Colophons in Sanskrit manuscripts

A study of the Sanskrit Library manuscript catalogue of manuscripts at Harvard University, the University of Pennsylvania, and Brown University

Anuja P. Ajotikar, Tanuja P. Ajotikar, and Peter M. Scharf

Introduction

The constitution of the text of a work in Sanskrit or in any language depends upon consulting original manuscripts. Critical editing generally seeks to establish the text that is the common ancestor witnessed in numerous manuscripts of that text. This common text may include rubrics and final rubrics that provide information about the text, its author or the circumstances of its authorship, i.e. headers such as अथ श्रीमद्भगवद्गीता and अथ प्रथमो ऽध्यायः, and trailers such as इति श्रीमद्भगवद्गीतासूपनिषत्सु ब्रह्मविद्यायां योगशास्त्रे श्रीकृष्णार्जुनसंवादे ऽर्जुनविषादयोगो नाम प्रथमो ऽध्यायः and इति समाप्तं योगदर्शनम्. Yet manuscripts generally additionally include a colophon that is neither part of the text itself nor a heading or trailer. Colophons provide information about the scribe and the origin of the manuscript. While the entire text, headings and trailers are merely copied by the scribe, unlike these, the colophon is composed by the scribe himself. Since scribes are frequently not learned in Sanskrit, these colophons contain unique variations not found in the body of the manuscript. These variations include not only Sanskrit synonyms and abbreviations for standard content included in these passages, but also accidental corruptions, deliberate regional language variants, and their corruptions and abbreviations. To make sense of these passages requires intimate familiarity with these variations; knowledge of standard Sanskrit is generally insufficient. Most of the corruptions and variants remain opaque to existing digital Sanskrit dictionaries, parsers and other tools. To facilitate the deciphering of colophons, we describe the creation of a digital resource to provide access to a wide range of the variations found in them.

Critical editions use the information in colophons to determine the history of the transmission of the text and to construct a stemma. While good critical editions convey the results of their analysis of these passages, they are not primarily concerned with process of deciphering the passages. Unlike critical editions which are concerned more with the common text than with the idiosyncrasies of their witnesses, descriptive manuscript catalogues are principally concerned with colophons for essential information about the manuscripts they contain. Descriptive catalogues scrupulously record these passages and often index important information in them. These catalogues thus present a corpus of original compositions in the form colophons whose authors are scribes. These colophons constitute a unique genre of literature with its own elements of style that deserve some attention in their own right.

A project conducted 2009–2013 by the Sanskrit Library designed a detailed digital manuscript catalogue based on the Text-Encoding Initiative manuscript cataloguing guidelines, the alignment of digital images with corresponding digital editions, and the linking of manuscript images and digital texts with the digital catalogue. While the 2009–2013 project digitally catalogued and indexed about a hundred and sixty Sanskrit manuscripts at Brown University and the University of Pennsylvania, a project 2013–2016 catalogued and indexed the entire collection of about eighteen hundred Sanskrit manuscripts at Harvard University. While the overwhelming majority of these manuscripts were written in Devanāgarī script, several are in Telugu, Bengali, Śāradā and Nandīnāgarī. An index to the catalogue is available at `http://sanskritlibrary.org` under *Reference* or directly at `http://sanskritlibrary.org/catindex.html`. Scharf (2015, n.d.[a],[b],[c]) described in detail the structure and content of the catalogue and the process of its creation. The entries in this catalogue include the original text of colophons as well as their interpretation. The present paper analyzes the colophons available in this collection and proposes some ways to use this data effectively. In particular, we propose to develop an XML database of colophons that strictly categorizes the data available in them and exposes the variations in key components so as to assist scholars in the interpretation of colophons elsewhere.

Components of a colophon

While the term *colophon* may denote any statement at the end of a work that provides information about its authorship or production, in accordance with the Text-Encoding Initiative, we distinguish a colophon from a final rubric, defining the latter as containing information about the author and work, and defining the former as containing information about the scribe and manuscript.[1] For example, the final rubric इति श्रीभविष्योत्तरपुराणे अनन्तव्रतकथा सम्पूर्णा समाप्ता। states that the work entitled *Anantavratakathā* which occurs in the larger context of the *Bhaviṣyottarapurāṇa* is complete. The colophon संवत् १८६९ चैत्रमासे कृष्णपक्षे तिथौ ७ बुधवासरे, occurring immediately after the final rubric, gives the date on which the scribe completed writing this copy of the work. The date includes the era, year, month, fortnight, date (typically in units of 1/30th of a month called tithi), and day of the week, here literally, "on Wednesday, the seventh tithi in the dark fortnight in the month of Caitra in the year 1869 of the Vikrama era".

