-3-030-86159-9_24

Originally published at:
Serbaeva, Olga; White, Stephen (2021). READ for Solving Manuscript Riddles: A Preliminary Study of the Manuscripts of the 3rd ṣaṭka of the Jayadrathayāmala. In: Barney Smith, Elisa; Pal, Umapada. Document Analysis and Recognition – ICDAR 2021 Workshops, Lausanne, Switzerland, September 5–10, 2021 Proceedings, Part 2. Cham: Springer, 339-348.
DOI: https://doi.org/10.1007/978-3-030-86159-9_24

READ for solving manuscript riddles: a preliminary study of the manuscripts of the 3rd *ṣaṭka* of the *Jayadrathayāmala*

Olga Serbaeva[1][0000−0002−0103−8284] and Stephen White[2]

[1] University of Zurich, Switzerland
[2] Ca'Foscari University of Venice, Italy olga.serbaeva@aoi.uzh.ch; stephen.white@unive.it

Abstract. This is a part of an in-depth study of a set of the manuscripts related to the *Jayadrathayāmala*. Taking JY.3.9 as a test-chapter, a comparative paleography analysis of the 11 manuscripts was made within READ software framework. The workflow within READ minimized the effort to make a few important discoveries (manuscripts containing more than one script, identification of the manuscripts potentially written by the same person) as well as to create an overview of the shift from Nāgarī to Newārī and, finally, to Devanāgarī scripts within the history of manuscript transmission of a single chapter. Exploratory statistical analysis in R of the syllable frequency in each manuscript, based on the paleography analysis export from READ, helped to establish that there are potentially two lines of manuscript transmission of the JY.3.9.

Keywords: Sanskrit manuscripts · Indic scripts · Paleography · Virtual Research Environment

1 Introduction

1.1 Brief history of the manuscript transmission of the *Jayadrathayāmala*

The *Jayadrathayāmala* (further JY) is a text compiled in Northern India around the end of the 9th-10th century, and it was cited on multiple occasions by Kashmiri polymath Abhinavagupta in his *Tantrāloka*. However, none of the manuscripts from Kashmir seem to have reached the academia. The manuscript transmission is almost exclusively Nepalese: the Nepal-German Manuscript Preservation project (NGMPP) lists about 30 *Jayadrathayāmala*-related codices. Prof. Dr. Diwakar Acharya, University of Oxford[3], suggested that the oldest existing manuscript of this text comprising all four books, palm-leaf JY_B here, was written in Vārāṇasī around the time of the reign of the king Jayachandra (1170-1194). Theoretically, all Nepalese transmissions of the JY could stem from this single palm-leaf.

[3] Pers. comm. 05.04.2021

1.2 Aim and methods of the present study

The aim of the present study is to verify if all Nepalese transmissions indeed could have come from this single source, or if the paleography and statistical facts would allow any other representation of the history of transmission of JY. A secondary aim of this phase of the study is to test, whether Research Environment for Ancient Documents (READ)'s paleography features can be helpful for resolving the above named issue.

The preliminary results of this phase are based on a single chapter of the 3rd book, namely, JY.3.9, dealing with an invocation and the worship of the goddess Pratyaṅgirā. This is a short chapter, comprising some 73 verses, and it is situated rather towards the middle of the text, i.e. the script shall be much less calligraphic as it is on the initial folios of a book. The chosen chapter survived in 8 manuscripts of the JY_A, B, C, D, E, G, I, K and in 3 manuscripts of the text called the *Jayadrathayāmalamantroddhāraṭippanī*, (further JYMUṬ_A, B, O). It is an independent compilation of JY materials, which often includes the whole chapters.

The overall methodology can be divided into two procedures READ[4] Paleography analysis and R statistical analysis. We use READ to segment and group the individual characters of the text, as well as hand writing characterization through side by side visual inspection of each set of characters across the different transmissions using READ's generic tagging/classification feature. We use the export of these results as the input to the R-based statistical analysis. Having mapped the syllable frequency within each manuscript, we shall cluster the manuscripts by their proximity in relation to their particular syllabic portrait. This should point to the number of independent lines of transmission.

2 Procedure

2.1 READ as a Virtual Research Environment (VRE) for working with manuscript materials

Producing a curated digital edition is at the center of READ's design along with a data model in which heavily interlinked data used to support the variety of workflows for the different language types. READ has projects in (1.) syllabic Indic languages, (2.) alphabetic Latin, Greek, Hebrew, (3.) logo-syllabic Mayan. The workflow consists of adding a manuscript image (via upload or url), marking segments on the manuscript image, adding a transcript, linking the transcript and the marked segments together, and visually inspecting and classifying characters using the paleography tool.

