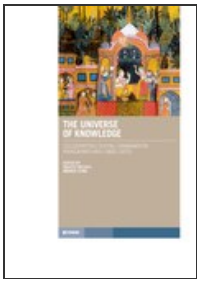
FORUM EDITRICE UNIVERSITARIA UDINESE



Giovanni Ciotti e Marco Franceschini

Experimenting with Digital Palaeography: The First Application of the Handwriting Software Tool (HAT 3.5) to Indian Scripts

Abstract: HAT (Handwriting Analysis Tool) is a software tool developed by Hussein Adnan Mohammed at the Centre for the Study of Manuscript Cultures (Universität Hamburg) that analyses and compares handwriting styles provided as input in the form of digital reproductions of handwritten texts. Thanks to some innovative features of the method on which it is based, HAT is easy to use autonomously, even by those with no particular digital expertise, and it is not tied to a specific script or a particular script typology. This paper presents the results of the first attempt at applying HAT to two Indian scripts, namely the Tamil and the Tamilian Grantha scripts, in the form in which they are attested in palm-leaf manuscripts from Tamil Nadu. Six tests were carried out, different in mode and objective from one another. The first four tests are aimed at verifying the ability of HAT to assess the similarity between writing styles and to use such an ability to establish or verify the identity of a particular scribe. The last two tests explore the possibility of exploiting HAT's potential for studying the diachronic development of the two scripts in question.
Keywords: Colophons, Palm-leaf manuscripts, Pattern analysis and recognition, Sanskrit, Tamil

Contenuto in: The Universe of Knowledge. Celebrating Shiyali Ramamrita Ranganathan (1892-1972)
Curatori: Fausto Freschi e Andrea Cuna
Editore: Forum
Luogo di pubblicazione: Udine
Anno di pubblicazione: 2024
Collana: Tracce. Itinerari di ricerca/Area umanistica e della formazione
ISBN: 978-88-3283-428-4
ISBN: 978-88-3283-486-4 (versione digitale/pdf)
Pagine: 159-192
DOI: 10.4424/978-88-3283-486-4-06

Per citare: Giovanni Ciotti e Marco Franceschini, «Experimenting with Digital Palaeography: The First Application of the Handwriting Software Tool (HAT 3.5) to Indian Scripts», in Fausto Freschi e Andrea Cuna (a cura di), *The Universe of Knowledge. Celebrating Shiyali Ramamrita Ranganathan (1892-1972)*, Udine, Forum, 2024, pp. 159-192
Url: https://forumeditrice.it/percorsi/lingua-e-letteratura/tracce/the-universe-of-knowledge/experimenting-with-digital-palaeography-the-first

FARE srl con socio unico Università di Udine
Forum Editrice Universitaria Udinese
via Larga, 38 - 33100 Udine
Tel. 0432 26001 / Fax 0432 296756 / forumeditrice.it

CC BY NC ND

Experimenting with Digital Palaeography: The First Application of the Handwriting Software Tool (HAT 3.5) to Indian scripts

Giovanni Ciotti and Marco Franceschini*

Abstract

HAT (Handwriting Analysis Tool) is a software tool developed by Hussein Adnan Mohammed at the Centre for the Study of Manuscript Cultures (Universität Hamburg) that analyses and compares handwriting styles provided as input in the form of digital reproductions of handwritten texts. Thanks to some innovative features of the method on which it is based, HAT is easy to use autonomously, even by those with no particular digital expertise, and it is not tied to a specific script or a particular script typology. This paper presents the results of the first attempt at applying HAT to two Indian scripts, namely the Tamil and the Tamilian Grantha scripts, in the form in which they are attested in palm-leaf manuscripts from Tamil Nadu. Six tests were carried out, different in mode and objective from one another. The first four tests are aimed at verifying the ability of HAT to assess the similarity between writing styles and to use such an ability to establish or verify the identity of a particular scribe. The last two tests explore the possibility of exploiting HAT's potential for studying the diachronic development of the two scripts in question.

1. Introduction

The authors of this article (henceforth "we") have been engaged for several years in a wide-ranging research that aims at collecting and studying the paratexts (in particular scribal colophons, ownership and loan notes) recorded in manuscripts written in Tamil and Grantha scripts. Our research does not end with the study of the paratexts themselves, but ideally aims to contribute to the reconstruction of the processes of manuscript production and circulation in Tamil Nadu between the early 17th and early 20th centuries. In the course of our work, we are often compelled to compare possibly similar writing styles that are found in different

* This article is the result of the joint and inseparable work of the two authors. By mere editorial convention, the scientific responsibility for the first half of the article (159-175) is attributed to Giovanni Ciotti, that of the second half (176-192) to Marco Franceschini.

manuscripts, in order to determine whether they were in fact written by the same scribe: the identification of the scribe can, in fact, provide essential information about the history of a manuscript, e.g. its temporal and/or spatial location. To date, there are no palaeographic reference studies on Tamil and Grantha scripts, and the comparison of writing styles can be an extremely difficult task, as anyone who has attempted it knows very well. Moreover, the conclusion is often undermined by serious doubts about its correctness and is more than often tantamount to an educated guess rather than a final verdict. Today, however, it is possible to make use of digital tools for analysing handwriting that can provide an effective help in this regard.

Over the last two decades, the contribution made by computer science to palaeographic and codicological studies has grown to such an extent that it has now become de facto indispensable. The use of digital images is now the norm for the use, storage, circulation, optimisation and study of manuscripts. Among the numerous tools that computer science has made available to researchers dealing with these issues, there is software that aims to identify the scribe who copied a given manuscript[1]: this article will deal precisely with one such software, namely the Handwriting Analysis Tool (HAT). This paper will present the results of the first attempt at applying HAT to an Indian script, the Tamil/Grantha script[2]. HAT was developed by Hussein Adnan Mohammed at the Centre for the Study of Manuscript Cultures (Universität Hamburg)[3]. Compared to similar digital handwriting analysis tools, HAT produces results of excellent reliability (Mohammed & al. 2018: 538-539). In addition, thanks to some innovative features

[1] For a historical overview of the different digital methods of recognising handwriting styles devised since the 1970s, see the sources labelled [1], [2] and [3] in Mohammed & al. 2018: 534-5, 539.

[2] Strictly speaking, Tamil and Grantha scripts are two separate scripts. Nevertheless, in this article they will be considered as one single script, the "Tamil/Grantha script": the reasons for this are explained below, "Notes on Tamil/Grantha script and manuscripts in Tamil Nadu".

[3] The HAT software (DOI 10.25592/uhhfdm.900) has reached version 3.5 (HAT-3.5), released in 2021 (https://www.fdr.uni-hamburg.de/record/9179). Since its first release in 2018, HAT is licensed under the Creative Commons Attibution-NonCommercial-NoDerivatives 4.0 International (Mohammed & al. 2018: 538). HAT-3.5 has been developed as a part of sub-project RFA05 "Pattern Recognition in 2D Data from Digitised Images and Advanced Aquisition Techniques". The research for this software was funded by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC 2176 "Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures" (https://www.fdr.uni-hamburg.de/record/9179). HAT is part of a set of six software tools collectively called "Pattern Analysis Software Tools (PAST)", all developed by H. Mohammed (Mohammed & al. 2022); they are all licensed under the Creative Commons Attribution Non Commercial No Derivatives 4.0 International and available here: https://www.csmc.uni-hamburg.de/publications/software.html.

of the method on which it is based[4], HAT offers itself to codicologists as a broad and easy-to-use tool: it is easy to use autonomously, even by those with no particular digital expertise, as it is based on a learning-free algorithm and makes use of an easy-to-use GUI (Graphical User Interface); it can be used offline; it is not tied to a specific script or a particular script typology[5], as the analysis method on which it is based is text-independent and segmentation-free (Mohammed & al. 2018: 534). The latter feature allows HAT to be used even if only a small number of samples of the writing styles to be analysed are available (Mohammed & al. 2017: 1014).

HAT is a software tool that analyses and compares handwriting styles provided as input in the form of digital reproductions of handwritten texts; the result is returned in the form of percentage values of relative similarity between the writings. More precisely, HAT compares a handwriting style, defined as "unknown style", with two or more "predefined" or "known" styles, which are provided as input subsequent to the first, and calculates the relative percentage similarity value of each of the predefined styles with respect to the unknown style[6]. Thus, when loading images of the different writing styles, the question asked to HAT is not "Which and how many of the known styles A, B [C, etc.] resemble the unknown style X?" but rather "To what extent does each of the known styles A, B [C, etc.] resemble unknown style X in relation to the other known styles?"; in other words, the question posed to the software is not polar, and the answer received is not polar either, but rather a relative index of similarity in percentage form, the value of which must be interpreted by the researcher who asked the question.

2. Notes on Tamil/Grantha Script and Manuscripts in Tamil Nadu

All texts used for the tests presented in this article are in Tamil and/or Grantha script and are written by engraving the text on folios made from palm leaves.

[4] For the method on which HAT is based, called Normalised Local NBNN (Normalised Local Naïve Bayes Nearest-Neighbour Classifier), see Mohammed & al. 2017.

[5] For example, HAT software was employed for the analysis of Arabic script by Mohammed & al. 2020 and, more recently, by El-Khatib 2023, as well as of the Ume (dbu med, "headless") script of Tibet by Mohammed & Helman-Wazny 2022.

[6] "HAT is a software tool that can be used to analyse handwriting styles. Multiple and different handwriting styles can be analysed concurrently and sorted according to their similarity to a questioned or unknown style. A similarity score can be calculated for each predefined style to create a relative comparison between them with respect to an unknown style. The handwriting styles (scribes) do not need to be known, only discriminative labels must be used for each different style (hand)" (https://www.csmc.uni-hamburg.de/publications/software/hat.html; accessed September 28, 2023).

The Tamil and Grantha scripts have been used for centuries in the region where the Tamil language is spoken, roughly coinciding with today's Indian state of Tamil Nadu and some neighbouring areas. Although they are very similar to each other[7], in principle the two scripts have clearly separate functions, since they are used to write the Tamil and Sanskrit languages, respectively. Nevertheless, they are frequently found used together in the same text: sometimes because a Sanskrit text is accompanied by a commentary in the Tamil language, more often because the text transmitted by the manuscript is composed in a hybrid register called *maṇipravālam*, which combines Sanskrit lexemes with morpho-syntactic elements of one of the Dravidian languages – in this case, Tamil. The co-presence of the two scripts is extremely frequent in the paratexts that are the subject of our research, particularly in the scribal colophons, which were composed by the scribes themselves (see below, § 3). It is, therefore, reasonable to assume that all Tamil-speaking scribes were able to write in both scripts[8].

For the above-mentioned reasons, no distinction will be made in this article between Tamil and Grantha scripts (henceforth "Tamil/Grantha script"): the texts used for the tests described below are written in either or both scripts.

