A chronometric approach to Indian alchemical literature

Oliver Hellwig
Institut für Sprachen und Kulturen Südasiens, Freie Universität, Berlin

Abstract

Indian alchemy, a branch of traditional Indian medicine (Āyurveda), has produced a corpus of texts that are difficult to date using regular philological techniques. This article describes a contents-based computational method that is capable of calculating the relative chronology of these texts. Central parts of alchemical literature are encoded in a language model that can be understood by a computer and then compared with an alignment algorithm. Phylogenetic trees derived from these alignments show regularities in the ordering of alchemical texts, and these may be interpreted as temporal patterns. Processing these patterns with a minimization algorithm, we are able to compute a relative chronology of the corpus, which is largely consistent with results obtained using traditional philological techniques.

Correspondence:
Oliver Hellwig
Institut für Sprachen und Kulturen Südasiens,
Freie Universität,
Koenigin-Luise-Strasse 34a,
D-14195 Berlin, Germany.
E-mail:
hellwig7@gmx.de

Despite 200 years of philological research, the chronology of classical Indian literature is still full of uncertainties.[1] A particularly problematic area is a large bulk of metrical texts that can roughly be assigned to a ‘medieval’ period of literary production (500–1500 AD).[2] These texts come from fields such as religious, narrative or, as in our case, scientific literature. Nevertheless, they have two important traits in common: (1) most of these texts share a religious and sociological background that reflects mature Hinduism and (2) the texts are written in easy, unpolished Sanskrit, which had already become a ‘dead language’ only used by literati. Dating texts from this medieval period is complicated by several factors. Most importantly, these texts are (in most cases) not dated or attributable to (non-legendary) authors. Although Indian writers were highly interested in the tradition (*saṃpradāya*), in which a text was handed down, they did not care about the absolute chronology of these traditions (Gerow, 1977, 217–219). Therefore, dating Sanskrit texts may be seen as a genuinely Western approach to Indian literature, which is not supported by the inherent structure of this literary tradition. In addition, most manuscripts are only a few hundred years old, while the texts may, for instance, have been composed in the seventh or eighth century. Therefore, dated manuscripts can only serve as an approximation for the *terminus ante quem* of a text. Finally, texts are often compiled from different sources, and this makes them a mixture of heterogeneous elements that are difficult to separate.

If we want to date texts from this class of literature using chronometric methods, two restrictions apply. First, we do not have reliably dated texts to compare an undated text with. The same holds true for strategies such as authorship attribution.

Literary and Linguistic Computing, Vol. 24, No. 4, 2009.

doi:10.1093/llc/fqn043 Advance Access published on 8 January 2009

Since the concept of (non-legendary) authorship is almost absent from this class of texts, disputed works cannot be assigned to the œuvre of a known writer. As datable reference points are almost totally absent, chronometric studies of medieval Indian literature must focus on heterogeneity, or developments discernible at different levels of a text. The second restriction concerns the accuracy of datings found with chronometric methods. In many cases, traditional philological datings of a single Sanskrit text differ by hundreds of years. The case of ŚĀRṂGADHARASAṂHITĀ (sketched as in section 4) is a good, but not exceptional example of these controversies in Indian philology. Therefore, if it is possible to establish a relative chronology of the alchemical corpus using a chronometric method, this would be a considerable advance when compared to philological research, which is often based on limited sets of features and subjective conclusions drawn from these features.

The chronometric method proposed in this article differs from other computational approaches in two respects. First, the method is based on the hypothesis that the *content* of a well-defined group of texts changes with time in a similar way to the linguistic features of those texts. Differences in content are therefore interpreted as differences in time. Second, many studies use tests such as χ^2 (Trautmann, 1971) or (non-parametric) MANOVA (Tse *et al.*, 1998) clustering strategies [Hoover, 2002], or data reduction techniques, such as principal component analysis (Burrows, 1992) to detect homogeneity in texts. Apart from the fact that results using these techniques are often quite hard to interpret,[3] these methods are not suited to detect heterogeneity in alchemical text-content for two practical reasons. First, any kind of multivariate method needs input vectors of the same (high) dimension. As will become apparent later, such feature vectors cannot be extracted from the encoded content of alchemical texts. Second, most of these methods rely on rather strict statistical requirements, such as a multivariate normal distribution of features. Since contents analysis of Indian alchemical procedures is usually performed on a very small number of samples, these requirements are not fulfilled in most cases. Therefore, we propose a combination of bioinformatic algorithms, simple non-parametric tests, and minimization strategies in order to detect heterogeneity in alchemical texts.

In the first step, certain groups of alchemical instructions, the so-called 'basic procedures' found in almost any text of the alchemical corpus, are encoded with a simple language model. This results in a database of computer-understandable alchemical procedures, which currently contains over 1.200 procedures. If we want to be able to detect changes in content, the content has to be compared. This is the main issue of the next step. Encoded procedures of one thematic group are compared with an algorithm originally developed for the alignment of genetic sequences. As a by-product of this alignment, trees are generated that describe the temporal distribution of a group of alchemical instructions under certain mathematical conditions. From the layout of these trees, time rules are derived, which store regular patterns in the temporal distribution of texts. In a final step, statistically significant time rules are processed with a minimization algorithm that calculates the relative chronology of the alchemical corpus.