READ is centered around the concept of a transcription (interpretation) encoded to match the original markings on the artifact. Markings are annotated by segments (boundaries). The linkage between segments and encoding of each letter/syllable/glyph of the document is preserved throughout the workflow (Fig. 1).

[4] READsoftware 2021: https://github.com/readsoftware/read/wiki

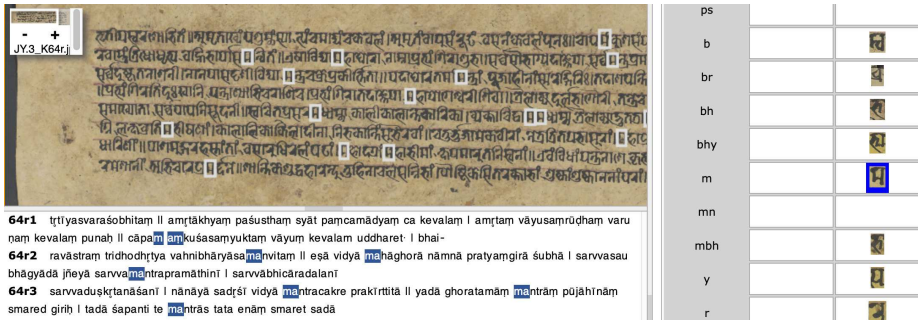

Fig. 1. Direct links between Paleography report (right), Document (top) and Edition (bottom) for the syllable "ma" of JY.3.9_K selected. Manuscript image: Digitale Sammlungen, Staatsbibliothek zu Berlin

2.2 Workflow in READ

Having obtained the manuscript images, one should find out if there are any missing folios, wrongly numbered pages, and work out the manuscript structure by assigning coherent filenames to the images of each side of the folio.

For importing transcriptions, READ has a data/constraint driven parser to validate the transcription before creating the highly link data model. One should ensure the transcription adheres to the defined encoding constraints for the language variation used (here Sanskrit, for example, transcription 1 will validate, while transcription 2 - will not.). During import READ will identify deviations

Word on the surface	Translation	Transcr. 1	Transcr. 2
	"nectar"	amṛ̣taṃ	amṛtaṃ

Fig. 2. Various transcription standards of the same word, of which one is in the data model, and the other is not.

which should be adjusted in the READ constraint data or changed in the transcription before import. This constitutes the one of three transcription coherency verification flowlets built into READ.

Having imported the document and uploaded the image(s), segmentation of every glyph (letter/syllable and punctuation mark) necessary for the research is completed by a semi-automated linking process. For manual segmentation, one should expect to spend about 10 minutes for marking up 500 glyphs.

The second stage of built in verification happens during the process of linking. When hovering over letter/syllable/word in the transcription with the mouse, the corresponding glyph(s) highlight on the image. Having located any misalignments, one should correct the edition or relink.

The third, and the final of level of verification comes when one works with the Paleography table. It is a grid where each cell represents the set of a single type of glyph found in the document. The exact number of rows and columns in the table is defined by the type of language. For example, for Greek one will only see the alphabet in one column, for Sanskrit a consonant vs vowel matrix will be generated (Fig. 3).

Fig. 3. Paleography report: partial grid showing consonants as rows and vowels as columns, JY.3.9_B. Digitally modified segments of the manuscript JY_B, copy obtained via NGMPP.

READ shows one example of the glyphs for a given cell, clicking on it will open a presentation of all linked glyphs. In this side-by-side presentation one will immediately see any intruders (clicking on the intruder identifies its location in the document). Thanks to this triple coherency check, close to 100 percent correctness of reading can be achieved. This level of precision as well as preserved links between segments and glyphs constitute a solid basis for further analysis. Fig. 4).

Fig. 4. Paleography for Newārī "a" of BT-1 type by order, JY.3.9_A. Digitally modified segments of the manuscript JY_A, copy obtained via NGMPP.

In this side-by-side presentation, all segments can be tagged with one of several values in each of up to 3 categories (work in progress to extend this to 6 categories), a segment belongs to "unknown" until it is classified. All categories are independent from one another, thus one has sufficient number (10) of values per category (3) to mark any desired glyph-particularities.