Returning to the medium, Tamil Nadu manuscripts are traditionally written on palm leaves; the manuscripts on paper that exist today are invariably 19th-century or, much more often, 20th-century copies of apographs written on palm leaves. Once palm leaves were adequately prepared for hosting texts, writing was performed by engraving letters by means of a metal stylus; the text thus written was then blackened (inked) with lampblack mixed with vegetal oil. Returning to the medium, the leaves of two palms were mainly used in Tamil Nadu: Talipot (*Corypha umbraculifera*) and Palmyra (*Borassus flabellifer*). The folios obtained from the Talipot palm are larger, softer and more flexible than those obtained from the Palmyra and allow the scribe to engrave, with less effort, more minute (although less marked) characters, on more closely spaced lines; but Talipot leaves were rarer then Palmyra leaves and, for this reason, were used sparingly.

[7] The Tamil and Grantha scripts have a remarkable degree of similarity, but also have important differences: some graphemes represent sounds that are exclusive to one of the two languages and are therefore only used in one of the two scripts, while other graphemes, although representing the same phonemes, have a different graphic appearance in the two scripts. In addition, the Grantha script makes extensive use of consonantal ligatures (i.e., it tends to graphically join one or more consecutive consonants), which appear very rarely in the Tamil script.

[8] It should be noted that, while the Tamil script is still used today to write the Tamil language, handwriting in Grantha actually fell into disuse in the first decades of the 20th century, with the abandonment of the practice of manually copying texts onto palm-leaf manuscripts; however, the Grantha script is still used today in printed text publications.

For the tests presented in this article, manuscripts written on both media were used, although those on Palmyra constitute the vast majority of the total[9].

3. Methodological Observations

Before describing the tests and illustrating their results, a few methodological considerations are in order.

All the texts used for the tests presented in this article are closed by a scribal colophon that we studied and analysed in the research mentioned above. These paratexts were composed and appended at the end of the texts by the scribes themselves: although usually succinct, they can be of varying length and can contain various pieces of information, the most frequent of which are the date of the end of the copying and the name of the scribe. In this respect, the data already collected in our ongoing research on colophons proved very useful, since, as will be seen, the dating of the copying of the text and/or the name of the scribe who copied it is almost always indispensable information for deciding whether a text is appropriate for a test[10].

From a chronological point of view, our database contains approximately one thousand colophons distributed over a span of three centuries, from the beginning of the 17th century to the second decade of the 20th century[11]. The chronological distribution of the dated manuscripts in our database is, however, far from uniform: around 80% of the dated texts were copied in the 19th century. This imbalance is reflected in the texts used for the tests in this article, which are also chronologically placed, for the vast majority, in the 19th century (see Appendix B).

As far as the digital reproductions of the manuscripts is concerned, the quality of the images available to date is rather uneven. The method on which HAT is based (Normalised Local NBNN, see footnote 4) was designed to enable the software to obtain reliable results even where digital images of texts are "degraded". Providentially, the Normalised Local NBNN method has proven to be par-

[9] Although it is not always possible to distinguish the two media based on the mere analysis of digital images, the texts RE55825 and RE55827β, both copied by a scribe called Sundaravatiyar, are certainly written on Talipot.

[10] The only test for which this information is not indispensable is Test 1c.

[11] Manuscripts older than those we identified certainly did exist, and it cannot be ruled out that they still exist, albeit in extremely small numbers, surviving the hot and humid climate of southern India, the voracity of insects and mice, and neglect. Instead, we can assume that the abandonment of the practice of copying texts on palm-leaf manuscripts at the beginning of the 20th century can be attributed to the spread of movable type printing, which would have definitively supplanted the manual copying of texts.

ticularly effective in counteracting two common types of degradation found in digital reproductions of manuscripts, which also affect the images at our disposal: lack of contrast and low resolution (Mohammed & al. 2018: 536-538). The lack of contrast between the support and the text is a problem that we often have to deal with in the case of palm leaf manuscripts. It stems mainly from two factors: the dark colour of the support, which may have been such originally or may have darkened over time due to various factors (fumigation, dirt, etc.), and the attenuation of the intensity of the inking. For our tests, we tried to select images that were not affected by this problem[12]. As for resolution, as one might expect, the values are uniform for reproductions of manuscripts belonging to the same collection that were photographed within the scope of the same digitising initiative, but may vary considerably from collection to collection. In order to carry out the tests presented here, we drew on four collections: Bibliothèque nationale de France (BNF), École française d'Extrême-Orient, Centre de Pondichéry (EFEO), Institut français de Pondichéry (IFP) and U.V. Swāmināthaiyar Library, Chennai (UVSL). The four collections have considerably different resolution values, the BNF's being considerably higher, the EFEO's lower. In order to compare images that are qualitatively consistent with each other in terms of resolution (but also in terms of brightness, contrast and colour dominance), manuscripts belonging to only one collection were used in each test[13].

HAT compares an "unknown style" with two or more "predefined" or "known" handwriting styles and assigns each predefined style a percentage value indicating the relative degree of similarity of each to the unknown style. Each style, the unknown and the predefined styles, are represented by one or more digital reproductions of handwritten text. In the tests carried out for this article, each style is represented by ten digital images. In order to obtain samples as representative as possible of an entire text, images of folios positioned at different points in the manuscript were chosen, trying to select an equal number of images of the recto and verso of the folios[14]; for obvious reasons, we avoided

[12] It is not uncommon to come across uninked manuscripts in Tamil/Grantha script: for obvious reasons related to their problematic legibility (even more so in photographic reproductions), these manuscripts have not been considered in the present study.

[13] To be precise, what may adversely affect the test result is not the overall resolution of the image, but that of the script characters reproduced in the image. In this respect, there are considerable (albeit infrequent) differences even between digital reproductions of manuscripts belonging to the same collection: for example, the characters of the UVSL1044 have approximately twice the pixel size of those of the other manuscripts in the UVSL collection, whereas, in the IFP collection, the characters of RE10857β, RE10900α, RE43835β, RE45802, RE47718 have considerably smaller pixel sizes – and those of RE15386 considerably larger – than those of the other manuscripts.

[14] Both the appropriateness of using ten digital images representing each writing style and

choosing images with insufficient contrast as well as images reproducing severely damaged folios (broken or partially devoured by insects).

Before carrying out the actual tests, we carried out a few test runs, supplying HAT with digital images that reproduced the folios in their entirety: the observation of the position of the keypoints generated by HAT to analyse the writing styles[15] highlighted, however, that the software recognised as part of the text portions of the image of a different nature (i.e. non-textual), such as the edges of the folios and the holes drilled in it to pass the binding cord. Following this finding, in order to eliminate disturbing graphic elements and reduce noise, we cropped all the images, retaining only the part of the text delimited laterally by the two holes and above and below by the top and bottom edges of the folio[16].

4. The Tests: Preliminary Notes

In the following pages, we will present the six tests, different in mode and objective, that we carried out with HAT on manuscripts in Tamil/Grantha script.

Test 1 has the sole, but fundamental, purpose of verifying the correct functioning of HAT with texts engraved on palm leaves and written in Tamil/Grantha script. The test was therefore carried out on a sample of 14 texts for which we know the name of the scribe and the date of writing, asking HAT to recognise pairs of texts written by the same scribe.

Instead, in Tests 1a, 1b and 1c we submitted three case studies to HAT, i.e. real doubts that arose in the context of our study on paratexts: in these tests we asked HAT to corroborate (or, conversely, to exclude) the identification of two writing styles or to verify the identity of a scribe[17].

Finally, in Tests 2a and 2b, we employed HAT for investigations with different purposes than in the previous tests, shifting the focus from the identity of the scribes to the development of Tamil/Grantha script and its styles over time.

the opportunity to select the ten images from different parts of the manuscript are suggested by Mohammed & al. 2018: 535.

[15] Keypoints are the crucial points in writing that allow a style to be characterised and distinguished from other styles (see Mohammed & al. 2020: 81). For a technical discussion of the keypoints detection and feature extraction method in Normalised Local NBNN (and, thus, in HAT) see Mohammed & al. 2017: 1014 and 2018: 535-536.

[16] Cropping images in order to remove "irrelevant information" is recommended in the Remarks section of the HAT-2 User Manual (Mohammed 2018: 8).

[17] Mohammed & al. (2017: 1013) define and distinguish the processes of "writer identification" and "writer verification" as follows: "The task of writer identification can be defined as the process of assigning a writer with known reference handwriting samples to an unknown handwriting sample, while writer verification is the task of measuring the similarity between two samples of handwritings".

The names assigned to the texts to identify and distinguish them from each other were designed differently in the different tests, in order to highlight the most significant information for that specific test. For example, in Test 1, the purpose of which is the identification of scribes, the texts are identified by labels composed of the scribe's name followed by the year of writing (e.g. Atankan 1867). In contrast, in test 2b, the purpose of which is detecting diachronic changes in a scribe's style, texts are identified by the year of copying followed by its manuscript's shelfmark (e.g.: 1814 EO0036). A general concordance of these identification labels is offered in Appendix B.

4.1. Test 1: Recognition of Scribes' Writing Style

The first test we carried out aimed at ascertaining whether HAT is able to operate correctly on texts engraved on palm leaves and written in Tamil/Grantha script. In concrete terms, the aim of the test is to verify whether, by analysing and comparing digital reproductions of texts written by different scribes, HAT is able to produce numerical results that show a greater stylistic proximity between texts copied by the same scribe compared to those copied by different scribes. In the event of a positive result of the test, the quantitative aspect, i.e. the ratio of numerical similarity values assigned to texts by the same scribe compared to those assigned to texts by different scribes, will also assume considerable importance. Indeed, it must be remembered that HAT does not assertively assign a style to a particular scribe, but assigns a percentage value of similarity relative to the "predefined" or "known" styles as opposed to the "unknown" style, which serves as the subject of the comparison: it follows that the evaluation and interpretation of the test result is the task of the test taker[18].

The texts used for Test 1 were selected as follows. In our database, we selected the scribes of whom we possess at least two texts[19]; we considered only those texts that we are sure were written by a specific scribe, due to the presence in the colophon of a self-attribution formula, such as "[name of scribe] *svahastalikhitam*", "[name of scribe] *eḻuti mukiñcatu*" and the like. Furthermore, we only

[18] It is clear that the numerical aspect of the test result is of great importance: in future, on the basis of the results obtained from a sufficient number of tests, it will hopefully be possible to establish a numerical ratio between the values assigned to the two best-match writing styles (i.e., the two styles most similar to the one under examination) beyond which it is reasonably safe to assume that the "unknown style" and the "predefined style" that received the highest similarity score are the work of the same scribe.

[19] The codicological units used in the tests described in this article have been called "texts" (and not "manuscripts") because two or more texts can come from one and the same manuscript. For example, texts EO0076α and EO0076β were both taken from manuscript EO0076 (see Appendix B).

considered dated texts, i.e. those with colophons that have a date convertible to a specific day of the Gregorian calendar. Strictly speaking, the presence of the date would not have been necessary to carry out this test; however, since it is customary in Tamil Nadu to assign the same name to the male children of every other generation, the chronological proximity between the dates of a scribe's texts is a guarantee that they were actually written by the same person and not by two homonymous scribes – in this case, grandson and paternal grandfather.

Altogether, we selected seven pairs of texts, written by as many scribes: all of the selected texts were taken from manuscripts belonging to the collection of the Institut français de Pondichéry and were all written over a period of approximately one century, between 1809 and 1904. Using HAT, we then compared, in turn, the writing style of each of the 14 texts with the remaining 13, considering the former as the "unknown style", the others as the "predefined styles": the expected result was that, for each of the texts tested, HAT detected a greater similarity with the other text written by the same scribe than with the remaining 12 texts, written by other scribes.