The rest of the article is organized as follows: Section 1 introduces the language model that encodes the alchemical texts. The following section describes how encoded alchemical procedures are compared with an alignment algorithm. In Section 3, we discuss the derivation of 'time rules' from alignments. Finally, in Section 4, we present a minimization algorithm, which transforms time rules into a stratification of the alchemical corpus.

1 Encoding Alchemical Texts

Indian alchemy arises from a group of about twenty important works written in regular, easy Sanskrit verses.[4] It flourished from 900 AD to 1600 AD, although alchemical treatises continued to be produced until the beginning of the twentieth century.[5] The Indian alchemical tradition has two main objectives: (1) the production of precious metals, especially of gold, and (2) the perfection of the human body. In order to achieve these aims, alchemical texts list a large number of experiments or *procedures*, which may be divided into basic and advanced procedures. The advanced procedures deal with the two main objectives of Indian alchemy, the transformation of base metals and of the fragile human body. Since many texts focus either on metallurgy or on body optimization, advanced procedures are quite unequally distributed over the alchemical literature. However, almost any alchemical text mentions the experiments which may be termed as ‘basic procedures’. They describe fundamental operations performed with mineral substances used in Indian alchemy, e.g. the purification (*śodhana*) and ‘oxidation’ (*mārana*) of these minerals, or the extraction of their ‘essences’ (*sattvapātana*). Minerals prepared in this way can later be used in the advanced procedures, however, as mentioned above, this topic is not covered equally by all alchemical texts. If we want to deduce age from changes in content, the first way to do this would be to compare basic procedures.

Basic procedures can be distinguished through two criteria: (1) the processed mineral (e.g. mercury) and (2) the type of procedure itself (e.g. purification). Grouping the procedures with these criteria, we get about ninety *thematic groups* labelled, for instance, ‘purification of mercury’ or ‘oxidation of copper’. Each procedure is then manually encoded in a simple time-linear language model, which has been developed for this project.[6] This model is derived from the verb-valence-model of linguistics, and each procedure is interpreted as a sequence of *verbal phrases*, i.e. verbs which are specified by their necessary valences (mainly actor, object, and instrument) and modifiers (mainly place). Concepts used in this language model are stored in a hierarchical ontology that combines the ideas of semantic nets (Quillian, 1986) and frames (Minsky, 1975). The following line from the Bhāvaprākāśa, a late text containing alchemical passages is used as an example (BhPr, 2, 3, 62):

> *dhrtvā tadgolakaṃ bhāṇḍe śarāveṇa ca rodhayet//* ‘(The alchemist) puts this ball (made of different ingredients) in a vessel and closes it with a *śarāva* (a kind of lid).’

$$\cdots \longrightarrow \begin{bmatrix} \textsc{Verb} & \text{put} \\ \textsc{Actor} & [\text{*alchemist}] \\ \textsc{Obj.} & [\text{ball}] \\ \textsc{Place} & [\text{vessel}] \end{bmatrix} \longrightarrow \begin{bmatrix} \textsc{Verb} & \text{close} \\ \textsc{Actor} & [\text{*alchemist}] \\ \textsc{Object} & [\text{vessel}] \\ \textsc{Ins.} & [\text{śarāva}] \end{bmatrix} \longrightarrow \cdots$$

Fig. 1 Encoding the line BhPr, 2, 3, 62.2. Entities marked with ‘asterick’ are supplemented from the context

The content of the line is split into two instructions (putting the ball in the vessel; closing this vessel with the lid) and then encoded as shown in Fig. 1.

2 Comparing Alchemical Procedures

To detect changes in one thematic group, we have to compare the procedures that the group contains. Sequence alignment, a technique from bioinformatics, is an adequate tool for such a comparison. Sequences of verbal phrases can be interpreted as symbol sequences. The aim is to align procedures, or the symbol sequences of one thematic group in such a way that as many similar or identical symbols (i.e. verbal phrases) as possible can be written ‘one under the other’, and as few gaps as possible have to be inserted. An example may clarify this method. Figure 2 shows three procedures describing the purification of sulphur.[7] The procedures are built from a small set of symbols consisting of the verbal phrases ‘melt ...’ = A, ‘filter ...’ = B, and ‘pour ...’ = C. Therefore, they can be rewritten in the form

$$\begin{array}{llll} p_1 & A & B & C \\ p_2 & A & C & \\ p_3 & B & C & \end{array}$$

and aligned as

$$\begin{array}{llll} p_1 & A & B & C \\ p_2 & A & - & C \\ p_3 & - & B & C \end{array}$$

