Development of a TEI standard for digital Sanskrit texts containing commentaries: A pilot study of Bhaṭṭi's *Rāvaṇavadha* with Mallinātha's commentary on the first canto

Tanuja P. Ajotikar[1], Peter M. Scharf[2]
[1] Maharishi Vedic University, Vlodrop, The Netherlands
[2] IIIT Hyderabad
`tanuja@sanskritlibrary.org, scharf@sanskritlibrary.org`

Abstract

Sanskrit commentarial literature is important because it pervades all fields of Sanskrit literature and provides rich information in the form of references, explanation and derivation of specialized vocabulary, explanation of metrical patterns, syntax, figures of speech, etc. This information is invaluable to linguists, philologists, and other humanities scholars. In the digital age, it is possible to identify, extract, and reorganize this information easily if contemporary techniques of digital humanities are employed. The extensible markup language (XML) has emerged as the most widely used method of data-interchange, and the Text-Encoding Initiative (TEI) has become recognized as the source of global guidelines for the production of digital editions of texts. Yet there is a need to develop conventions for employing TEI in standard ways for Sanskrit texts and commentaries to maximize the universal accessibility of these texts. After explicating the structure and elements present in commentaries, we describe templates to encode them in accordance with the TEI guidelines to represent relevant structural and semantic information in addition to document strucure. We employ these templates to encode Mallinātha's commentary on the twenty-seven verses of the first sarga of Bhaṭṭi's *Rāvaṇavadha*. We then develop extensible stylesheet language transformations (XSLT) to extract Mallinātha's rendering of the verse in prose syntax (*anvaya*) and to create a dictionary of synonyms and derivations, and indices to words, citations, and definitions of metrical patterns and figures of speech. Finally, we offer some initial observations on the bearing of Mallinātha's anvaya on Sanskrit syntax.

Keywords: XML, TEI, Sanskrit, commentary, text-encoding, syntax

1 Introduction

Scharf (2018) reviewed the history of the development of the extensible markup language (XML) and the Text-Encoding Initiative (TEI). He noted that the principal purpose of XML is to separate the semantic markup of a digital text from software that implements its presentation digitally or in print. Given that the XML metalanguage permits the development of an unlimited variety of markup languages, TEI provides guidelines to standardize such markup to facilitate reuse and interchange. Given that TEI itself offers numerous alternative methods to encode texts, to maximize the universal accessibility of digital Sanskrit texts, there is a need to develop standards for employing TEI. Scharf (2015) and Scharf (2017) developed a TEI template for the digital descriptive cataloguing of Sanskrit manuscripts. Such a standard template needs to be developed for Sanskrit texts. This task is complicated by the fact that Sanskrit literature is vast and diverse. Each discipline has its own special features that require special handling. Sanskrit literature includes not only base texts but also commentaries and sub-commentaries. This commentarial literature is important because it pervades all fields of Sanskrit literature and provides rich information in the form of references, explanation and derivation of special-

ized vocabulary, explanation of metrical patterns, syntax, figures of speech, etc. Hence commentaries are filled with quotations, derivations, metrical definitions, syntactic reordering and much more additional information which needs to be properly encoded. In a paper presented at the international conference on the tradition of commentary and the stream of knowledge (*Bhāṣyaparamparā Jñānapravāhaśca*), Ajotikar and Scharf (2020) explicated the structure and elements present in commentaries by examining Mallinātha's commentaries on the first verse of the five principal Sanskrit major poetic works (*mahākāvya*s) and devised preliminary templates in accordance with the TEI guidelines to represent relevant structural and semantic information in addition to document structure. They concluded by outlining procedures to build a library of digital Sanskrit commentarial literature. We review the features of commentary and templates to encode them here and describe our use of them to encode Mallinātha's commentary on the twenty-seven verses of the first sarga of Bhaṭṭi's *Rāvaṇavadha*. We then demonstrate the utility of our encoding for the creation of a dictionary of synonyms and derivations, and indices to words, citations, and definitions of metrical patterns and figures of speech using extensible stylesheet language transformations (XSLT) and for extracting Mallinātha's rendering of the verse in prose syntax (*anvaya*). We conclude by offering some initial observations on the bearing of Mallinātha's anvaya on Sanskrit syntax.