The results of Test 1, shown in Table 1, allow us to confidently state that the test was largely successful. In twelve out of fourteen cases, HAT assigned the highest percentage of similarity to the text under examination (the "unknown style") to the other text by the same scribe, thus proving that it was able to recognise his writing style. The quantitative aspect of the test results is also very significant and extremely encouraging: in fact, the similarity value assigned to the other text of the scribe who copied the text under examination, which was always the highest in these twelve tests, is often well over twice as high as that assigned to the runner-up, i.e. to the text of another scribe that most closely resembles the text under examination[20]; in only four cases does it fall below the 2:1 ratio, but only slightly[21]. The test was negative only for the two texts written by a single scribe, Deyvasikhamanibhattar; in these tests, in fact, HAT assigned a mutual similarity value to the two texts of this scribe that was lower than that of three (in the case of the text Deyvasikhamanibhattar 1816) or even five (Deyvasikhamanibhattar 1831) values assigned to writing styles belonging to other scribes.

The fact that the only two negative test results concern the texts of a single scribe (out of seven) suggests that the problem lies either in a particular characteristic that distinguishes this scribe from the others, or in the state of preserva-

[20] Almost triple in the case of Muddaya 1809 (26.1:9.4=2.8) and Muddaya 1836 (25.9:9.5=2.7), even more than triple in the case of Atankan 1880 (29.4:8.7=3.4), Venkatacalan 1882 (37.1:11.6=3.2) and Venkatacalan 1883 (35.7:11.5=3.1).

[21] Atankan 1867 (18.6:10.6=1.8), Sundaravatiyar 1886 (18.8:10.9=1.7), Sundaravatiyar 1834 (19.5:12.2=1.6), Venkatanarayanasastri 1904 (27.1:14.4=1.9).

tion of the folios on which the two representative texts of this scribe are written. In fact, one could argue, for the results of these two tests, that the folios on which Deyvasikhamanibhattar wrote his texts contain tunnels dug by parasites; as evidenced by the analysis of the keypoints, HAT interprets these tunnels as signs belonging to the text – and, thus, as "traits" that help determine the scribe's writing style. It must be said, however, that the folios of other texts tested also show the same tunnels and, more generally, a similar state of preservation to those of the Deyvasikhamanibhattar texts (e.g. Venkatacalan 1882)[22]. It will be necessary to further investigate the reasons why HAT was unable to recognise the similarity between the two texts written by Deyvasikhamanibhattar, while it recognised the style of the other six scribes tested with such broad values.

Altogether, the result of Test 1 demonstrates that HAT is able to detect (almost always) the characteristics that make a scribe's style recognisable as opposed to those of other scribes. One might be tempted to infer information from the results of Test 1 that goes beyond its main purpose, in particular similarity relationships between the styles of scribes active in the same or contiguous periods; in this regard, caution is advised, as better described below (Test 2a and Test 2b), where the use of HAT for analysing the diachronic development of Tamil/Grantha script will be discussed.

Finally, the results concerning two scribes, Venkatacalan and Venkatanarayanasastri, seem noteworthy: not only are their styles considered to be very similar to each other, but the style of at least one of these two scribes turns out to be the most similar to that of all other scribes, with the sole exception of Muddaya (and one text by Atankan, namely Atankan 1880). These findings are to some extent surprising, but not immediately interpretable: further investigation will be needed to establish what caused them and whether a specific meaning can be attributed to them.

Unlike Test 1, whose sole objective was to verify the reliability of HAT with Tamil/Grantha script, the following three tests (1a, 1b and 1c) are inspired by real problems that arose in the course of our work on colophons in Tamil/Grantha manuscripts: therefore, on the one hand they aim at providing us with useful elements for solving such problems, and on the other hand they present possible practical applications of the ability shown by HAT in Test 1 to "recognise" the writing style of a scribe in opposition to that of other scribes.

[22] More generally, however, what may invalidate the test result is not the mere presence of noise elements (such as, for example, tunnels dug into the medium by parasites), but the ratio between the keypoints that HAT places on the actual script track and those it places on the noise elements. An examination of the arrangement of keypoints on the images suggests that this ratio, called the signal to noise ratio (S/N), is significantly high in the case of the texts copied by Deyvasikhamanibhattar.

4.2. Test 1a: Are Annacamiyuppattiyar and Annacami One and the Same Scribe?

The database of our work on colophons includes four texts, all belonging to the collection of the U.V. Swāmināthaiyar Library (UVSL) in Chennai, which were copied by a scribe named Annacamiyuppattiyar, son of "A." and inhabitant (or native) of the village of Tirumayilai[23]. From the same collection is a multiple text manuscript (UVSL67) containing four colophons in which the scribe laconically signs himself off as Annacami[24]. The similarity between the names of the two scribes suggests that this may be the same person, but the question remains uncertain and difficult to determine. Firstly, Annacami (or Annasvami) is a very common name in Tamil Nadu. Furthermore, upon visual comparison, Annacami's writing style seems similar to that of Annacamiyuppattiyar, but the similarities are not conclusive – nor are they uniform across the four texts written by Annacamiyuppattiyar. Indeed, even assuming that Annacami and Annacamiyuppattiyar are the same person, a certain degree of dissimilarity of style in the different texts would be natural, in light of the fact that these texts were copied at times even far apart: in fact, the text copied by Annacami is dated to the years 1837-1838[25], whereas one of those copied by Annacamiyuppattiyar (UVSL1) was written more than 40 years later, in 1880 – and it seems reasonable to assume that Annacamiyuppattiyar's writing style changed, at least to some extent, over such a long period of time[26].

We therefore resorted to HAT in order to acquire a further, more objective assessment of the possible identity of Annacami and Annacamiyuppattiyar. The test involved six texts, all belonging to the UVSL collection: the text copied from Annacami (UVSL67), two of the four texts copied by Annacamiyuppattiyar[27] and

[23] In the four colophons, the scribe gives us the following (scanty) information about his own identity: *tirumayilaiyil - ā - kumāraṉ - aṇṇācuvāmiy upāttiyāl - eḻuti niṟaiveriṉatu*, "[The text] was fully copied by Aṇṇācuvāmi Upātti son of Ā. in Tirumayilai" (UVSL1); *tirumayilai aṇṇācāmiyuppāttiyar eḻutiyatu*, "Copied by Aṇṇācāmippāttiyar of Tirumayilai" (UVSL5); *itai mayilai aṇṇācuvāmiy upāttiyāl*, "The teacher Aṇṇācuvāmiy from Mayilai copied this" (UVSL41c); *iḵtu mayilai aṇṇācāmiy upātti eḻutiyatu*, "This was copied by the teacher Aṇṇācāmiy from Mayilai" (UVSL231a).

[24] These are UVSL67β, UVSL67γ, UVSL67δ and UVSL67ζ.

[25] The colophon at the end of text UVSL67γ records two dates, 14 July 1837 and 7 December 1838, which presumably refer to the beginning and end of copying.

[26] Thus, if Annacami and Annacamiyuppattiyar are the same person, his period of activity would have spanned a period of at least 43 years. This would be the longest career recorded in our database, although other scribes were active for considerably longer periods: for instance, of another scribe, Venkatarya, we possess texts copied over a period of 34 years, from 1814 to 1848 (see below, Test 2b).

[27] Annacamiyuppattiyar UVSL1 (1880) and Annacamiyuppattiyar UVSL41c (date unknown). Annacamiyuppattiyar UVSL1 was chosen because it was copied in 1880, more than

a control group consisting of three texts[28]. The latter were selected from texts of which we know the name of the scribe (in order to rule out the possibility that they had been copied by Annasvami or Annacamiyuppattiyar) and the date of copying, so that the results would not be distorted by a (hypothetical) lack of homogeneity due to temporal distance. The test consisted of three trials: by turns, the text copied by Annacami and the two texts copied by Annacamiyuppattiyar were given as input to HAT as "unknown styles" and compared with the remaining texts.

The results of the test (Table 1a) leave no room for doubt: Annacamiyuppattiyar and Annacami are the same person. The writing style of one is always the most similar to that of the other, with the percentage values of mutual similarity always being very high (between 30% and 50%). But the fact that removes all doubt is that in the two tests in which the role of "unknown style" was assigned to Annacamiyuppattiyar's texts, HAT attributes an even higher similarity value to Annacami's text than to the other text copied by Annacamiyuppattiyar. In other words, HAT notes that the writing style of each of the two Annacamiyuppattiyar's texts more closely resembles Annacami's style than that of the other text by the same Annacamiyuppattiyar. On the basis of these data, the identification between the two scribes is incontrovertible.

4.3. Test 1b: Did Sundaravatiyar Write RE55844?

The activity of a scribe called Sundaravatiyar (also known as Sundaravaddhyar and, sometimes, Sundaram, in abbreviated form) is known to us thanks to several colophons distributed in three manuscripts in the collection of the Institut français de Pondichéry (RE55825, RE55827 and RE55853)[29], in which the scribe recorded, in addition to his own name, those of his father (Ramasvamivaddhyar), of his village of origin or residence (Marutvakuti) and, on six occasions, the dates of the end of copying, which are distributed between 1864 and 1886[30].

A fourth manuscript from the same collection (RE55844) contains two colophons (RE55844α and RE55844β) placed at the end of the two texts it transmits,

40 years after Annacami's text (1837-1838): given the large chronological gap separating the two texts, a high degree of mutual similarity would lend greater strength to the hypothesis of an identification between the two scribes.

[28] Arunacalakkaviraya UVSL1044 (copied in 1851), Mu. Kumarettu UVSL1080c (copied in 1873), Renkanata UVSL107 (copied in 1824).

[29] Texts from two of these three manuscripts were also used in Test 1: RE55827β (called "Sundaravatiyar 1864" in Test 1) and RE55853β ("Sundaravatiyar 1886" in Test 1).

[30] The scribe's name, as well as other elements (invocations and symbols) that play a central role in this test and that will be described below, recur in six colophons, distributed in the three manuscripts: RE55825, RE55827α, RE55827β, RE55853β, RE55853γ, RE55853θ.

in which the scribe did not record his own name[31]: the two colophons, however, contain two clues that indirectly suggest that the texts contained in this manuscript were also written by Sundaravatiyar. The first clue is an invocation to Airāvateśvarasvāmin[32] and his consort Abhirāmī, recorded in the first colophon (RE55844α): in our database, similar invocations are attested elsewhere only in the colophons of two texts written by Sundaravatiyar (RE55827β and RE55853β)[33]. The second clue is two symbols that appear in the second colophon (RE55844β), used in the date as markers to label the Jovian year and the solar month: in our database, these two symbols occur exclusively (and always in pairs) in six colophons of texts copied by Sundaravatiyar.

The question that arises, therefore, is: was the scribe who copied the two texts contained in manuscript RE55844, respectively closed by colophons RE55844α and RE55844β, written by the same scribe who copied the texts in manuscripts RE55825, RE55827 and RE55853, namely Sundaravatiyar?