This kind of alignment can, for example, be performed with the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970), which is based on the same idea as the well-known Levenshtein algorithm for string comparison. To align sequences of verbal phrases which are more complicated than

$$p_1 \quad \begin{bmatrix} \textsc{Verb} & \text{melt (tr.)} \\ \textsc{Actor} & \text{alchemist} \\ \textsc{Obj.} & \begin{bmatrix} \text{[sulphur]} \\ \text{[ghee]} \end{bmatrix} \\ \textsc{Place} & \begin{bmatrix} \text{[fireplace]} \\ \text{[vessel]} \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \textsc{Verb} & \text{filter} \\ \textsc{Actor} & \text{alchemist} \\ \textsc{Object} & \text{[sulphur]} \\ \textsc{Ins.} & \text{[piece of cloth]} \end{bmatrix} \longrightarrow \begin{bmatrix} \textsc{Verb} & \text{pour} \\ \textsc{Actor} & \text{alchemist} \\ \textsc{Object} & \text{[sulphur]} \\ \textsc{Place} & \text{[milk]} \end{bmatrix}$$

$$p_2 \quad \begin{bmatrix} \textsc{Verb} & \text{melt (tr.)} \\ \textsc{Actor} & \text{alchemist} \\ \textsc{Obj.} & \begin{bmatrix} \text{[sulphur]} \\ \text{[ghee]} \end{bmatrix} \\ \textsc{Place} & \begin{bmatrix} \text{[fireplace]} \\ \text{[vessel]} \end{bmatrix} \end{bmatrix} \longrightarrow \begin{bmatrix} \textsc{Verb} & \text{pour} \\ \textsc{Actor} & \text{alchemist} \\ \textsc{Object} & \text{[sulphur]} \\ \textsc{Place} & \text{[milk]} \end{bmatrix}$$

$$p_3 \quad \begin{bmatrix} \textsc{Verb} & \text{filter} \\ \textsc{Actor} & \text{alchemist} \\ \textsc{Object} & \text{[sulphur]} \\ \textsc{Ins.} & \text{[piece of cloth]} \end{bmatrix} \longrightarrow \begin{bmatrix} \textsc{Verb} & \text{pour} \\ \textsc{Actor} & \text{alchemist} \\ \textsc{Object} & \text{[sulphur]} \\ \textsc{Place} & \text{[milk]} \end{bmatrix}$$

Fig. 2 Three procedures for purification of sulphur ('ghee' is a kind of butter)

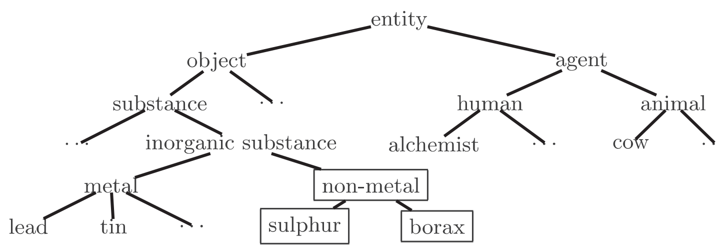


Fig. 3 Extract from a possible hierarchical ontology for alchemical texts

in this example, the alignment algorithm has to be modified in one important aspect. The source algorithm operates with a very small alphabet of symbols, and penalties for all symbol pairs can be pre-computed (Needleman and Wunsch, 1970, p. 447), whereas the alphabet of alchemical verbal phrases is unlimited. Calculating with pre-computed penalties unduly simplifies the comparison of similar verbal phrases. However, if we use distance information derived from the structure of the hierarchical ontology, similar verbal phrases can be compared at the time of running the program. Considering an example with two sample phrases 'he mixes sulphur and mercury' and 'he mixes borax and mercury', the concepts 'sulphur' and 'borax' are both children of the parent concept 'inorganic substance' → 'non-metal', and therefore have a very small 'ontological distance' (see Fig. 3). Generally speaking, in the modified Needleman–Wunsch algorithm, two symbols, or verbal phrases can be compared and also combined, if at least their verbal components are identical. In this way, the rather strict penalty function of the source algorithm is replaced by a variable similarity value ranging continuously from 0 (no similarities) to 1 (identical).

3 Deriving Time Rules

When more than two procedures are compared, the exact alignment algorithm soon becomes too computationally expensive. An easy heuristic solution for this problem first groups sequences into clusters, and then derives the order of alignment from this clustering (Thompson *et al.*, 1994). More specifically, alignments of all possible pairs of procedures are first calculated with the exact algorithm described in the last section. The pairwise distances resulting from the modified Needleman–Wunsch alignment are stored in a symmetric distance matrix, from which a hierarchical tree is computed. This 'guide tree' describes the order in which to perform the actual alignment. Contrary to the approach in Thompson *et al.* (1994), the guide tree is constructed with the UPGMA (Sokal and Michener, 1958), and not the Neighbour-Joining method. This results in a rooted tree that supplies the source information for the dating of alchemical texts. If the matrix of pairwise distances, from which this tree is constructed, is ultrametric,[8] the tree can be interpreted as a phylogenetic tree of one thematic group. The arrangement of elements in such a tree indicates their temporal relations. Elements branching off this tree at nodes that are closer to its root are earlier than those branching off later. From a biological point of view, UPGMA is motivated by the partly obsolete theory that successful mutations in a population occur with a constant rate ('molecular clock property'). If alchemical procedures also 'mutate' with an approximately constant rate, the clustering tree constructed with UPGMA may be used to date procedures or the texts that contain these procedures. Assuming a constant rate of mutation, of course, grossly simplifies the complicated process of text production and transmission. Scientific development may proceed in a non-linear way, thereby, invalidating the basic assumption of this method. In addition, the texts of Indian alchemy are interconnected by a dense network of citations and (from a European perspective) plagiarism. Two texts describing similar alchemical procedures need not be contemporary, but may cite or even, as Bendixen has put it in the case of RRS and RCūM (Bendixen, 1990), re-edit each other. Traditional philology, however, has based the chronology of Indian alchemy on the same foundation as this computational method: when the alchemical tradition is examined as a whole, differences of content become apparent, which can be interpreted as time markers. In this system, early alchemy deals with metallurgy, while late texts combine alchemical and āyurvedic teachings. The dates of texts showing features from both classes can only be estimated (cf. Ray, 1956, p. 98, White, 1996, p. 78–170). Therefore, the computational method proposed in this article may be seen as a reinterpretation of a typical philological technique, which, however, tries to replace the fuzziness of personal impressions by a (perhaps too) strict formalism.