2 Principal features of commentary

Each mahākāvya begins with a benedictory verse. Although the benedictory verses differ in each mahākāvya, after a brief introduction, the commentary on each begins by describing the importance of a benedictory verse. The commentary on the benedictory verse and each subsequent verse begins with the citation of the incipit (*pratīka*) of the verse. The commentary then proceeds to explain each word in the original verse in prose order. At relevant junctures, the commentary explains metrical patterns, figures of speech, and often concludes by summing up the purport of the verse. These explanations include a number of features which are summarized as follows:

1. Most words in the base text are provided with a contextually appropriate synonym (*paryāya*) that may reveal its morphological structure as well as its semantic appropriateness.
2. Most commentaries provide words with a derivation (*vyutpatti*).
3. The metrical patterns (*chandas*) used in poetic works are explicitly named and explained by reference to works in the highly sophisticated discipline of Sanskrit poetics (*chandaḥśāstra*).
4. The figures of speech (*alaṅkāra*) used in poetic works are explicitly named and explained by reference to works in the highly sophisticated discipline of Sanskrit rhetoric (*alaṅkāraśāstra*).
5. In providing derivations and explanations of metrical patterns and figures of speech, as well as discussing more general aspects of the verse or the work as a whole, commentators frequently cite the authoritative works of their predecessors.
6. Commentators generally transform verse into prose in the process of their explanation. Mallinātha explains each verse by reordering the words in the syntactic order in which they would appear in ordinary Sanskrit prose (*anvyaya*).

3 TEI elements and attributes employed

In order to capture all of the information described in the preceding section, a number of TEI elements and attributes are employed. These include particular usage of elements and attributes besides the basic text encoding elements and attributes for verse and prose encoding described by Scharf (2018), Scharf (2020), Scharf (2015) and Scharf (2017) in his explanation of manuscript text transcription. We describe how to encode various types of analysis of the base text and its commentary below.

We analyze the base text morphologically in a separate file. Each inflected word is put in a `w` element. We identify the nominal base (*prātipadika*) of each nominal, and the verbal root (*dhātu*) of each finite verb in a `lemma` attribute, the lexical identifier in a `type` attribute, and

the inflectional identifier in a `subtype` attribute. Morphemes are indicated by putting the form of the basic element in subordinate `m` elements and their lexical identifier in a `type` attribute. An analytic paraphrase (*vigraha-vākya*) is provided in a `seg` element with `type="vigraha"` as the first child of each compound. Each `w` and `m` element is given an `xml:id` attribute as is each verse and chapter.

The commentary is entered in a separate file from the base text on which it comments. The commentary on each verse is linked with the verse on which it comments by using identical values of the verse's `xml:id` in the base text file and in the commentary file. The explanation of each major division of the commentary on the verse is put in an `ab` element which is supplied with an `xml:id`. The explanation of each word in the commentary is put in an `ab` element with the `type` attribute given the value `pada` and a `corresp` attribute given the value of the `xml:id` of the sandhi-analyzed word in the morphological analysis file. Each word in the base text commented upon in the commentary is put in a `w` element as a child of this `ab` element and is provided with an `n` attribute indicating the sequence in which the word appears in the commentary. Each word is also provided with an `xml:id` that coincides with its `xml:id` in the morphological analysis file and the `corresp` attribute of its parent `ab` element. If a word is not explicitly mentioned yet is commented upon or supplied with a derivation, it is given a `type` attribute with the value `inferred`. The few words that are neither mentioned explicitly nor commented upon are put in a `w` element in the expected sequence in the commentary with `type="supplied"`.

Where Mallinātha supplies a word with a synonym, the synonym is put in a `seg` element as a sister to the `w` element containing the word. The `seg` element is supplied with the attribute-value pair `type="synonym"` and a `corresp` attribute with the value of the `xml:id` of the `w` element. Where Mallinātha supplies the antecedent of a pronominal, we put the antecedent in a `seg` element with the attribute-value pair `type="antecedent"`.

Where Mallinātha supplies a word with a derivation, the entire derivation is put in a `seg` element as a sister to the `w` element containing the word and is supplied with the attribute-value pair `type="derivation"` and a `corresp` attribute with the value of the `xml:id` of the `w` element. Citations of sūtras are tagged with `ref` elements with `corresp` attributes with the value of the xml:id of the original Pāṇinian rule in the Sanskrit Library's sūtra file. The `ref` element is immediately followed by a `note` element containing a `bibl` element which identifies the exact edition, and which has a `biblScope` element which again identifies the sūtra number in the *Aṣṭādhyāyī*.

All quotations are tagged. The `quote` element is used for long passages, and the `q` element for short passages. We analyze the prosodic changes (*sandhi*) at the beginning and end of the quote and put element tags immediately before and after the quote, i.e., before *iti*. We use the `bibl` element within a `note` element immediately following the quote to identify the source of the citation. The entire phrase containing the quote and its reference is put in a `seg` element where classification of the quotation is added in the `type` attribute. Many quotations occur within longer derivation segments which bear the `type="derivation"` attribute-value pair. Most others are citations from lexicons, so their containing `seg` element bears the attribute-value pair `type="lexicon"`. Until the need arises for greater specificity, the rest, which are citations to dharmaśāstra and literary texts, occur within `seg` elements with the attribute-value pair `type="quotation"`.