In addition to the date, a colophon often includes the name of the scribe, the place in which it was written, the purpose for which it was written, sometimes including the person for whom it was written, and the owner. The colophon may also include additional details about the scribe or owner such as genetic and scholastic lineage, and native place. For example, the colophon सं० १८४१ कार्त्तिमा० कृष्णप० ४ शनिवासरे आत्मपठनार्थम् अलिखत् includes the purpose. Literally this colophon states, (the scribe) wrote (the manuscript) on Saturday, the fourth day of the dark fortnight in the month of Kārttika in the year 1841 of the Vikrama era for the sake of his own reading.

Colophons may be composed in the verse as well as in prose. For example, the following colophon, which mentions owner and his father, is composed in verse:

> जगन्नाथः पिता यस्य सम्राड् इत्य् उपनामकः।
> दीनानाथाभिधेयस्य जगज् जानातु पुस्तकं॥

[1] David Pingree used the terms *colophon* and *post-colophon* for our final rubric and colophon respectively.

"Let the world know that this book belongs to the one
whose name is Dīnānātha whose father is Jagannātha Samrāṭ."

Colophons include some unusual vocabulary. Scribes sometimes use real-world terms (*bhūtasaṅkhyā*) to represent numbers designating the year in dates, and epithets for days of the week. For example, संवत् इन्दुरसबाणगुणसंख्यामिते वर्षे बाहुलमासे बहुलपक्षे दीपालिकादिने रजनीवल्लभवारे मंगलपुरमद्ह्ये लिखितं स्वकृते।"(The manuscript) was completed on Monday, the last day of the dark fortnight in the month of Kārttika in the year 1653 of the Vikrama era, in Maṅgalapura for the sake of his own reading." In this colophon, the year 1653 is designated by using the real-world terms indu 'moon' for 1 because the earth has just one moon, rasa 'taste' for 6 because Āyurvedic medicine enumerates six tastes, bāna 'arrow' for 5 because in Indian mythology the god love Kāma has five arrows, and guṇa 'constituent' for 3 because in ontological system of the Sāṅkhya darśana there are three fundamental constituents of nature. In this colophon, the day Monday is referred to using the word *rajanīvallabha* which literally means 'dear to the night', an epithet of the moon. No Sanskrit dictionary includes this word, although there are 84 words listed by Monier Williams that end in the word *vallabha*, one of which is *rohiṇīvallabha* 'lover of Rohiṇī, the Moon'.

Tagging scheme for colophons

Tokenization is needed in order to tag segments of the colophon. Although Sanskrit manuscripts generally leave no spaces at all between words, our transcriptions supply these where Romanization permits. As is well-known, Sanskrit sandhi obscures word boundaries possibly even combining the final sound of a preceding word and the initial sound of a following word in a single character. Therefore, analysis of sandhi is a prerequisite to tagging segments of a string.

We segment the sandhi-analyzed colophon into phrases and individual words and tag segments and subsegments using XML tags. As described above, the colophon contains phrases containing the following information:

- date
- scribe
- place
- owner
- purpose
- verb

Each of these phrases contains subordinate components. The date contains the following parts:

- day of the week
- date
- fortnight
- month
- year
- era

The scribal phrase or owner phrase may contain a proper name, family name, father's name, teacher's name, professional title, etc. The place phrase may contain the name of a city, region, or geographical feature such as a river or mountain. The purpose phrase may contain names and pronouns. The verb phrase often consists of a single verb. Where TEI provides tags for these items we adopt them. We add several additional tags not available in TEI. We therefore suggest the following structure to tag colophons:

1. datephrase
    a. dayseg
    b. dateseg
    c. fortnightseg
    d. monthseg
    e. yearseg
    f. eraseg
2. scribephrase
    a. persName
        i. forename
        ii. surname
    b. roleName
3. placephrase
    a. placeName
    b. geogName
4. ownerphrase
    a. persName
        i. forename
        ii. surname
    b. roleName
5. purposephrase
    a. persName
        i. forename
        ii. surname
    b. roleName
    c. pronoun
6. verbphrase
    a. verb

In addition to these elements, we use the element `keyword` with a `type` attribute to categorize keywords commonly found in most of the terminal categories shown above.