When this theoretical model is applied to real alchemical data, two important restrictions have to be made. The first and most serious restriction is the fact that phylogenetical methods in biology compare the genetic material of contemporary taxa. This condition is also expressed in the claim that distances from the root to each leaf must be identical in a phylogenetic tree. Contrary to this claim, it cannot be guaranteed that each alchemical text is a 'contemporary taxon'. Some of the editions may preserve the state of a text from hundreds of years ago, while others may offer a recent version. Second, distance matrices for a group of procedures are very often far from ultrametric. Trees constructed from such matrices cannot be expected to represent the temporal relationship of the procedures/texts correctly. A practical way of working around this problem is based on the observation that in many cases procedures collected in one thematic group (such as 'purification of mercury') are obviously alternatives, and are therefore not 'phylogenetically' related to each other. One of the numerous examples for this claim is the purification of copper, for which two sub-groups, G_1 and G_2, can be distinguished (Fig. 4). In the first sub-group G_1, copper is smeared with plant extracts and salts, heated and dipped into plant juices. In G_2, the metal is cooked in cow's urine. The claim that sub-groups are alternatives and therefore not derived from a common ancestor is supported by the fact that many texts mention procedures which belong to several subgroups (in the case of copper, procedures from both sub-groups G_1 and G_2 are, for instance, listed in RRS and RCINT). A look at the guide tree in Fig. 4 shows that both sub-groups G_1 and G_2 are concentrated in well-definable sub-trees. The sub-matrices of pairwise distances corresponding to these sub-trees are in most cases at least approximately ultrametric.[10] Therefore, temporal relationships deduced from coherent sub-trees can be expected to contain less errors than those deduced from the complete tree of a thematic group.

To obtain an impression of how procedures are ordered in sub-groups, Fig. 5 shows the alignment of G_2 in a structural diagram. Items printed in bold are referenced in all procedures of G_2, otherwise the

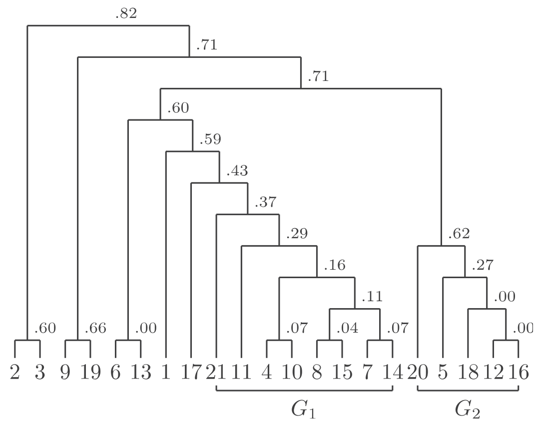


Fig. 4 Guide tree for clustering the thematic group 'purification of copper'—References: **1:** BhPr, 2, 3, 55–56; **2:**RArṇ, 12, 138–139; **3:** RArṇ, 17, 95–97; **4:** RCint, 6, 10; **5:** RCint, 6, 11; **6:** RCūM, 14, 45–46; 7: RCūM, 14, 46–47; **8:** RCūM, 14, 47–48; **9:** RCūM, 14, 48–56; **10:** RHT, 9, 13; **11:** RMañj, 5, 26–27; **12:** RMañj, 5, 28; **13:** RRS, 5, 49; **14:** RRS, 5, 50; **15:** RRS, 5, 51; **16:** RRS, 5, 52; **17:** RRĀ, R.kh., 8, 47–49; **18:** RRĀ, R.kh., 8, 50; **19:** RRĀ, V.kh., 3, 106–107; **20:** RSK, 2, 18; and **21:** RRS, 1, 270–271

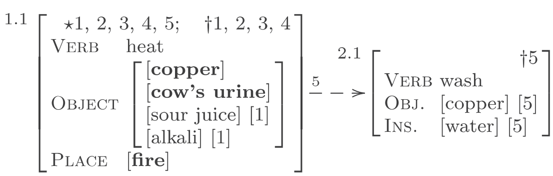
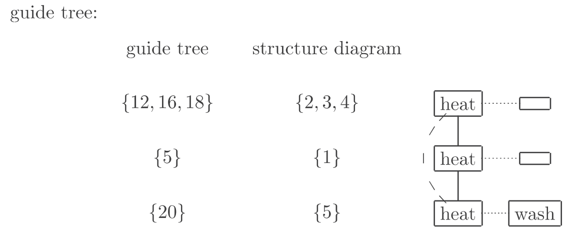