On this question, HAT's ability to discern the writing styles of different scribes can provide us with an objective and very useful assessment. But, before proceeding with the test, some clarifications are necessary.

First, in the state in which they have come down to us, both texts copied by the anonymous scribe (RE55844α and RE55844β) are now composed of two sets of folios written and inserted into the manuscript at different times: this is evident from a visual analysis of the writing media. In both texts, two types of folios can be distinguished, which are different perhaps because of the botanical variety of the palm used to produce them, perhaps because of the treatment the folios received before being written upon, or perhaps made different by the frequency, intensity and duration of use: the first type is visually darker and shinier, probably stiffer, the second type lighter and softer[34]. We can assume that the first type of

[31] Each of the two colophons contains a date, but the calendar data recorded in them does not allow us to convert them to specific dates in the Gregorian calendar. More specifically, the two dates are "cyclical": in fact, they are based on the system of "Jovian years", a cycle consisting of 60 solar years (each year distinguished by a name) which, having come to an end, starts again from the first year of the series. It follows that the calendar elements recorded in the two dates under consideration (Jovian year, month and day) recur together cyclically once every sixty years.

[32] Airāvateśvarasvāmin is a form of Śiva worshipped in the famous Darasuram temple near Kumbakonam (Tamil Nadu).

[33] The invocations have a largely overlapping form in the three colophons: *abhirāmisameta-airāvateśvarasvāmisahā⟨y*m⟩* and *abhirāmmīsameta-aiai(sic!)ravateśvarasvāmisahāyaṃ*, "The support of Airavateśvarasvami together with Abhirami" (in RE55827β and RE55853β respectively, where Sundaravatiyar mentions his own name), *abhirāmīsameta-airāvateśvarasvāmine namaḥ*, "Honour to Airāvateśvarasvami together with Abhirami" (in RE55844α, where the scribe, whoever he is, does not record his own name).

[34] The two types of folios are also distinguished by the different distance between the two

folios is the one the entire manuscript was originally made of: on the one hand because they look more worn and tattered, and on the other hand because the text copied on these folios is complete with the notation of the Vedic accent, which is not found in the text copied on the second type of folios[35]. It can be surmised that some time after the manuscript was copied onto the folios of the first type, some of the folios of both texts deteriorated or were lost. A scribe (perhaps the same as the one who copied the text in the first draft, perhaps another scribe: see below) would then have rewritten the text to be replaced on new folios (those of the second type), without taking care to accompany the text with the Vedic accent markings.

For the sake of simplification, we have arbitrarily ascribed all the text written on the folios of the first type to a single hypothetical scribe called Scribe A and, similarly, all the text written (presumably later) on the second type of folios to a single hypothetical scribe called Scribe B; it is worth emphasising that the two types of folios (and, therefore, the two writing styles we have attributed to the hypothetical Scribe A and Scribe B) are present in both texts under examination (RE55844α and RE55844β)[36].

Visual examination of the writing styles suggests that only one scribe wrote on the first type of folios, both those in the first text and those in the second, as the style of writing on these folios, which is very elegant, appears considerably uniform: in other words, Scribe A could really be a single person. It is much more difficult to visually assess whether the same can be said of Scribe B, as well as whether Scribe A and Scribe B are the same person and, again, whether the scribe (or one of the scribes) of the two anonymous texts RE55844α and RE55844β can, in turn, be identified with Sundaravatiyar. Apparently, the styles of Scribe A and Scribe B show remarkable similarities, but any assessment based on visual analysis is inevitably influenced by the different style of engraving of the text (lighter that of Scribe A), the different intensity of inking (much more intense and

holes through which the binding cord passes, which is approximately 18.7 cm in the first type of folios, 16.2 in the second type. The outer dimensions of the folios are, however, approximately the same: 38.5-39.0x3.5-4 cm.

[35] The two texts that we call RE55844α and RE55844β are two successive sections of the *Taittirīyasaṃhitā*, TS.6.1.1-6.1.11.6 (RE55844α) and TS.6.2.1.1-6.1.3.5 (RE55844β). To be precise, the Vedic accent is annotated only in the folios of the first type found in the first text.

[36] In the first text (RE55844α), which occupies folios 1-14, the folios of the first type (written by Scribe A) are the first five, the ninth and the tenth; in the second text (RE55844β), which occupies folios 15-72, the folios of the first type (written by Scribe A) are those foliated with the numbers 46 and 49-72. The remaining folios are those of the second type, which we have conjecturally ascribed to Scribe B. Note that the colophons, which contain the invocations and symbols elsewhere recorded only in the colophons of the scribe Sundaravatiyar, were written respectively by Scribe B in the first text (RE55844α, folio 14 verso) and by Scribe A in the second text (RE55844β, folio 72 recto).

recent that of the text written by Scribe B) and the different number of lines of writing per folios (5 or 6 in the text written by Scribe A, 7 to 11 in that of Scribe B) and, consequently, the density of the writing.

Against this background, we thought we would "interrogate" HAT in order to acquire a further element of evaluation that would help us dispel these doubts – an element of evaluation that is precious precisely because it is free from the prejudicial influences mentioned above.

We collected 12 sets of images for the test. For the manuscript under examination (RE55844), we collected four sets of images, subdivided by type of folios (first type = Scribe A, second type = Scribe B) and by text (RE55844α and RE55844β): this was done in order to be able to compare the styles of Scribe A and Scribe B in the two different texts, so as to be able to verify whether the text we arbitrarily ascribed to them can be attributed to a single scribe. The writing styles of Scribe A and Scribe B in the two texts under examination were compared with three texts written by Sundaravatiyar[37], in order to reveal possible identities between these scribes, and with a control group consisting of the five texts whose writing styles were found, in Test 1, to be the most similar to Sundaravatiyar's[38]. In this way, any greater similarity of the texts under examination to the Sundaravatiyar style would gain in value, as it would emerge from a batch of similar styles of writing.

The results of the test (Table 1b) are partly expected, partly surprising, and shed new light on the relationships between the scribes under investigation. The first, important conclusion, deduced from the values in the first and third columns of the table, is that Scribe A is a single scribe and can be identified with Sundaravatiyar. In manuscript RE55844, the writing style on the oldest folios, those we have called "of the first type", is the same in the first and second text (RE55844α and RE55844β) and is identifiable with Sundaravatiyar's writing style. In other words, the two texts in manuscript RE55844 were originally written by Sundaravatiyar, as was suggested to us by the occurrence in the colophons of the invocation and symbols that, in our database, are attested exclusively in the colophons of other texts copied by this scribe.

In contrast, the folios "of the second type", i.e. those that, according to our hypothesis, were written as replacements for deteriorated or missing original folios, were written not by a single hypothetical Scribe B, but by two different scribes in the two texts. According to the results provided by HAT, the replacement folios in the second text (RE55844β) were written by Scribe A alias

[37] Sundaravatiyar 1864 (=RE55827β), Sundaravatiyar 1886 (=RE55853β), Sundaravatiyar 1886(bis) (=RE55825). The first two texts were also used in Test 1.

[38] Atankan 1867, Venkatacalan 1882, Venkatacalan 1883, Venkatanarayanasastri 1899, Venkatanarayanasastri 1904.

Sundaravatiyar (see the fourth column in Table 1b, where maximum similarity values were assigned to the styles of Scribe A and Sundaravatiyar); the replacement folios in the first text (RE55844α), on the other hand, were supposedly written by another scribe, who will remain anonymous (see the second column in Table 1b, where the three highest similarity values are assigned to scribes in the control group). These conclusions, to which we were led by the test results, imply a further interesting consideration: unlike the colophon of the second text (RE55844β), which would have been written by (Scribe A alias) Sundaravatiyar, the colophon of the first text (RE55844α), which is on a replacement folio, would have been written by the anonymous scribe (see footnote 36). To verify this attribution, we performed a further test with HAT, comparing the writing style of the single folio containing the colophon of the first text to the other texts: the result (see the fifth column of Table 1b) confirms that the style in which the colophon is written is different from that of Sundaravatiyar. The colophon of the first text contains the invocation for Airāvateśvarasvāmin and Abhirāmī, which, in the current state of our study of colophons, is employed exclusively by Sundaravatiyar: from this it seems inevitable to deduce that the anonymous scribe copied not only the text but also the colophon from the original folio. If so, it would be the first case known to us in a manuscript written in Tamil/Grantha script of a scribe copying the scribal colophon composed and originally written by another scribe[39]. But this is probably not the case.

In fact, it is possible to formulate a different, more plausible hypothesis. As mentioned above, each of the colophons of texts RE55844α and RE55844β contains a date: the two dates are cyclical, so they cannot be converted to a specific day in the Gregorian calendar (see footnote 31), but it is possible to establish that the minimum distance separating them is about 17 years. Such a wide temporal distance means that the two colophons cannot have been composed during the same copying process. In all likelihood, the colophon at the end of the second text (preserved on a folio "of the first type") was composed and written by Sundaravatiyar at the end of the original text[40]; on the other hand, the colophon at the end of the first text (folio 14) was *composed* and written by the anonymous scribe who rewrote (on folios "of the second type") folios 6-8 and 11-14 of that

[39] It is worth emphasising that scribal colophons mainly contain information not about the transmitted text, but about the manuscript itself, in particular the process of its production and circulation and the actors involved in these processes. In this case, the colophon of RE55844α contains the (cyclical) date of the end of copying and a series of invocations, starting with the one in honour of Airāvateśvarasvāmin and Abhirāmī.

[40] If this were the case, the cyclic date in the colophon of RE55844β would correspond to a day in December 1886 or January 1887, which is the chronologically closest date to the other dates recorded in Sundaravatiyar's colophons known to us.

text to replace deteriorated or lost original folios. In other words, at the end of the copying process, the scribe composed his own colophon, in which he inserted the actual date of the end of his work, 17 years later than Sundaravatiyar's original drafting of the text[41]. This hypothesis, however, implies that we cannot simply ascribe to Sundaravatiyar the exclusive use of the invocation to Airāvateśvarasvāmin and Abhirāmī, since another scribe made use of it. In response to this objection, one could assumed that the anonymous scribe belonged to Sundaravatiyar's family circle: based on the distance between the two dates, he could have been a son. Indeed, the invocation to Airāvateśvarasvāmin and Abhirāmī suggests a special devotion of Sundaravatiyar (and his family!) to this particular aspect of Śiva – and, perhaps, a special relationship with the temple of Darasuram, where these deities are worshipped (see footnote 32). Although lacking any supporting evidence, this hypothesis seems plausible: it should be remembered, in fact, that in India the professions, and that of scribe is no exception, are caste-related and are handed down from father to son. Moreover, this conjecture makes it possible to account for the fact that HAT seems to recognise a certain degree of similarity between the writing styles of the anonymous scribe and Sundaravatiyar (see in particular the fourth column of Table 1b), while remaining far from identifying them. Incidentally, in future, it would be worth using HAT to study the similarity between the writing styles of scribes belonging to the same family circle in greater depth and to ascertain whether this similarity could generate false identifications by the software.