In Figure 1, we mark up the last example cited above in the previous section, namely, संवत् इन्दुरसबाणगुणसंख्यामिते वर्षे बाहुलमासे बहुलपक्षे दीपालिकादिने रजनीवल्लभवारे मंगलपुरमद्ह्ये लिखितं स्वकृते।

```
<colophon>
 <datephrase>
  <eraseg>
   <era>saMvat</era>
  </eraseg>
  <yearseg>
   <year>indurasabARaguRa</year>
   <keyword type="year">saMKyAmite varze</keyword>
  </yearseg>
  <monthseg>
   <month>bAhula</month>
   <keyword type="month">mAse</keyword>
  </monthseg>
  <fortnightseg>
   <fortnight>bahula</fortnight>
   <keyword type="fortnight">pakze</fortnight>
  </fortnightseg>
  <dateseg>
   <date>dIpAlikA</date>
   <keyword type="date">dine</keyword>
  </dateseg>
  <dayseg>
   <day>rajanIvallaBa</day>
   <keyword type="day">vAre</keyword>
  </dayseg>
 </datephrase>
 <placephrase>
  <placeName>maMgalapura</placeName>
  <keyword type="place">maDye</keyword>
 </placephrase>
 <verbphrase>
  <verb>liKitaM</verb>
 </verbphrase>
 <purposephrase>
  <pronoun>sva</pronoun>
  <keyword type="purpose">kfte</keyword>
 </purposephrase>
</colophon>
```

Figure 1: XML markup of a colophon

Variations

Each of the elements described in the previous section may contain a variety of content. Obviously, the content may vary due to actual differences in the object described, such as different months of the year, different scribes, and different owners. Where actual differences are limited, such as in the months of the year, we indicate the intended object in a separate attribute; for example, we enumerate the months of the year in an `n` attribute with values 1–12. Yet even where the object described is the same, variations are encountered due to the use of synonymous Sanskrit terms, terms in vernacular languages, abbreviations, and errors. We indicate the type of variation in a type attribute with the following values:

1. sanskrit
2. vernacular
3. sktabbr
4. verabbr
5. skterror
6. vererror

We collected about 700 colophons from the Sanskrit Library manuscript catalogue of manuscripts at Harvard University, the University of Pennsylvania, and Brown University, and propose to tag them according to the scheme mentioned above. On the basis of a preliminary survey of these, we collected some sample variations which we describe in the examples below.

Variation in the names of months

Among thirty occurrences of reference to the seventh month, namely, the month of *āśvina*, in colophons, we find an astonishing variety of terms used for it: the following sixteen different terms are found:

1. *Āśvina*
2. *Aśvayuk*
3. *Āśvinya*
4. *Aśvana*
5. *Āśvana*
6. *Aśvina*
7. *Āsvina*
8. *Aśvī*
9. *Āśvana*
10. *Aśvan*
11. *Asuna*
12. *Āsoja*
13. *Āsaja*
14. *Āsosa*
15. *Kuāra*
16. *Āso*

Of these, the first two are Sanskrit terms, the first of which is the common name of the month and the second of which is an older term referred to by Pāṇini in *A*. 4.3.36. The third through tenth are corruptions of the first. The eleventh through fifteenth are vernacular terms; the eleventh is derived from (1), (12)–(14) are derived from (2); and (15) is a vernacular synonym. The sixteenth term is an abbreviation of the vernacular derived from the second. Only the first two are found in a Sanskrit dictionary. Only one of the vernacular terms, (15), is found in an on-line Hindi dictionary (`pustak.org`). The terms that are corrupt versions of a Sanskrit term or vernacular terms in pre-modern regional languages are not available in dictionaries. The compilation of these variants is necessary to permit their automated identification.