Fig. 5 Alignment of the procedures from the sub-group 'purification of copper, G_2'—References: 1: RCint, 6, 11.1–2; **2:** RMañj, 5, 28.1–2; **3:** RRS, 5, 52.1–2; **4:** RRĀ, R.kh., 8, 50.1–2; and **5:** RSK, 2, 18.1–2

number(s) of the procedure(s) that contain these items are printed right after an item (e.g. '[copper] [5]' = copper is mentioned in procedure 5).[11] If we compare this diagram with the sub-tree G_2 in Fig. 4, we can view the relationship between changes in the procedure and the order in the tree. Three texts (RMañj, RRS, and RRĀ) describe exactly the same procedure (copper and cow's urine are heated on a fire). The RCint adds sour juice and alkaline substances to the mixture, and therefore follows in a small (temporal) distance. Finally, the RSK adds a second step to the procedure, and is therefore placed farthest away from the group {2, 3, 4} in the guide tree:



About 120 sub-groups can be extracted from the alignment trees of thematic (sub)groups contained in the alchemy database. As mentioned above, the position of the procedures or texts is strongly influenced by ultrametric properties of the respective (sub)matrix. Since the (sub)matrices are not perfectly ultrametric in most cases and the basic requirement of a constant rate of mutation may be violated, the guide trees are not directly transformed into date informations. Instead, we look for statistically significant patterns in the placement of texts in these trees. To find these patterns, we extract two kinds of 'time rules' from (sub)trees. The first kind of rule records how often two texts follow each other in a guide tree [order rule (OR)]. For example, the rule RArn<RRS = 10:2 means that the RArn precedes the RRS in 10 of 10 +2=12 cases. The second kind of rule stores the distance between text pairs as given by the distance of their respective clusters calculated with UPGMA The distance of a text pair is assigned to one of three bins ('small', 'medium' and 'large' temporal distance). This results in distance rules (DR) such as RArn<>RRS = 2:5:40: the pair RArn and RRS is found in forty-seven (sub)trees, and they have a small temporal distance in only two of the forty-seven cases. Next, these rules are examined by using the binomial (OR) and multinomial (DR) test, with significance assigned at the 10% level. Rules which are not statistically significant at this level are discarded, and statistically significant distance rules are transformed into average distances. A group of external chronological restrictions that can be deduced from manuscript tradition and citations is added to the significant rules (Tables 1 and 2).

Table 1 Statistically significant time rules extracted from (sub)trees; $\alpha = 0.1$

OR	*P*-value	DR	Average distance	*P*-value
RCint<RMañj	0.0038	RCint<>RSK	0.43	0.0205
RHT<RCint	0.0625	RMañj<>RRS	0.14	0.0410
RCint<BhPr	0.0195	RMañj<>RRĀ	0.21	0.0597
RCint<ŚdhSaṃh	0.0313	RMañj<>RArṇ	0.35	0.0056
RSK<RCint	0.0625	RMañj<>RSS	0.13	0.0132
RPSudh<RMañj	0.0059	RMañj<>RSK	0.44	0.0781
RRS<RMañj	0.0717	RHT<>RPSudh	0.32	0.0781
RRĀ<RMañj	0.0216	RPSudh<>RRĀ	0.38	0.0087
RArṇ<RMañj	0.0021	RPSudh<>RArṇ	0.41	0.0260
RCūM<RMañj	0.0461	RPSudh<>ŚdhSaṃh	0.38	0.0888
RSK<RMañj	0.0625	RRS<>RCūM	0.07	<0.0000
RHT<RRS	0.0156	RRĀ<>RArṇ	0.16	0.0042
RHT<RSS	0.0313	RRĀ<>RCūM	0.35	0.0355
RPSudh<RRS	0.0145	RArṇ<>RCūM	0.40	0.0053
RArṇ<RPSudh	0.0730	BhPr<>ŚdhSaṃh	0.08	<0.0000
RPSudh<RSS	0.0461			
RPSudh<RCūM	0.0000			
RRĀ<RRS	0.0004			
RArṇ<RRS	0.0002			
RRS<RCūM	0.0717			
RRĀ<RSS	0.0121			
RRĀ<RCūM	0.0001			
RRĀ<BhPr	0.0065			
RRĀ<ŚdhSaṃh	0.0193			
RArṇ<RSS	0.0001			
RArṇ<RCūM	0.0003			
RArṇ<BhPr	0.0156			
RArṇ<ŚdhSaṃh	0.0313			

4 Dating the Alchemical Corpus

In the final step, time rules listed in Tables 1 and 2 are used to derive a relative chronology of the alchemical corpus. We use a technique that is motivated by multi-dimensional scaling. The texts have to be placed on a (one-dimensional) timeline in such a way that as many time rules as possible are fulfilled. In the beginning, each text is assigned to a random position (i.e. dating) that lies inside the global time frame of Indian alchemy (~900–1600 AD).[12] The goodness S of a single constellation of all texts on the timeline is calculated as follows. All texts T_i (dating: d_i) and their associated time rules are inspected one after another. If a time rule is fulfilled, nothing is added to S. Otherwise, a penalty γ is added to S:

(1) In the case of DRs, the absolute difference δ between the desired temporal distance Δ_{ij} of texts T_i and T_j as given in Table 1 (right half) and the actual difference of their datings ($|d_i - d_j|$) is calculated: $\delta = ||d_i - d_j| - \Delta_{ij}|$. If δ is larger than a threshold $\Theta = 100$, $\gamma = \delta \cdot (1 - P_{ij})^2$ is added to S. P_{ij} is the P value of the DR that connects text T_i with text T_j. The inverted and squared value of P_{ij} incorporates the statistical significance of the rule that connects T_i and T_j into the resulting penalty, because the violation of more significant rules results in a higher penalty.