4.4. Test 1c: Was BN-INDIEN 337 Written by One or Two Scribes?

The manuscript BN-INDIEN 337, belonging to the BNF collection, transmits the *Upatēcakāṇṭam*, the seventh book of the *Kantapurāṇam*. In the colophon, the copying of the text is first attributed to a scribe called Ampalavanavattiyar, then, later on, it is stated that the first 712 stanzas (of the 4333 that make up the text) were supposedly copied by another scribe, whose name is not recorded[42]. The

[41] Following this hypothesis, the date in the colophon of RE55844α would correspond to 17 May 1904.

[42] The text of the colophon is as follows: *⟨āka⟩c ceyyuḷ - 4333 ๛ tuṉmuk[i] ⟨Jovian year⟩* āvaṇi *⟨month⟩ 8 ⟨day⟩ cukkiṟavāramu makālaṭcuminoṉpum uttirāṭanaṭcattiramun tiṟaiyoteci piṟatoṭapuṇṇiyakālamuṅ kūṭiṉa cupatiṉattile cellamaṇipiḷḷaiyavarkaḷukku upatēcakāṇṭam eḻuti niṟaiveṟiyatu muṟṟum ๛ ivai eḻutiṉa naṉmaikku tiruveṇṇainallūr tillaiyappāvāttiyār kumārar* [[…]] *kūṭappākkam vāttiyār tampi ampalavāṇavāttiyār kai eḻuttu ๛ ivaiyil [u]ruttirākkam māṉmiya varaikkum ceyyuḷ 712 †?† nī†kki† ceyyuḷ 3621m ampalavāṇavāttiyār kai eḻuttu ๛* śrī *kokilāmpāḷ kirupai ๛ civamayam*. A possible translation is: "The illustrious Ciṟṟampalam. The essence of Śiva. Total stanzas 4333. On the auspicious day when the Jovian year Tuṉmuki, the month of Āvaṇi, the 8th day, Friday, the Makālaṭcumi festival, the

supposed point of caesura between the work of the two scribes has been precisely identified in the text, roughly in the middle of folio 44 verso ([44v4]): but the actual presence of a change of hand at that point is not visually evident and different scholars have expressed differing opinions on the matter. In order to ascertain whether the text is the work of a single scribe or whether, starting in the middle of the 44th folio, the copying was taken over by a second scribe, it was decided to acquire the impartial opinion of HAT. To this end, ten images were collected from the first 43 folios of the text and ten images from the following folios: the two groups were labelled by adding the letters [A] and [B] to the manuscript abbreviation, respectively. The control group consisted of four texts also belonging to the BNF collection, written by four different scribes, each text written entirely by a single scribe[43]. The four texts were halved: ten images were collected from the first half and ten images from the second half of each text, distinguishing the two groups as above, i.e. labelling them by adding the letters [A] and [B] to the text abbreviation. In total, therefore, we created 10 groups of images, five labelled with the letter [A] and five with the letter [B]. We then carried out ten tests, assigning the role of "unknown style" in turn to each group ([A] and [B]) and asking HAT to compare it with all the other groups ("predefined styles").

In the first eight tests, we assigned the role of "unknown style" to the groups [A] and [B] of the four control texts: the purpose of these tests is to verify that HAT assigns the highest similarity value to the other half of the text proposed as "unknown style", thus demonstrating that it "recognises" the style of that particular scribe in relation to (or rather: in opposition to) those of the other scribes.

Since all eight tests were positive (see the first eight columns of Table 1c), we submitted the two parts of the text under examination, BN-INDIEN 337[A] and BN-INDIEN 337[B], to HAT as "unknown style". In both tests (last two columns of Table 1c), HAT assigned the highest similarity value to the other part of the text under examination, thus supporting the hypothesis that the text in manuscript BNF Indien 337 was copied by a single scribe. It should be noted, however, that the gap between the highest similarity value (invariably assigned by HAT to the

constellation of Uttirāṭam, the thirteenth [lunar day], the auspicious time of the evening (*piṟatoṭa*°) come together, the *Upatēcakāṇṭam* was copied and completed for Mr Cellamaṇipiḷḷai. The handwriting of Ampalavāṇavāttiyār, son of Tillaiyappavāttiyār of Tiruveṇṇainallūr [and] younger brother of Kūṭappākkam-vāttiyār, is for the goodness (*naṉmaikku*) for which (sic!) these [stanzas] are copied (*eḻutiṉa*). Among these [stanzas] (*ivaiyil*), excluding (*nīkki*) the 712 stanzas up into (*varaikkum*) the *Uruttirākkam Māṉmiyam* (=the section of the *Upatēcakāṇṭam* called *Rudrākṣamahātmya*), 3621 stanzas are in the handwriting of Ampalavāṇavāttiyār. The compassion of the illustrious Kokilāmpāḷ. The essence of Śiva".

[43] These are BN-INDIEN 2 (copied from Cittampalavan), BN-INDIEN 184 (Cupapti), BN-INDIEN 318 (Cokkalinkan), BN-INDIEN 531 (Caminatan).

other half of the texts under examination) and the immediately lower value (attributed to a text written by another scribe), is very high for the control texts (first eight columns of Table 1c), but much smaller in the case of the text which is being investigated in this test, namely BN-INDIEN 337 (last two columns of Table 1c).

On the other hand, confirming the consistency of the results output by HAT, it can be seen that, in test pairs in which the two halves of the same text are examined, the order of similarity of the "predefined styles" tends to be stable. As further proof of the general reliability of the results provided by HAT, it may be noted that the similarity values assigned to the two halves of the same text are extremely close to each other in all tests. The deviation between the values assigned to the two halves of the same text is always less than 2%: the only interesting exception is the text subject of this test, BN-INDIEN 337, for which the deviation is 4.9% and 2.5% respectively in the two tests in which the role of "unknown style" was assigned to the two halves of the text BN-INDIEN 531 (fifth and sixth columns of Table 1c); incidentally, the style of the scribe of the latter text turns out to be the most similar to that of the scribe of BN-INDIEN 337 (last two columns of Table 1c). The "interference" between the styles of these two scribes will certainly merit further investigation.

4.5. Test 2a: Diachronic Evolution of the Tamil/Grantha Script

The tests presented so far were aimed at verifying the ability of HAT to assess the similarity between writing styles and to use this ability to establish or verify the identity of a particular scribe. The two tests that follow (2a and 2b), on the other hand, are aimed at ascertaining the possibility of exploiting HAT's potential for studying the diachronic development of the Tamil/Grantha script.

In particular, Test 2a represents an attempt to ascertain the possibility that HAT can detect the existence of a "period style" for the Tamil/Grantha script, i.e. that it can find particular similarities between texts written in the same years, even though they were written by different scribes and, in all likelihood, in different areas of Tamil Nadu.

To this end, we selected 12 texts whose copying dates were concentrated in three periods sufficiently distant from each other. We then created three groups, which we named "1780s", "1830s-1840s" and "1900s" respectively: each group contains digital reproductions of four texts whose copying ended in the years suggested by the name assigned to the group[44]. Finally, we submitted the images

[44] Each group therefore contains forty images, ten for each text. The group "1780s" contains digital reproductions of texts whose copying ended in 1781 (RE15386), 1782 (RE20088), 1787 (RE15535), 1789 (RE47718). The group "1830s-1840s" contains images of texts copied in 1836-1837 (RE15447), 1837 (RE436430), 1840 (RE10900α), 1842 (RE45802). Finally, the

of each text to HAT as "unknown style" and asked it to compare them with those of the three groups, obviously excluding the images of the text under examination from those of his own group. The expected result was that the writing style of the text under investigation would be more similar to that of the three texts of their own group, which were written in the same years in which it was written, than to that of the texts of the other two groups, which were written several decades before or after the text under investigation.

The results of the test are controversial, as can be seen in Table 2a. In contrast to four cases in which HAT assigned very high similarity values to the group to which the text under investigation belongs[45] and two cases in which the value of the group to which the text under investigation belongs barely prevails over the other two[46], there are six cases (out of 12!) in which HAT assigned the highest similarity value to texts copied in years farther apart than the manuscript under investigation[47]. A careful evaluation of the assumptions, method and results of this test suggests two reflections.

A first consideration concerns the narrowness of the statistical basis on which the test was carried out: in all likelihood, four different texts (i.e. four different writing styles) do not constitute a sufficiently large sampling to represent the "style of the period" (a period that, in our test, extends almost to a decade) and do not provide HAT with sufficient data to carry out the test correctly. At present, however, we do not have a larger number of texts dated in years close to each other and belonging to the same collection.

Secondly, a reflection on the very theoretical assumptions on which we have based the test is in order. Can we really imagine that the diachronic dimension of Tamil/Grantha script is scanned by a temporal succession of synchronic writing styles recognisable and distinguishable from one another? More generally: can one actually postulate the existence of a "period writing style"? To do so is to assume that particular features of the Tamil/Grantha script emerged in different periods that were able to spread and establish themselves over the entire area of use of the script (or over a large part of it), and that these "styles of the period" prevailed not only over the writing styles that had emerged in earlier periods, but also over the styles characteristic of a particular sub-area and those peculiar to different scribal families[48] – without forgetting the personal style of

group "1900s" contains texts copied in 1899 (RE09826), 1902 (RE04090β), 1904 (RE04137), 1905 (RE20202).

[45] These are 1787 RE15535, 1837 RE15447, 1899 RE098267 and 1904 RE04137.

[46] This refers to the results of 1781 RE15386 and 1840 RE10900α.

[47] See 1782 RE20088, 1789 RE47718, 1837 RE436430, 1842 RE45802, 1902 RE04090β and 1905 RE20202.

[48] Within the caste system in India, the profession of scribe was passed down from generation

each scribe. The answer to these questions may come, in the future, from HAT itself, once we have a statistically adequate base of writing styles representing different periods of use of the Tamil/Grantha script to submit to its algorithm for analysis.

4.6. Test 2b: Chronological Development of a Single Scribe's Style

The last test we present concerns, like the previous one, the diachronic aspect of Tamil/Grantha writing; but unlike the previous one, this one focuses on the writing style of a single scribe. The test aims to verify whether it is possible to use the HAT software to trace the evolution of a single scribe's writing style over the decades: more specifically, we wanted to check whether HAT detects a greater similarity between the chronologically closest writing styles than between the most distant ones. To this end, we submitted the writing styles of a fair number of texts copied by a scribe called Venkatarya to HAT. Venkatarya is the scribe of whom we have the largest number of works: in the course of our research on colophons, we collected 31 colophons "signed" by him, distributed across 18 manuscripts, all of which are preserved in the EFEO collection[49]. Of these 31 colophons, 21 are dated over a span of 35 years, from 1814 to 1848: such an abundance of dated texts and the breadth of the chronological window in which they are distributed made him an ideal candidate for this test. If the test were successful, it would be possible to use HAT to date, albeit approximately, texts written by Venkatarya but lacking a date; furthermore, it would eventually be possible to apply this method of chronological collocation to texts by other scribes of whom we possess a fair number of dated works.

We performed 18 tests on as many dated texts[50]: in each test, one text was provided as input to HAT as "unknown style", the remaining 17 as "predefined

to generation within the same family from father to son. It is therefore likely that, along with the art of writing, the father (more or less consciously) imprinted his son's writing style with his own, with the result of giving a "family air" to the styles of the scribes in his family.