3 Selected discoveries

3.1 Manuscripts with multiple scripts/letter variants

While checking each letter in the paleography tables, some important manuscript characteristics became clear. JY_B and JYMUṬ_B within a single chapter and often within a single word mix two different scripts. For JYMUṬ_B there is a switch from Devanāgarī to probably one that is native to the scribe, namely, Newārī script (distribution of the two is about 50-50).

As for JY_B, going through the paleography report glyph by glyph made it clear that under the same syllables (consonant plus "e" or "o") there were two very different variations of glyphs (Fig. 5). The actual script is yet to be deter-

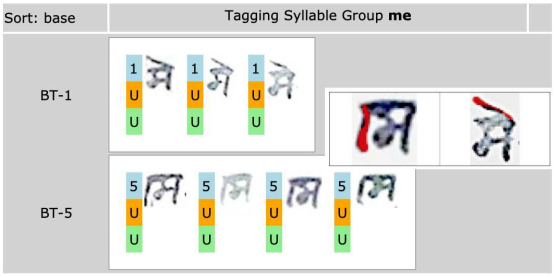

Fig. 5. Variation of "me" in JY.3.9_B made apparent in READ. In the close-up vowel sign "e" is highlighted in red for each variant. Digitally modified segments of the manuscript JY_B, copy obtained via NGMPP.

mined, but it has close similarities with the North-Indian variety of Nāgarī. The documents written in Nāgarī between the 10th and the 12th centuries, besides having similar shapes of glyphs, demonstrate the same variations concerning the writing of the consonant combinations with "e" and "o" [5]. This fascinating script inconsistency can point to a combination of the two different sub-types of the Nāgarī script attested within a single manuscript. If the same incoherence is found in the other 3 *ṣaṭkas* of the palm-leaf manuscript of JY_B), it would argue for the idea of Diwakar Acharya that all palm-leaf manuscripts of JY indeed constitute a set belonging to the same period and probably written by the same scribe.

3.2 Identification of the same or very similar handwriting

By simply opening up two paleography reports for each combinations of two of the selected 11 manuscripts side-by-side in READ, it became apparent that JY_D and JYMUṬ_O, separated in time by some 22 years according to their dating,

[5] Based on Gupta [n.d.]. See also Fig. 4 here.

are most probably written by the same person. The initial idea in READ was further confirmed by the similarity of the syllabic portrait of the two in the complete paleography set comprising about 690 different syllables (Fig. 6).

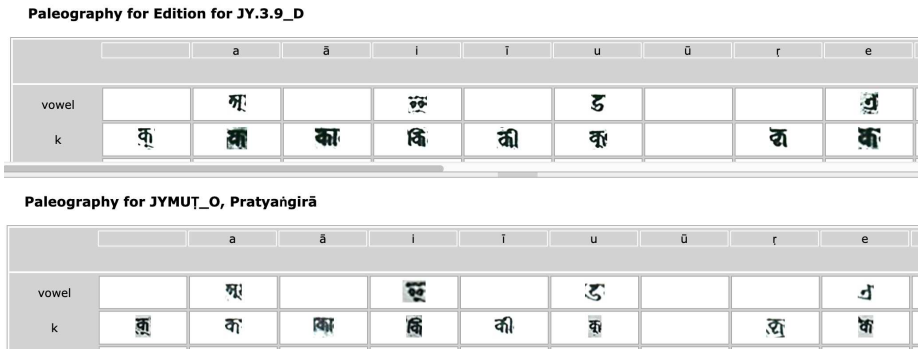

Fig. 6. Paleography of JY_D and JYMUṬ_O side-by-side. Digitally modified segments of the manuscripts, copies obtained via NGMPP.

3.3 Overview of the JY scripts from 12th to the 20th centuries

Exporting the encoded and categorised segments from READ has allowed the creation of a compiled overview table, where the manuscripts are arranged according to their dates (when known). The table also represents the main differences between (1.) Nāgarī, (2.) Newārī and (3.) Devanāgarī scripts in their historical transition from 1 to 3 based on the 11 manuscript variants of the same chapter of the JY (Fig. 7).

3.4 Exploratory statistical analysis in R of the data produced by READ Paleography report

It is still impossible to cluster the manuscripts based on the varieties of handwriting, but we are in the process of designing the methodology that uses Computer Vision (the present data set (JY.3.9) will serve as training data) to process all 201 chapters of the JY. However, even the paleography report constitutes a solid basis for the statistical analysis in R code adapted for the project.[6] For example, having taken just the file names of each exported segment (of the type bha_bt1_63r4.10.png, where the syllable "bha", classified under "bt1", appears on the folio "63r", line "4", syllable number "10"), we have first generated a syllabic portrait of each manuscript, i.e. how many times a syllable appears in a

[6] For clustering function dfm_weight, library quanteda in R. Visualisation - pheatmap. We thank Dmitry Serbaev, Omsk, Russia, for his expert advise concerning R.