(2) An OR can be interpreted as a temporal limit, which is set up by the dating of T_j for the dating of T_i (if $T_i < T_j$), or as an inequality condition $d_i < d_j$. A quadratic barrier function can be used as a good and easily differentiable approximation for this inequality condition [see Boyd and Vandenberghe (2004, p. 562ff.) for a similar, but logarithmic barrier]: if $d_i \geq d_j$, a penalty

$\gamma = (d_i - d_j)^2 \cdot (1 - p_{ij})^2$ is added to S. If a temporal limit D constituted by dated manuscripts (see Table 2), or the global time frame are transgressed, γ is calculated in a similar way: $\gamma = (d_i - D)^2$ (in these cases, P is not used because these rules are not assessed statistically).

The minimization of S, which leads to an optimal arrangement of datings, is performed with a gradient descent method. Apart from the amount function in DR,[13] the calculation is straightforward and yields a list of texts sorted by decreasing age. Gradient descent is not guaranteed to find a global minimum of the error function or the optimal order of texts. Therefore, the whole optimization is repeated several times. Since the data have inherent numerical uncertainties, we are not interested in absolute datings, but only in a relative chronology of the alchemical corpus. The resulting rank counts after fifty repetitions are reported in Table 3. The table shows that the alchemical literature has a clear tripartite structure, which is largely consistent with results of traditional philological research as reported in Meulenbeld (2000), White (1996), or Ray (1956):

(1) 'Early texts' comprise RHT, RArṇ, RAdhy, and RRĀ. The fact that the RSK is also contained in this group is the only important deviation from the traditional scheme of Indian alchemy. This is probably a misplacement since the RSK shows a combination of alchemical and āyurvedic theories, which is typical for the middle and late periods of Indian alchemy (see, e.g. the medical preparations in Chapter 4 of the RSK). As can be seen from Table 1, RSK is only connected to RCint and RMañj. It is possible that the low number of constraints was responsible for the misplacement of the text.
The remaining four texts certainly belong to the early group. It is an almost undisputed fact that RHT and RArṇ are founding texts of Indian alchemy. RRĀ is perhaps a compilation, but nevertheless contains basic and old material. Finally, the RAdhy comes from a seemingly isolated alchemical tradition which contains many archaic or at least extraordinary features. Among these special

Table 2 Fixed time rules deduced from dated manuscripts and citations

Rule	Source
BhPr<1558	manuscript (Meulenbeld 2000, IIA, 246)
RMañj<1546	manuscript (Meulenbeld 2000, II A, 638)
RRS<1699	manuscript (Meulenbeld 2000, II A, 670/71)
RRĀ<1473	manuscript (Meulenbeld 2000, II A, 665)
RAdhy<1386	manuscript (Meulenbeld 2000, II A, 616)
RCūM<RRS	citations
RArṇ<RRS	citations
RRĀ<RCint	citations
RHT<RCint	citations
RArṇ<RCint	citations
RHT<RCūM	citations
RCūM<RPSudh	citations

Table 3 Distribution of texts on ranks after 50 repetitions of the optimization algorithm

Text	1	2	3	4	5	6	7	8	9	10	11	12	13	avg. rank
BhPr							3		4	2	16	14	11	11.3
RAdhy	12	8	3	8	15	2	1	1						3.4
RArṇ	9	10	23	7	1									2.6
RCint						16	2	16	12	4				7.7
RCūM					4	30	13	3						6.3
RHT	19	13	12	5	1									2.1
RMañj								34	11	5				10.4
RPSudh						2	30	15	3					7.4
RRS							1	11	30	5	3			9.0
RRĀ			8	20	22									4.3
RSK	10	19	4	10	7									2.7
RSS								4	1	5	15	21	4	11.2
ŚdhSaṃh											5	10	35	12.6

features are alchemical apparatus of a quite primitive character, an unusual terminology using many 'alchemical hapax legomena' and an irregular arrangement of the alchemical *saṃskāras*.

(2) The 'middle layer' contains the group RCūM, RRS, and RPSudh. RRS, although held in highest esteem in Indian alchemy, is a compilation of previous alchemical material in large parts, and cites intensively from RCūM. The RPSudh formulates the same ideas as RCūM and RRS with slightly different words. In addition, RCint belongs to the middle group. This result is consistent with the style and content of this work.