[49] Unfortunately, such a large number of colophons do not provide us with an equivalent amount of information about his person and life. All we know of Venkatarya (aka Venkata and Venkatacarya) is that he was the son of Sriramgacarya and grandson of Venkatacarya, and that he belonged to the Mahāpūrṇa lineage, founded by the Śrīvaiṣṇava saint of the same name, who was Rāmānuja's teacher. It must be said, however, that Venkatarya is the only scribe of whom we know the date of death: from a note written by an anonymous scribe on a flyleaf, we know that Venkatarya, "a lake of erudition and first among the commentators of the *Rāmāyaṇa*", died on the 22nd of April 1849.

[50] These are: EO0007a (1827), EO0009b (1824), EO0014 (1846), EO0021 (1848), EO0036 (1814), EO0069γ (1829), EO0076α (1845), EO0076β (1845), EO0078γ (1841), EO0083β (1828), EO0111b (1832), EO0115a (1839), EO0119c (1829), EO0134aβ (1825), EO0138 (1838), EO0143 (1847), EO0152 (1835), EO0408 (1818).

style", and HAT returned to us as output the relative similarity value of each of the 17 "predefined styles" to the "unknown style". Soon, however, we realised that fragmenting the result into 17 separate percentage values, each referring to a single text, would not provide us with meaningful results for tracing Venkatarya's "stylistic parabola" through the decades. We therefore decided to proceed differently: in each of the 17 tests, we merged the 17 "predefined styles" into sets defined on the basis of the temporal distance from the "unknown style": ±5 years, ±10 years, ±15 years, ±20 years, ±25 years or more[51]. In this way, the comparison does not take place between two individual texts (the "unknown style" and each of the 17 "predefined styles"), but between a single text (the "unknown style") and a group consisting of texts that Venkatarya wrote in years close to each other and placed at approximately the same temporal distance from the date of the text entered as "unknown style". This procedure allowed us to receive more aggregated and, we believe, meaningful results from HAT.

The theoretical assumptions of this test seem less controversial than those of the previous test: it is, in fact, reasonable to assume that the style of a given scribe changes over the years and that these changes are detectable by the use of HAT. Yet again, the results of the test are not satisfactory (see Table 2b). The set of styles labelled as "±5 years", i.e. the group of writing styles chronologically closest to the one given to HAT as "unknown style", obtained the highest similarity value in only eight out of 18 tests[52]. In three further tests, HAT assigned the highest similarity value to the "±10 years" group of styles, while giving the "±5 years" group a slightly lower value[53]. Overall, therefore, the test can be said to be successful in eight out of 18 cases and moderately satisfactory in three others; but in the remaining seven cases[54], the results seem to disprove the existence of a progressive change in Venkatarya's writing style over the decades of his long career as a scribe – or, perhaps, HAT is unable to detect it.

[51] The sets are defined by progressive difference (or exclusion) ("B minus A", "C minus (A joined to B)" etc.): e.g. if the text under examination is dated 1830, the group of styles "±5 years" will be formed by the texts dated 1825 to 1835 (with the obvious exclusion of the text under examination), the group of styles "±10 years" will be formed by the texts dated 1820-1824 and 1836-1840 and so on.

[52] These are the test in which the role of "unknown style" was assigned to 1814 EO0036, 1825 EO0134aβ, 1832 EO0111b, 1838 EO0138, 1841 EO0078γ, 1845 EO0076α, 1845 EO0076β, 1848 EO0021.

[53] See the results for 1829 EO0119c, 1846 EO0014 and 1847 EO0143.

[54] See results for 1818 EO0408, 1824 EO0009b, 1827 EO0007a, 1828 EO0083β, 1829 EO0069γ, 1835 EO0152, 1839 EO0115a.

5. Conclusions

In this article, we presented the results of the first application of the HAT software (v3.5) to an Indian script, the Tamil/Grantha script. The results are, in general, comforting and encouraging.

The first test (Test 1) showed that HAT can identify the writing styles of different scribes in manuscripts written in the Tamil/Grantha script. The next five tests provided examples of the use of HAT for solving actual problems that arose in the course of our research on paratexts in palm leaf manuscripts from Tamil Nadu. Taken together, these five tests represent, at least in part, the wide variety of potential applications of HAT to palaeographic and codicological studies. The results obtained in these tests are also encouraging, particularly those of the tests whose aim was to identify and verify the identity of the scribe (Tests 1a, 1b, 1c). More controversial, on the other hand, are the results of the attempts to employ HAT for diachronic research, both concerning the development of Tamil/Grantha writing as a whole (Test 2a) and the evolution of a specific scribe's personal writing style (Test 2b). In all likelihood, the results of both tests were negatively influenced by the limited size of the statistical population on which they were conducted. In the case of Test 2a, it is also necessary to reflect on the correctness and soundness of the theoretical assumptions on which the test is based. Overall, we can certainly state that HAT is an effective and useful tool for palaeographic and codicological studies. For the future, it would be desirable and important to establish a threshold-value beyond which it is reasonable to consider successful the identification of the "unknown style" with one of the "predefined styles": this value could be represented, for example, by the ratio between the values of the best-match and the second best-match. However calculated, it seems evident that it will only be possible to define this threshold-value once a sufficiently large number of *verifiable* results have been collected and analysed. Mohammed & al. (2018: 539) close their article by writing: "analysing and understanding the method behaviour using real-life data is more important than gaining improvement on synthesised or randomly chosen test data. This is particularly true for practical implementations of computational methods". For our part, we hope that this article of ours will contribute to the development and improvement of HAT.

Acknowledgements
The sections of this article authored by Giovanni Ciotti were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2176 "Under-standing Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures", project no. 390893796. The research was conducted within the scope of the Centre for the Study of Manuscript Cultures (CSMC) at Universität Hamburg, in particular the Palm-Leaf Manuscript Profiling Initiative (PLMPI).

Appendix A. Tables of results

Table 1. Test Results 1. In the table, the texts are identified by the name of the scribe and, in the line immediately below, by the year of copying (all in bold), for example: Atankan 1867, Atankan 1880.

Atankan		**Deyvasikhamanibhattar**		**Muddaya**		**Sundaravatiyar**
1867	**1880**	**1816**	**1831**	**1809**	**1836**	**1864**
Atankan 18,6 1880	Atankan 29,4 1867	Venkata-calan 12,4 1882	Venkata-calan 13,9 1883	Muddaya 26,1 1836	Muddaya 25,9 1809	Sundara-vatiyar 21,9 1886
Venkata-calan 10,6 1882	Sundara-vatiyar 8,7 1886	Venkata-calan 11,6 1883	Tyaka-racar 12,5 1840	Deyva-sikhamani-bhattar 9,4 1816	Deyva-sikhamani-bhattar 9,5 1816	Venkata-calan 9,3 1882
Venkata-calan 9,6 1883	Sundara-vatiyar 8 1864	Venkata-narayana-sastri 10,3 1904	Venkata-calan 12,5 1882	Atankan 9 1880	Atankan 8 1867	Venkata-calan 8,7 1883
Venkata-narayana-sastri 9 1899	Muddaya 7,8 1836	Deyva-sikhamani-bhattar 10 1831	Tyaka-racar 12,3 1834	Sundara-vatiyar 8 1864	Atankan 7,8 1880	Atankan 7,7 1867
Tyaka-racar 8,8 1834	Deyva-sikhamani-bhattar 7,7 1816	Venkata-narayana-sastri 8,4 1899	Venkata-narayana-sastri 10,6 1904	Sundara-vatiyar 7,5 1886	Sundara-vatiyar 7,3 1864	Venkata-narayana-sastri 7 1899
Venkata-narayana-sastri 8,2 1904	Muddaya 7,7 1809	Atankan 7,4 1867	Deyva-sikhamani-bhattar 8,8 1816	Atankan 6,8 1867	Sundara-vatiyar 7,3 1886	Deyva-sikhamani-bhattar 6,7 1816
Sundara-vatiyar 8,2 1886	Venkata-calan 6,2 1882	Tyaka-racar 7,4 1834	Sundara-vatiyar 6,6 1864	Venkata-calan 6,7 1882	Venkata-calan 6,9 1882	Venkata-narayana-sastri 1904 6,6
Deyva-sikhamani-bhattar 6,8 1816	Venkata-narayana-sastri 5,7 1899	Sundara-vatiyar 7,1 1864	Venkata-narayana-sastri 6,4 1899	Venkata-narayana-sastri 5,3 1899	Venkata-narayana-sastri 6,8 1899	Deyva-sikhamani-bhattar 6,2 1831
Sundara-vatiyar 6,5 1864	Venkata-calan 5,3 1883	Sundara-vatiyar 6,9 1886	Sundara-vatiyar 5,2 1886	Deyva-sikhamani-bhattar 1831 5,3	Venkata-calan 5,4 1883	Atankan 6 1880
Muddaya 4,6 1836	Venkata-narayana-sastri 1904 4,6	Muddaya 5,4 1836	Atankan 3,7 1867	Venkata-calan 5,2 1883	Venkata-narayana-sastri 1904 5	Tyaka-racar 6 1834
Deyva-sikhamani-bhattar 1831 4,2	Tyaka-racar 3,5 1834	Atankan 4,9 1880	Muddaya 2,7 1836	Venkata-narayana-sastri 1904 4,7	Tyaka-racar 4 1834	Muddaya 5,4 1836
Muddaya 3,1 1809	Deyva-sikhamani-bhattar 1831 3,5	Muddaya 4,6 1809	Muddaya 2,4 1809	Tyaka-racar 3,5 1834	Deyva-sikhamani-bhattar 1831 4	Tyaka-racar 4,4 1840
Tyaka-racar 1,9 1840	Tyaka-racar 1,8 1840	Tyaka-racar 3,7 1840	Atankan 2,4 1880	Tyaka-racar 2,4 1840	Tyaka-racar 2 1840	Muddaya 4 1809