Explanations of metrical patterns are tagged within `ab` elements with the attribute-value pair `type="vftta"`. An `ana` attribute names the meter (*chandas*) the commentator specifies in his explanation.

Explanations of figures of speech are tagged within `ab` elements with the attribute-value pair `type="alaNkAra"` and an `ana` attribute that names the figure of speech (*alaṅkāra*).

The incipit (*pratīka*) cited by the commentator at the beginning of the commentary on each verse is put in an `ab` element with the attribute-value pair `type="pratIka"`. Comments made in the commentary that tell the purport of the verse are set in an `ab` element with the attribute-value pair `type="tAtparya"` unless intertwined with the explanation of a word in which case

it is set in a `seg` element with `type="tAtparya"`. Where Mallinātha supplies words not in the verse by stating *iti śeṣaḥ*, we put the phrase in an `ab` element with the attribute-value pair `type="Seza"`. Other comments that explain aspects of the verse that do not fit in any of these categories are put in an `ab` element with the attribute-value pair `type="vyAKyAna"`.

We also capture syntactic information provided by the commentators reordering of the verse in ordinary prose. Once we have completed tagging the commentary on a versified text, each word in the original verse appears in a `w` element with an `xml:id` attribute indicating the sequence of the word in the original verse, and with an `n` attribute indicating its sequence in the commentator's reordering of the syntax in ordinary prose. We automatically extract the words of each verse in the commentator's prose order from the commentary file and place them in an `s` element within a `p` element in that order in a separate file to obtain a file that shows the words of the verse in the prose order in which they are explained by the commentator. In order to show the variance of the prose syntax from the syntax of the verse, a numbered version of this prose file is created in which each word is supplied with the sequence number of the word in the original verse as indicated by the `xml:id`. Where words are inferred or supplied they are inserted in square brackets and given an explicit `"inferred"` or `"supplied"` tag.

Finally, we use the `seg` element to encode segments of text subject to textual criticism. Three values of the `type` attribute distinguish them. The `variant` value indicates that the segment of text discusses a reading that differs from the reading in the base text. The `error` value indicates that the editor has erred in establishing the text. Lastly, the `interpolation` value, while also an editorial error, indicates that the passage is interpolated in the commentary by a later redactor.

4 File preparation procedure

Preparation of the base verse file and commentary file would most efficiently be undertaken in two phases. The first phase consists of data-entry in structured text files that number major text divisions, verses, and, in the base text file, lines of verse. A data-entry specialist types the base text and commentary in separate files in the Sanskrit Library phonetic basic (SLP1) encoding. SLP1 is described by Scharf and Hyman (2011) and more briefly by Scharf (2014). Help pages at `https://sanskritlibrary.org/help-text.html` describe the basic characters. A Sanskrit scholar analyzes the sandhi in copies of each of the two files. The TEITagger program, described by Scharf (2018), is used to transform these files to TEI-tagged versions. This program employs a driver file of sets of regular expressions and replacement expressions written based upon an analysis of the text file structure to make the transformation to a file that conforms to the Text-Encoding Intitiative guidelines.

The sandhi-analyzed file of the base text is programmatically converted to a shell of the morphological analysis file by putting each word in a `w` element within its `seg` element that marks its verse quarter and `l` element that marks its verse line within its `lg` element that marks its verse. Each `w` element is provided with `n` and `xml:id` attributes. The `n` attribute numbers each word in sequence from the beginning of the verse. The `xml:id` attribute is formed by adding to the corresponding `xml:id` of the original verse a separator dot followed by a `w` followed by the sequence number. The TEITagger program also can leave `type` and `subtype` attributes with dummy values in the `w` elements. The Sanskrit scholar then just replaces the dummy values with appropriate lexical identifiers and morphological identifiers.

Alternatively, a work-flow can be designed to utilize the Sanskrit Heritage Site's Sanskrit Reader Companion at `https://sanskrit.inria.fr/DICO/reader.fr.html`. A Sanskrit scholar can use the site to filter possible solutions and save the correct analysis in the Corpus Manager described by Huet and Lankri (2018). Software described by Goyal et al. (2012) and utilized by Scharf et al. (2015) could be revised to scrape the HTML pages of these saved morphological analyses. This software converts the Sanskrit Heritage lexical and morphological identifiers to Sanskrit Library tag-sets, loads the analysis into a flexible editing interface that permits corrections and writes the morphological analysis to an XML file. While at present the XML file the software writes does not conform to TEI guidelines, it could be revised to do so.