Fig. 7. Selected digitally modified segments (all copies obtained via NGMPP, except for JY_K, belonging to Staatsbibliothek zu Berlin) from the 11 chosen manuscripts demonstrating the instability of scripts in JY_B (switch from Nāgarī to Newārī) and JYMUṬ_A and B (from Newārī to Devanāgarī) as well as the outline of change from Nāgarī to Newārī and to Devanāgarī script in the process of transmission of the JY.3. JYMUṬ_A, contains Newārī characters only occasionally.

given manuscript. From that the mean frequency was calculated, and based on a difference from that mean, we have found that a set of just 59 syllables (punctuation excluded) was sufficient to describe the major differences between the 11 manuscripts for the same chapter. These include the preferences for single and duplicated letters (*sarva* versus *sarvva*), for rendering the nasals (Pratyaṅgirā versus Pratyaṃgirā), and the final *sandhi*, i.e. the letters at the end of the word.[7]

One could argue that orthography could be accidental and therefore cannot be used to identify the manuscript relationships. However, the present case is a particular one: we are talking about 11 manuscripts of the same chapter written in verses, which means that the structure is particularly stable, and these manuscripts are the same for about 98 percent of their material. Thus, the tiny differences found were rechecked for stability and significance, i.e. singular scribal mistakes were separated from the reoccurring, statistically significant patterns that were made apparent during the clustering process. All this is represented here as heatmap (Fig. 8).

The difference in the manuscript cluster visualisation is represented by height. One line is stemming from JY_B based on syllabic portrait similarity and including JY_A, E, I, of which only the last two are very closely related; the other is rather related to JYMUṬ transmission variant, and including the remaining 7 manuscripts, of which JYMUṬ_B and JYMUṬ_O and JY_K and JY_D are particularly close to each other. This is a preliminary result based on a single chapter, however, it already shows that this chapter has at least two independent lines of transmission, one rooted in JY_B, but another one is coming from an independent manuscript, from which both JYMUṬs and JY_C, G, D, K stem. Most likely it was very close to JY_C.

4 Conclusion

We have discovered important variations in the scripts of two manuscripts, for JY_B it confirms the dating around the 12th century suggested by Diwakar Acharya, however, further work is needed to ascertain the precise varieties of the Nāgarī script; the JYMUṬ_B attests the tendency to change the script from Newārī to Devanāgarī at the beginning of the 18th century. Paleography analysis has allowed the establishment that JY_D and JYMUṬ_O were likely written by the same scribe. We have also compiled a table representing the evolution of the Newārī script from 1593 to 1723. The syllabic portrait of each manuscript, established with precision, has produced an exploratory clustering, which highlights the changes in orthography. The clustering results tend to support the idea that there are at least 2 different lines in the manuscript transmission, and we can suggest that another variation of JY, not rooted in JY_B, existed in Nepal which gave rise to the JYMUṬ variant of JY.3.9. The fact that JY_B stands very far from the rest of the manuscripts both in time and by its syllabic portrait, also

[7] For the historical changes in Sanskrit orthography in a close comparison to Middle-Indic languages see Edgerton 1946:199, 202-203.