	Tyakaracar		**Venkatacalan**		**Venkatanarayanasastri**	
1886	**1834**	**1840**	**1882**	**1883**	**1899**	**1904**
Sundara-vatiyar 1864 18,8	Tyaka-racar 1840 19,5	Tyaka-racar 1834 32,7	Venkata-calan 1883 37,1	Venkata-calan 1882 35,7	Venkata-narayana-sastri 1904 30,6	Venkata-narayana-sastri 1899 27,1
Venkata-narayana-sastri 1899 10,9	Venkata-narayana-sastri 1904 12,2	Deyva-sikhamani-bhattar 1831 16	Venkata-narayana-sastri 1904 11,6	Venkata-narayana-sastri 1904 11,5	Venkata-calan 1882 12,7	Venkata-calan 1883 14,4
Venkata-narayana-sastri 1904 9,8	Venkata-calan 1883 11,1	Venkata-calan 1883 10,2	Venkata-narayana-sastri 1899 9,1	Venkata-narayana-sastri 1899 8,7	Venkata-calan 1883 12,2	Venkata-calan 1882 14,2
Atankan 1867 9,5	Venkata-calan 1882 10,5	Venkata-narayana-sastri 1904 9,9	Deyva-sikhamani-bhattar 1831 6,7	Deyva-sikhamani-bhattar 1831 7,8	Sundara-vatiyar 1886 7,2	Tyaka-racar 1834 9,5
Venkata-calan 1882 9,4	Deyva-sikhamani-bhattar 1831 9,9	Venkata-calan 1882 9,1	Tyaka-racar 1834 6,7	Tyaka-racar 1834 6,5	Atankan 1867 6,4	Deyva-sikhamani-bhattar 1831 6,9
Venkata-calan 1883 9,3	Atankan 1867 8,2	Sundara-vatiyar 1864 4,9	Deyva-sikhamani-bhattar 1816 6,1	Deyva-sikhamani-bhattar 1816 6,1	Deyva-sikhamani-bhattar 1816 6,1	Deyva-sikhamani-bhattar 1816 6,1
Deyva-sikhamani-bhattar 1816 6,1	Venkata-narayana-sastri 1899 8,1	Venkata-narayana-sastri 1899 4,4	Sundara-vatiyar 1886 4,7	Sundara-vatiyar 1886 5,1	Tyaka-racar 1834 6,1	Sundara-vatiyar 1886 5,4
Tyaka-racar 1834 5,5	Deyva-sikhamani-bhattar 1816 6,5	Deyva-sikhamani-bhattar 1816 4	Sundara-vatiyar 1864 4,6	Sundara-vatiyar 1864 4,7	Deyva-sikhamani-bhattar 1831 4,8	Atankan 1867 4,5
Atankan 1880 5,5	Sundara-vatiyar 1864 4,8	Sundara-vatiyar 1886 3,1	Atankan 1867 4,6	Atankan 1867 4,4	Sundara-vatiyar 1864 4,6	Tyaka-racar 1840 4,2
Deyva-sikhamani-bhattar 1831 4,7	Sundara-vatiyar 1886 4,5	Atankan 1867 1,9	Tyaka-racar 1840 3,7	Tyaka-racar 1840 4,4	Muddaya 1836 2,8	Sundara-vatiyar 1864 3,7
Muddaya 1836 4,6	Muddaya 1836 2	Muddaya 1836 1,5	Muddaya 1836 2,1	Muddaya 1836 1,9	Tyaka-racar 1840 2,5	Muddaya 1836 1,7
Muddaya 1809 3,2	Atankan 1880 1,4	Muddaya 1809 1,2	Atankan 1880 1,6	Atankan 1880 1,8	Atankan 1880 2,3	Atankan 1880 1,3
Tyaka-racar 1840 2,8	Muddaya 1809 1,3	Atankan 1880 1	Muddaya 1809 1,4	Muddaya 1809 1,4	Muddaya 1809 1,7	Muddaya 1809 1,1

Table 2. Results of Test 1a. In the table, texts are identified by the name of the scribe followed by the shelfmark of the manuscript in which they are found, for example: Renkanatan UVSL107.

Annacami UVSL67 (1837-1838)	**%**	**Annacamiyuppattiyar UVSL1 (1880)**	**%**	**Annacamiyuppattiyar UVSL41c (no date)**	**%**
Annacamiyuppattiyar UVSL1	43,7	Annacami UVSL67	49,9	Annacami UVSL67	40,7
Annacamiyuppattiyar UVSL41c	29,4	Annacamiyuppattiyar UVSL41c	29,5	Annacamiyuppattiyar UVSL1	35,9
Mu. Kumarettu UVSL1080c	22,4	Mu. Kumarettu UVSL1080c	14,3	Mu. Kumarettu UVSL1080c	16,7
Renkanatan UVSL107	3,2	Renkanatan UVSL107	5,4	Renkanatan UVSL107	5,3
Arunacalakkavirayan UVSL1044	1,4	Arunacalakkavirayan UVSL1044	0,9	Arunacalakkavirayan UVSL1044	1,4

Table 3. Results of Test 1b. In the table, texts are identified by the name of the scribe followed by the year of copying (e.g. Atankan 1867), with the exception of the texts found in the manuscript under examination, RE55844.

RE55844α Scribe A	**%**	**RE55844α Scribe B**	**%**	**RE55844β Scribe A**	**%**	**RE55844β Scribe B**	**%**	**RE55844α Scribe B (only colophon)**	**%**
RE55844β Scribe A	20,2	Venkata-calan 1883	12,9	Sundara-vatiyar 1886	19,5	RE55844β Scribe A	11,8	Venkata-calan 1882	14,9
Sundara-vatiyar 1886	18,1	Venkata-calan 1882	12,5	RE55844α scribe A	19,1	Sundara-vatiyar 1886(bis)	11,5	Venkata-calan 1883	14,6
Sundara-vatiyar 1886(bis)	15,5	Venkata-narayana-sastri 1904	11,6	Sundara-vatiyar 1886(bis)	16,2	RE55844α scribe B	10,4	Venkata-narayana-sastri 1904	11,4
Sundara-vatiyar 1864	10,5	RE55844β Scribe B	9,9	RE55844β Scribe B	8,9	RE55844α scribe A	9,9	Venkata-narayana-sastri 1899	10,1
RE55844β Scribe B	8,2	Venkata-narayana-sastri 1899	9,4	Sundara-vatiyar 1864	7,9	Sundara-vatiyar 1886	9,8	Sundara-vatiyar 1886(bis)	8,7
RE55844α scribe B	5,7	Sundara-vatiyar 1886(bis)	9,1	RE55844α scribe B	5,5	Sundara-vatiyar 1864	8,8	RE55844β Scribe B	7,5
Venkata-calan 1882	5,5	Atankan 1867	8,3	Venkata-narayana-sastri 1899	5,2	Atankan 1867	8,1	RE55844β Scribe A	7,5
Venkata-calan 1883	5,2	RE55844β Scribe A	7,7	Venkata-calan 1882	4,8	Venkata-narayana-sastri 1899	7,7	Sundara-vatiyar 1886	6,7
Venkata-narayana-sastri 1904	4,2	RE55844α scribe A	6,9	Venkata-calan 1883	4,4	Venkata-calan 1882	7,5	Atankan 1867	6,5
Venkata-narayana-sastri 1899	4	Sundara-vatiyar 1886	6,1	Venkata-narayana-sastri 1904	4,3	Venkata-calan 1883	7,4	RE55844α scribe A	6,4
Atankan 1867	2,8	Sundara-vatiyar 1864	5,5	Atankan 1867	4,2	Venkata-narayana-sastri 1904	7,2	Sundara-vatiyar 1864	5,7

Table 4. Results of Test 1c. In the table, texts are identified by the shelfmark of the manuscript in which they are found followed by [A], "first half of the text", or [B], "second half of the text", for example: BN-INDIEN 318[A], BN-INDIEN 318[B].

BN-INDIEN 2[A]	**%**	**BN-INDIEN 2[B]**	**%**	**BN-INDIEN 184[A]**	**%**	**BN-INDIEN 184[B]**	**%**	**BN-INDIEN 318[A]**	**%**
BN-INDIEN 2[B]	33,4	BN-INDIEN 2[A]	34,5	BN-INDIEN 184[B]	42,4	BN-INDIEN 184[A]	43,1	BN-INDIEN 318[B]	35,2
BN-INDIEN 531[A]	13	BN-INDIEN 531[A]	11,9	BN-INDIEN 531[B]	11,2	BN-INDIEN 531[B]	11,4	BN-INDIEN 184[A]	9,7
BN-INDIEN 531[B]	12	BN-INDIEN 531[B]	11,4	BN-INDIEN 531[A]	10,1	BN-INDIEN 531[A]	9,8	BN-INDIEN 531[B]	9,3
BN-INDIEN 318[B]	8,7	BN-INDIEN 318[B]	8,7	BN-INDIEN 337[B]	7,8	BN-INDIEN 337[B]	7,6	BN-INDIEN 184[B]	8,8
BN-INDIEN 184[A]	8,3	BN-INDIEN 184[A]	8,6	BN-INDIEN 337[A]	7,2	BN-INDIEN 337[A]	7,1	BN-INDIEN 2[B]	8,1
BN-INDIEN 318[A]	8,2	BN-INDIEN 318[A]	8,5	BN-INDIEN 318[A]	5,8	BN-INDIEN 318[A]	5,8	BN-INDIEN 531[A]	8,1
BN-INDIEN 184[B]	7,4	BN-INDIEN 184[B]	7,9	BN-INDIEN 2[A]	5,5	BN-INDIEN 318[B]	5,6	BN-INDIEN 2[A]	7,9
BN-INDIEN 337[B]	5,5	BN-INDIEN 337[B]	5,2	BN-INDIEN 2[B]	5,2	BN-INDIEN 2[A]	4,8	BN-INDIEN 337[A]	6,4
BN-INDIEN 337[A]	3,4	BN-INDIEN 337[A]	3,5	BN-INDIEN 318[B]	4,8	BN-INDIEN 2[B]	4,8	BN-INDIEN 337[B]	6,4

BN-INDIEN 318[B]	**%**	**BN-INDIEN 531[A]**	**%**	**BN-INDIEN 531[B]**	**%**	**BN-INDIEN 337[A]**	**%**	**BN-INDIEN 337[B]**	**%**
BN-INDIEN 318[A]	32,6	BN-INDIEN 531[B]	34,1	BN-INDIEN 531[A]	34,8	BN-INDIEN 337[B]	27,6	BN-INDIEN 337[A]	25,3
BN-INDIEN 531[B]	10,1	BN-INDIEN 337[B]	14	BN-INDIEN 337[B]	11,7	BN-INDIEN 531[B]	16,7	BN-INDIEN 531[A]	20,8
BN-INDIEN 531[A]	9,2	BN-INDIEN 184[A]	10,4	BN-INDIEN 184[A]	10,7	BN-INDIEN 531[A]	15,9	BN-INDIEN 531[B]	17,3
BN-INDIEN 2[A]	9	BN-INDIEN 337[A]	9,1	BN-INDIEN 184[B]	9,8	BN-INDIEN 184[A]	11,3	BN-INDIEN 184[A]	10,4
BN-INDIEN 184[A]	8,8	BN-INDIEN 184[B]	9,1	BN-INDIEN 337[A]	9,2	BN-INDIEN 184[B]	10	BN-INDIEN 184[B]	9,3
BN-INDIEN 2[B]	8,5	BN-INDIEN 2[A]	7,2	BN-INDIEN 2[A]	6,7	BN-INDIEN 318[B]	5,9	BN-INDIEN 318[B]	5
BN-INDIEN 184[B]	8,4	BN-INDIEN 2[B]	6,2	BN-INDIEN 2[B]	6,1	BN-INDIEN 318[A]	5,8	BN-INDIEN 318[A]	4,6
BN-INDIEN 337[B]	7	BN-INDIEN 318[B]	5,1	BN-INDIEN 318[B]	5,6	BN-INDIEN 2[A]	3,5	BN-INDIEN 2[A]	4
BN-INDIEN 337[A]	6,5	BN-INDIEN 318[A]	4,8	BN-INDIEN 318[A]	5,3	BN-INDIEN 2[B]	3,3	BN-INDIEN 2[B]	3,5

Table 5. Results of Test 2a. In the table, texts are identified by the year of copying followed by the shelfmark, for example: 1781 RE15386. The shelfmarks “1780s”, “1830s-1840s” and “1900s” classify groups of texts: for their meaning, see above, “Test 2a: Diachronic evolution of the Tamil/Grantha script”.