(3) In the 'late group', we find texts that are of an obvious medical character. Apart from RMañj and RSS, which are alchemical works with an āyurvedic orientation, the āyurvedic texts BhPr and ŚdhSaṃh, which treat alchemy only in a few chapters, belong to this group. The date of ŚdhSaṃh found with the numerical method differs from the traditional research opinion. Dates proposed for ŚdhSaṃh range from the eleventh to the sixteenth century [cmp. (Meulenbeld, 2000, II B, p. 233)]. Meulenbeld and Ray place the work in the fourteenth century (middle layer) for internal reasons (cf. Meulenbeld, 2000, II A, p. 206–07; Ray, 1956, p. 160–61). The alchemical procedures described in the ŚdhSamh closely resemble those in the BhPr, which is a late text [Sixteenth century; cf. (Meulenbeld, 2000, IIA, p. 245)] and cites from the ŚdhSaṃh. The alchemical teachings of both texts can be compared with those of other late treatises such as the RMañj, where alchemy is partly superseded by āyurvedic theories. In my opinion, the ŚdhSaṃh should, therefore, be placed within the late group or at least between the middle and the late groups.

5. Conclusion

This article describes a numerical framework to deduce the chronological stratification of a corpus of Indian alchemical texts from changes in their content. The method is based on the assumption that the content of alchemical procedures changes over the course of time at an approximately constant rate. To capture these changes, methods from bioinformatics (sequence alignment, phylogenetic trees) are combined with basic non-parametric statistical tests and a gradient descent minimization algorithm. The results found using this approach coincide with the main philological theories regarding the development of Indian alchemy. Although detailed philological knowledge of the encoded texts (and a lot of time) was required to build the database of alchemical procedures, the contents-based numerical approach has yielded very promising results, and may be applied to investigate the (relative) chronology of other literary traditions of medieval India.

References

Bendixen, A. (1990). Überlegungen zu *Ratnarasasamuccaya* (kapitel 1-11) und zu *Rasendracūḍāmaṇi*. *Journal of the European Āyurvedic Society*, **1**: 165–66.

Bhāvamiśra (1947). *Bhāvaprakāśa*, Jaya Krishna Das Hari Das Gupta, Benares.

Boyd, S. and **Vandenberghe, L.** (2004). *Convex Optimization*. Cambridge: Cambridge University Press.

Burrows, J. (1992). Not Unless you ask Nicely: The Interpretative Nexus between Analysis and Information. *Literary and Linguistic Computing*, **7**(2): 91–109.

Ḍhuṇḍukanātha (2000). *Rasendracintāmaṇi*, 1st edn, Varanasi: Caukhambha Orientalia.

Fosse, L. M. (1997). *The Crux of Chronology in Sanskrit Literature*, Oslo: Scandinavian University Press.

Gerow, E. (1977). *Indian Poetics, A History of Indian Literature, Volume V, Fasc. 3*. Wiesbaden: Otto Harrassowitz.

Gopālakṛṣṇa (n.d.). *Rasendrasārasaṃgraha*. Dillī: Caukhambā saṃskṛ t pratiṣṭhān.

Govindabhagavatpāda (1989). *Rasahṛdayatantra*. Vārāṇasī: Caukhambha Orientalia.

Hoover, D. L. (2002). Frequent word sequences and statistical stylistics. *Literary and Linguistic Computing* **17**: 157–80.

Kāyasthacāmuṇḍa (1984). *Rasasaṃketakalikā*. Chaukhamba Amarabharati Prakashan: Benares.

Meulenbeld, G. J. (2000). *A History of Indian Medical Literature*. Groningen: Groningen Oriental Studies, Egbert Forsten.

Minsky, M. (1975). A Framework for Representing Knowledge. In Winston, P. (ed.), *The Psychology of Computer Vision*. New York: McGraw-Hill.

Miśra, C. (1999). *Rasakāmadhenu*, 2nd edn. Varanasi: Chaukhambha Orientalia.

Needleman, S. B. and Wunsch, C. D. (1970). A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology*, **48**: 443–53.

Nityanātha (1985). *Rasaratnākara. Rasakhaṇḍa.*, 1st edn., Varanasi: Chaukhamba Amarabharati Prakashan.

Nityanātha (1991). *Rasaratnākara. Vādikhaṇḍa-Ṛddhikhaṇḍa*, *Vidyāvilāsa āyurveda granthamālā, 10*, 1st edn., Varanasi: Chaukhambha Orientalia.

Quillian, M. (1968). Semantic Memory. In Minsky, M. (ed.), *Semantic Information Processing*. MIT Press: Cambridge, Mass, pp. 227–70.

Ray, P. C. (1956). *History of Chemistry in Ancient and Medieval India*, Calcutta: Indian Chemical Society.

Ray, P. C. (ed.) (1910). *Rasārṇava.*, Calcutta: Satya Press.

Rocher, L. (1986). *The Purānas, Vol. II, Fasc. 3 of A History of Indian Literature.* Wiesbaden: Otto Harrassowitz.

Śālinātha (1995). *Rasamañjarī, Jaikrishnadas Ayurveda Series, 81*, Varanasi: Caukhambha Orientalia.

Śarman, R. (ed.) (1930). *Rasādhyāya, Chowkhamba Sanskrit Series*, North of Gopal Mandir, Benares.

Śārṅgadhara (2000). *Śārṅgadharasaṃhitā*. Krishnadas Academy, Benares.

Sastri, S. R. (ed.) (1952). *Ānandakanda*. (Edited with translation in Tamil, and introduction in Tamil and Sanskrit). Sri Vilasam Press, Tanjore.

Sokal, R. and Michener, C. (1958). A Statistical Method for Evaluating Systematic Relationships. *University of Kansas Science Bulletin*, **38**: 1409–438.