Similarly among the thirty-four occurrences of reference to the ninth month, namely, the month of *mārgaśīrṣa*, we find the following twelve different terms used by scribes:

1. *mārgaśīrṣa*
2. *mārgaśira*
3. *uttamamāsa*
4. *mārgasira*
5. *mārgaśīra*
6. *mārgāśara*
7. *magaśīrṣa*
8. *mārgaś*
9. *mārga*
10. *māgasarasa*
11. *agahana*
12. *māsara*

The first is the common term, the second a synonym, and the third an epithet ‘best month’. The fourth through sixth are corruptions of the second. The seventh is a corruption of the first. The eighth and ninth are abbreviations of the first. The tenth may be a corruption of *Māgasara*, the name found in an early Hindi dialect. The eleventh is a Hindi tadbhava word derived from an old Sanskrit synonym *āgrahāyaṇa* (Pāṇini *A*. 5.4.36). The last may be a corruption of an old vernacular of the first or second.

Variation in the verb

In most colophons some form of the verb meaning ‘write’ is used. The following twelve variations of this verb are found:

1. *likhitam* ppp (very common)
2. *alekhi* aor p3s (frequent)
3. *vyalekhi* aor p3s (infrequent)
4. *alikhat* ipf a3s (infrequent)
5. *lilekha* prf a3s (rare)
6. *likhāpitam* ppp of causative (rare)
7. *lipīkṛtam* (infrequent)
8. *lipy akarot* (rare)
9. *liṣitam* (frequent)

10. *liṣyate* (frequent)
11. *alekhit* (rare)
12. *liṣīk̥rtam* (rare)

The first six are grammatically correct forms. The seventh and eighth are Sanskrit-like grammatical abberations based upon Hindi compound verb forms. These can be analyzed as influenced by modern Indian language use of a complex verb construction because they are combinations of the word *lipi* 'script' and the verb *k̥r* 'do'. The ninth through twelfth are corruptions of the preceding. The ninth is a corruption of the first, the tenth of the passive third singular *likhyate* which does not occur. The eleventh is a cross between (2) and (4). The twelfth is probably a scribal error for (6). The grammatical abberations and corruptions are not described in any of the grammar of Sanskrit or found in any lexical source.

Conclusion

Colophons represent a body of literature composed by scribes, much of which is in non-standard language. Digital descriptive cataloguing of manuscripts has made available an extended corpus of this literature for systematic study. We have described here a plan to analyze this corpus for further investigation and practical use.

The bulk of the variations described in the previous section are non-standard forms not found in printed literature and not available in lexical and grammatical sources. We aim to create a database of of these variants and to develop software to permit access to them for all of the categories of information found in colophons described in our tagging scheme. Software might include a dictionary-like look up tool, a standardizer, and a translator. The last might be linked to software already available such as Grard Huet's *Sanskrit Reader Companion* (`http://sanskrit.inria.fr/DICO/reader.html`) Michio Yano's *Pancanga* (`http://www.cc.kyoto-su.ac.jp/~yanom/pancanga`). These products will assist in the cataloguing and editing of manuscripts which constitute the foundation of philological work in Sanskrit. They will also provide data for the study of linguistic history of Indian languages.

References

Scharf, Peter M. "Accessing manuscripts in the digital age: hypertext presentation, cataloguing, and text-image alignment." *unknown*. Tattvabodha 00. New Delhi: National Mission for Manuscripts. [forthcoming.]

—. "Accessing manuscripts in the digital age: a pipeline to create a hypertext catalogue, and searchable access to manuscript images." *unknown*. Tattvabodha 00. New Delhi: National Mission for Manuscripts. [forthcoming.]

—. "Aṅkīyayuge hastalikhitānāmabhigamaḥ: sūciracanaṁ pratimāvāgyojanā parasparasaṁbaddhānāṁ ca sūcipratimāsaṁskaraṇānveṣaṇatantrāṁśānāṁ jāle pradarśanam: Accessing manuscripts in the digital age: cataloguing, hypertext presentation, and internet display of interconnected catalogue, digital images, editions, and search software." *Proceedings of the National Seminar on the Application of Information Technology for Conservation, Editing and Publication of Manuscripts, 20–22 January 2013, Bangalore*. Mysore: Samskriti Foundation. [forthcoming.]

—. 2015. “Providing access to manuscripts in the digital age.” *From Mulberry leaves to silk scrolls: new approaches to the study of Asian manuscript traditions*, ed. by Justin Thomas McDaniel and Lynn Ransom, pp. 231–71. The Lawrence J. Schoenberg Studies in Manuscript Culture 1. Philadelphia: University of Pennsylvania Libraries.