1781		**1782**		**1787**		**1789**		**1837**		**1837**	
RE15386	**%**	**RE20088**	**%**	**RE15535**	**%**	**RE47718**	**%**	**RE15447**	**%**	**RE436430**	**%**
1780s	38,2	1900s	38,3	1780s	42,8	1830s-1840s	43	1780s	43,2	1780s	43,2
1900s	35,9	1780s	34,4	1830s-1840s	29,6	1780s	30,3	1900s	32	1900s	33,2
1830s-1840s	25,9	1830s-1840s	27,4	1900s	27,6	1900s	26,7	1830s-1840s	24,9	1830s-1840s	26,2

Table 6. Results of Test 2b. In the table, texts are identified by the year of copying followed by the shelfmark, for example: 1814 EO0036. The shelfmarks “±5”, “±10” etc. classify groups of texts: for their meaning, see above “Test 2b: Chronological development of a single scribe’s style”.

1814		**1818**		**1824**		**1825**		**1827**		**1828**		**1829**		**1829**		**1832**	
EO0036	**%**	**EO0408**	**%**	**EO0009b**	**%**	**EO0134aβ**	**%**	**EO0007a**	**%**	**EO0083β**	**%**	**EO0069γ**	**%**	**EO0119c**	**%**	**EO0111b**	**%**
±5	36,5	±20	30,5	±20	42,4	±5	32,3	±15	32,2	±15	33,6	±15	30	±10	27,8	±5	29,5
±20	22,4	±25 e oltre	23,7	±5	25,3	±20	19,7	±10	26,9	±10	27,8	±5	28,1	±5	25,1	±15	27,3
±15	16,9	±15	18,8	±15	13,8	±15	18,4	±5	22,1	±20	19,9	±10	25,1	±20	23,6	±20	22
±25 e oltre	12,5	±5	15	±10	10	±25 e oltre	14,9	±20	18,8	±5	18,7	±20	16,8	±15	23,4	±10	21,3
±10	11,7	±10	12	±25 e oltre	8,5	±10	14,8										

1840		1842		1899		1902		1904		1905	
RE10900α	**%**	**RE45802**	**%**	**RE098267**	**%**	**RE04090β**	**%**	**RE04137**	**%**	**RE20202**	**%**
1830s-1840s	36,9	1780s	39,2	1900s	56,5	1780s	45,8	1900s	56,4	1830s-1840s	35,8
1780s	36	1830s-1840s	32,1	1780s	26,2	1900s	33,5	1780s	24,9	1780s	32,8
1900s	27,1	1900s	28,8	1830s-1840s	17,3	1830s-1840s	20,7	1830s-1840s	18,7	1900s	31,3

1835		1838		1839		1841		1845		1845		1846		1847		1848	
EO0152	**%**	**EO0138**	**%**	**EO0115a**	**%**	**EO0078γ**	**%**	**EO0076α**	**%**	**EO0076β**	**%**	**EO0014**	**%**	**EO0143**	**%**	**EO0021**	**%**
±25 e oltre	27	±5	30,4	±25 e oltre	37,9	±5	23,1	±5	34,6	±5	42	±10	28	±10	32,3	±5	42,9
±5	24,4	±15	20,4	±10	22,3	±15	22,9	±15	25,4	±10	29,5	±5	27,6	±5	24	±10	22,1
±20	20	±25 e oltre	20	±5	21,8	±20	21,2	±10	20,7	±25 e oltre	16	±15	20,8	±20	18,1	±25 e oltre	12
±15	15	±10	16,9	±15	18	±10	20,6	±25 e oltre	11,8	±20	12,5	±20	15,7	±15	15	±20	11,8
±10	13,6	±20	12,2			±25 e oltre	12,2	±20	7,5			±25 e oltre	8	±25 e oltre	10,6	±15	11,2

Appendix B. Concordance of texts used in tests

In the following list, the texts used in the tests presented in this article are grouped according to the collection to which they belong. For each text is given, in order: the shelfmark, the name of the scribe (if known), the date of copying (if known), the tests in which it was used.

Collection of the Bibliothèque nationale de France (BNF)
- BN-INDIEN 2. Scribe: Cittampalavan. Date: 1707. Test 1c.
- BN-INDIEN 184. Scribe: Cupapti. Date: 1798. Test 1c.
- BN-INDIEN 318. Scribe: Cokkalinkan. Date: 1700. Test 1c.
- BN-INDIEN 337. Scribe: Ampalavanavattiyar. Date: 1896. Test 1c.
- BN-INDIEN 531. Scribe: Caminatan. Date: 1800. Test 1c.

Collection of the École française d'Extrême-Orient, Centre de Pondichéry (EFEO)
- EO0007a. Scribe: Venkatarya. Date: 1827. Test 2b.
- EO0009b. Scribe: Venkatarya. Date: 1824. Test 2b.
- EO0014. Scribe: Venkatarya. Date: 1846. Test 2b.
- EO0021. Scribe: Venkatarya. Date: 1848. Test 2b.
- EO0036. Scribe: Venkatarya. Date: 1814. Test 2b.
- EO0069γ. Scribe: Venkatarya. Date: 1829. Test 2b.
- EO0076α. Scribe: Venkatarya. Date: 1845. Test 2b.
- EO0076β. Scribe: Venkatarya. Date: 1845. Test 2b.
- EO0078γ. Scribe: Venkatarya. Date: 1841. Test 2b.
- EO0083β. Scribe: Venkatarya. Date: 1828. Test 2b.
- EO0111b. Scribe: Venkatarya. Date: 1832. Test 2b.
- EO0115a. Scribe: Venkatarya. Date: 1839. Test 2b.
- EO0119c. Scribe: Venkatarya. Date: 1829. Test 2b.
- EO0134aβ. Scribe: Venkatarya. Date: 1825. Test 2b.
- EO0138. Scribe: Venkatarya. Date: 1838. Test 2b.
- EO0143. Scribe: Venkatarya. Date: 1847. Test 2b.
- EO0152. Scribe: Venkatarya. Date: 1835. Test 2b.
- EO0408. Scribe: Venkatarya. Date: 1818. Test 2b.

Collection of the Institut français de Pondichéry (IFP)
- RE04090β. Scribe: Mahadevar. Date: 1902. Test 2a.
- RE04137. Scribe: Venkatanarayanasastri. Date: 1904. Tests 1, 1b, 2a.
- RE09826. Scribe: Venkatanarayanasastri. Date: 1899. Tests 1, 1b, 2a.
- RE10835γ. Scribe: Atankan. Date: 1880. Test 1.
- RE10838. Scribe: Atankan. Date: 1867. Tests 1, 1b.
- RE10857β. Scribe: Tyakaracar. Date: 1834. Test 1.
- RE10900α. Scribe: Tyakaracar. Date: 1840. Tests 1, 2a.
- RE15386. Scribe: Subrahmanyakuru. Date: 1781. Test 2a.
- RE15438. Scribe: Muddaya. Date: 1809. Test 1.
- RE15447. Scribe: Muddayya. Date: 1836-1837. Tests 1, 2a.

- RE15535. Scribe: Kurumurtti. Date: 1787. Test 2a.
- RE20047β. Scribe: Deyvasikhamanibhattar. Date: 1816. Test 1.
- RE20051. Scribe: Deyvasikhamanibhattar. Date: 1831. Test 1.
- RE20088. Scribe: Sritanunathanayaka. Date: 1782. Test 2a.
- RE20202. Scribe: Sankararamasiva. Date: 1905. Test 2a.
- RE43643θ. Scribe: Palucuvami Ayyar. Date: 1837. Test 2a.
- RE43833γ. Scribe: Venkatacalan. Date: 1883. Tests 1, 1b.
- RE43835β. Scribe: Venkatacalan. Date: 1882. Tests 1, 1b.
- RE45802. Scribe: Cuppiramaniyan. Date: 1842. Test 2a.
- RE47718. Scribe: Cuppiramaniyan. Date: 1789. Test 2a.
- RE55825. Scribe: Sundaravatiyar. Date: 1886. Test 1b.
- RE55827β. Scribe: Sundaravatiyar. Date: 1864. Test 1, 1b.
- RE55844α. Scribe: unknown. Date: unknown. Test 1b.
- RE55844β. Scribe: unknown. Date: unknown. Test 1b.
- RE55853β. Scribe: Sundaravatiyar. Date: 1886. Tests 1, 1b.

Collection of the U.V. Swāminātha iyar Library, Chennai (UVSL)

- UVSL1. Scribe: Annacamiyuppattiyar. Date: 1880. Test 1a.
- UVSL41c. Scribe: Annacamiyuppattiyar. Date: unknown. Test 1a.
- UVSL67. Scribe: Annacami. Date: 1837-1838. Test 1a.
- UVSL107. Scribe: Renkanatan. Date: 1824. Test 1a.
- UVSL1044. Scribe: Arunacalakkavirayan. Date: 1851. Test 1a.
- UVSL1080c. Scribe: Mu. Kumarettu. Date: 1873. Test 1a.

References

El-Khatib, A. (2023). *Editing mss Leiden Or. 14.545 b-c. The Oldest Qur'ān Fragments in Leiden University Library*, "Journal of Islamic Manuscripts", 14, 235-278.

Mohammed, H. (2018). *Handwriting Analysis Tool v2.0 (HAT-2). User Manual*, https://www.fdr.uni-hamburg.de/record/901.

Mohammed, H. & al. (2017). *Normalised Local Naïve Bayes Nearest-Neighbour Classifier for Offline Writer Identification*, in *2017 14th IAPR International Conference on Document Analysis and Recognition* (ICDAR, Kyoto, Japan, 9-15 November 2017), Piscataway (NJ), IEEE, 1013-1018.

Mohammed, H. & al. (2018). *Writer Identification for Historical Manuscripts: Analysis and Optimisation of a Classifier as an Easy-to-Use Tool for Scholars from the Humanities*, in [*Proceedings of the*] *16th International Conference on Frontiers in Handwriting Recognition* (ICFHR, Niagara Falls, New York, 5-8 August 2018), Piscataway (NJ), IEEE, 534-539, https://doi.org/10.1109/ICFHR44711.2018.

Mohammed, H. & al. (2020). *A Comparison of Arabic Handwriting-Style Analysis Using Conventional and Computational Methods*, "Manuscript Cultures", 14, 77-86.

Mohammed, H. & al. (2022). *Pattern Analysis Software Tools (PAST) for Written Artefacts*, in *Document Analysis Systems* (15th IAPR International Workshop, DAS 2022, La Rochelle, France, May 22-25, 2022. Proceedings), ed. by S. Uchida & al., Cham (Switzerland), Springer, 214-229.

Mohammed, H. & Helman-Wazny, A. (2022). *Understanding the Zhangzhung Nyengyu tsakali Collection using Computational Pattern Analysis*, in *CHAI 2022. Humanities-Centred Artificial Intelligence 2022* (Proceedings of the Workshop co-located with 45th German Conference on Artificial Intelligence, September 19 - September 23, 2022: Trier, Germany), ed. by S. Melzer & al., https://ceur-ws.org/Vol-3301/, 12-19.