Somadeva (1999). *Rasendracūḍāmaṇi*. Vārānasī: Caukhambhā Orientalia.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific Gap penalties and Weight Matrix Choice, *Nucleic Acids Research*, **22**(22): 4673–80.

Trautmann, T. P. (1971). *Kauṭilya and the Arthaśāstra*, Leiden: E.J. Brill.

Tse, E. K., Tweedie, F. J. and Frischer, B. D. (1998). Unravelling the Purple Thread: Function Word Variability and the Scriptores Historiae Augustae, *Literary and Linguistic Computing*, **13**: 141–47.

Vāgbhaṭa (1952). *Rasaratnasamuccaya*. Benares: The Chowkhamba Sanskrit Series Office.

White, A. G. (1996). *The Alchemical Body*, Chicago and London: The University of Chicago Press.

Yaśodhara (1940). *Rasaprakāśasudhākara*. Rasaśālā auṣadhāśram, Goṇḍal.

Notes

1 Results presented in this article are part of the research project 'Rasavidya. Computer based analysis of medieval Indian alchemy.' The project was funded by the Deutsche Forschungsgemeinschaft (DFG) in the year 2005–06.

2 For a general survey of the problems connected with this literature, see Rocher (1986, p. 100–3) or Fosse (1997, p. 138ff.).

3 The most obvious example is the use of frequent (function) words to distinguish between texts or text layers. Although these methods yield convincing results, it is sometimes difficult to interpret these results in terms of language history (what, for instance, is the philological meaning of the statement, that any component of a PCA contrasts the words 'and' and 'or'?).

4 The corpus investigated in this article consists of the following texts: Ānandakanda [ĀK; (Sastri, 1952)], Bhāvaprakāśa [BhPr; (Bhāvamiśra, 1947)], Rasakāmadhenu [RKDh; (Miśra, 1999)], Rasaprakāśasudhākara [RPSudh; (Yaśodhara, 1940)], Rasamañjarī [(RMañJ; (Sālinātha, 1995)], Rasaratnasamuccaya [RRĀ; (Vāgbhata, 1952)], Rasaratnākara [RRA; (Nityanātha, 1985, 1991)], Rasasaṃketakalikā [RSK; (Kāyasthacāmuṇḍa n.d.)], Rasahṛdayatantra [RHT; (Govindabhagavatpāda, 1989)], Rasādhyāya [Radhy; (Śarman, 1930)], Rasārṇava [RArn; (Ray, 1910)], Rasendracintamani [RCint; (Dhundukanātha, 2000)], Rasendracūḍāmaṇi [RCüM; (Somadeva, 1999)], Rasendrasārasaṃgraha [RSS; (Gopālakṛsṇsa n.d.)], Sārṅgadharasaṃhitā [SdhSamh; (Sārṅgadhara, 2000)].

5 The first scientific study of this field is Ray (1956). Meulenbeld (2000, IIA, ch 10, p. 583ff.) offers a comprehensive survey of alchemical sources and

secondary literature. White gives a shorter overview of the most important works (White, 1996, p. 144ff.).

6 This model, as well as any other program part mentioned later, is written in C++ and compiled with Microsoft Visual Studio .NET 2003.

7 $p1$ comes from BhPr, 2, 3, 205.1-206.3 and $p2$ from RCint, 5, 4.1-5.2. $p3$ is a fancy procedure added to make the example a bit more complex.

8 The distance matrix M is ultrametric if for any three items i, j, and k the triangle inequality of a standard metric can be replaced by the stronger ultrametric inequality $M(i, j) \leq \max(M(i, k), M(j, k))$. This can be shown to be equivalent to the condition that the two largest values from $M(i, j)$, $M(i, k)$, and $M(j, k)$ are equal.

9 The lengths of branches in this tree are not proportional to the distances of subordinate clusters! Numeric values at nodes give the inter-cluster distance calculated with UPGMA.

10 Even sub-matrices are only in a few cases really ultrametric. To test for approximate ultrametric properties, the strict ultrametric inequality (see Note 8 above) is relaxed to the following form: three items i, j, and k fulfil the relaxed condition if the absolute difference between the two largest values from $M(i, j)$, $M(i, k)$, and $M(j, k)$ is smaller than 0.1. In the complete distance matrix for the purification of copper, 74% of all triplets fulfil this condition (average distance between the two largest elements: 0.07); on the contrary, 90% of the triplets in the submatrix corresponding to G_2 fulfil the relaxed condition (average distance: 0.02). The sub-matrix is clearly closer to an ultrametric matrix than the complete distance matrix.

11 The numbering of procedures in Figs 4 and 5 does not correspond!

12 Inspite of many unclear details concerning the internal order of Indian alchemy, 900–1600 AD is a good estimation for the global time frame of this tradition. The upper limit can be fixed with dated manuscripts. The date of the lower limit is more speculative, values of 700 or 1000 are also possible (this is indeed a rather small variation in the history of Indian literature!). Nevertheless, very early beginnings ($\ll$500 AD) as proposed by some authors are probably wrong.

13 This function is not continuously differentiable at $d_i = d_j$ and $d_i = d_j \pm \Delta_{ij}$. The solution $dj \pm \Delta$ij can be ignored for $\Theta \neq 0$ because the function is not evaluated in an interval of size.