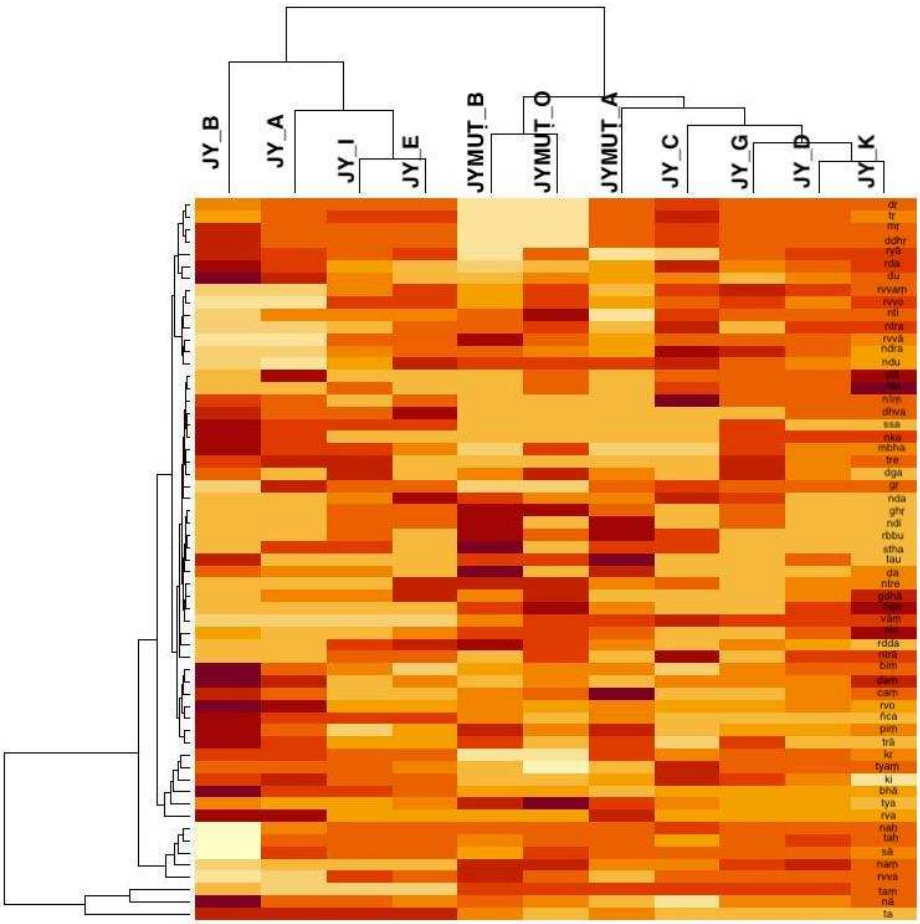

Fig. 8. Above heatmap is the visualisation of the clusters of the manuscripts by the closeness of their syllabic portrait. The "differentiating" syllables (listed on the right) of the 11 manuscripts are shown by the number of occurrences, where colour intensity indicates minimum (white) to maximum (dark read).

suggests that there existed more manuscripts in between, which, we hope, will surface in the future.

READ software paleography tools proved to be an important asset for this research, and its data model that maintaining the link between the manuscript surface, transcription, and various forms of interpretation, allows the scholar to see the higher levels of associations and to verify the coherence of the data. READ handles highly linked Big Data, which is visualised and made available in an interactive bite-size format for human scholars. This is a good example of what a VRE should be.

Besides clarifying the questions of the manuscript transmission, this IT-DH collaboration has also identified priorities for new features to develop in READ.

The next steps of the current study include the expanding to all chapters of JY using the current/ possibly modified methodology to extend the analysis of the transmission of JY. The analysis will be enriched by using data produced through other READ features (lexicography, syntax, compound analysis) with Computational Linguistics approaches.

References

1. Benoit K. [and al.]: "quanteda: An R package for the quantitative analysis of textual data." Journal of Open Source Software, **3**(30), 774 (2018). https://doi.org/10.21105/joss.00774 https://quanteda.io
2. Edgerton, F.: Meter, Phonology, and Orthography in Buddhist Hybrid Sanskrit. Journal of the American Oriental Society **66**(3), 197-206 (1946). https://www.jstor.org/stable/595566
3. Gupta, P. L.: Medieval Indian Alphabets: [Nagari, North India]. All Indian Educational Supply Company, Delhi [n.d., pre 2012]
4. [*Jayadrathayāmala*, *ṣaṭka* 3]: A NGMPP Microfilm A152/9; B NGMPP Microfilm B26/9; C NGMPP Microfilm A1312/25; D NGMPP Microfilm B122/3; E NGMPP Microfilm C72/1, G NGMPP Microfilm B121/13, I NGMPP Microfilm C47/03.
5. [*Jayadrathayāmala*, *ṣaṭka* 3]: K, https://digital.staatsbibliothek-berlin.de/werkansicht?PPN=PPN898781620&PHYSID=PHYS_0004. Last accessed 23 May 2021
6. [*Jayadrathayāmalamantroddhāraṭippaṇī*]: A NGMPP Microfilm B122/07; B NGMPP Microfilm A1267/3; O NGMPP Microfilm A152/8.
7. NGMPP (Nepal-German Manuscript Preservation Project) Catalogue, https://catalogue.ngmcp.uni-hamburg.de/content/search/ngmcpdocument.xed. Last accessed 22 May 2021