To create the TEI-tagged version of the commentary requires an expert Sanskrit scholar experienced in reading commentaries and familiar with Sanskrit grammatical, lexical, poetic, and rhetorical texts. To assist such a scholar in inserting the appropriate TEI elements and attributes in a copy of the sandhi-analyzed commentary file, an interface could be created that permits the scholar to select appropriate elements and attributes from menus that contain only the contextually appropriate tags.

5 Details of the data preparation

The verses of the entire text of Bhaṭṭi's *Rāvaṇavadha*, as in the edition of Śāstrin (1912), were typed, tagged and morphologically analyzed in a project entitled *Bhaṭṭikāvya-svādhyā-ya-pariyojanā* at the Rashtriya Sanskrit Sansthan, Ganganath Jha Campus, Prayagraj by Dr. Monali Das under the supervision of Prof. Lalitkumar Tripathi. In that project, Dr. Das manually constructed a prose ordering of words and fed the resulting sentences to the Saṁsā-dhinī morphological analyzer first described by Kulkarni and Shukl (2009) and most recently by Krishnan and Kulkarni (2019). The analyzer provides lexical and inflectional identifiers to words, analyzes compounds and separates preverbs from verbs. However, in most cases the analyzer failed to produce adequate results, and Dr. Das supplied the analysis by hand. Dr. Das typed the commentary of Mallinātha, as in the edition of Trivedi (1898), in WX encoding and then converted it to SLP1 by using Saṁsādhinī transcoding software. She then encoded the text of the commentary of the first canto in XML in accordance with the TEI guidelines under the supervision of Scharf and Ajotikar who made final revisions. Scharf transcoded the morphologically analyzed text to SLP1 using the Sanskrit Library's TranscodeFile software and restored sandhi to compounds by calling his revision of a Java version of sandhi software, originally written by him in Pascal in 1991, from an XSLT transformation. He added morphemic analysis of compound constituents in `m` elements as children of the `w` element containing the compound and provided them with `xml:ids` so that they could be referred to where needed by linking attributes in the commentary file.

6 Issues in tagging the commentary

6.1 Compound constituent analysis

Mallinātha dissolves compounds while commenting on the verse and provides synonyms and derivations of its constituents. For example, he comments on the compound *vibudhasakhaḥ* in the first verse by separating its constituents and commenting on each separately. His comment with our translation is as follows:

> विश्रेषेण बुध्यन्त इति विबुधा विद्वांसो देवाश्च । विबुधौ सुरपण्डितौ इति विश्वः ॥इगुपधलक्षणः कप्रत्ययः ॥ तेषां सखा विबुधसखो विद्वत्सेवी देवेन्द्रसहचरश्च विजयसहकारित्वात्तस्येति भावः ॥ राजाहःसखिभ्यष्टच् इति टच्॥

"Those who especially know are wise (*vibudha*), i.e. knowledgable people (*vidvat*) and deities (*deva*). Viśva says, 'The wise (*vibudha*) are gods (*sura*) and the learned (*paṇḍita*).' The affix **ka** is added by *A.* 3.1.135 *igupadhajñāprīkiraḥ kaḥ*. Their friend (*sakhin*) is called *vibudha-sakha* which means the one who serves knowledgable people (*vidvat-sevin*) and is the companion of Indra, the king of the gods (*devendra-sahacara*), because he helped him obtain victory. The compound-final affix **ṭac** is provided by *A.* 5.4.91 *rājāhaḥsakhibhyaṣṭac*."

Mallinātha begins his comment on the compound *vibudha-sakha* by implicitly dissolving the first constituent, *vibudha*, into its constituents *vi* and *budha* by providing its meaning in the phrase *viśeṣeṇa budhyante* 'Those who especially know'. He then provides two synonyms for the word *vibudha* 'wise', i.e. *vidvat* 'knowledgeable' and *deva* 'deity' and cites the lexicographer Viśva as evidence for these synonyms. He then gives the derivation of the word *vibudha* by the addition of the affix **ka** by citing the rule in Pāṇini's *Aṣṭādhyāyī* that provides the affix after roots with penultimate *i*, *u*, *r*, or *l* as well as after the roots *jñā* 'know', *prī* 'please', and *kr̄* 'skatter'.

He then shows that the compound *vibudhasakha* is a ṣaṣṭhī-tatpuruṣa compound equivalent to the paraphrase *vibudhānāṁ sakhā* by using the pronoun *teṣām* instead of the prior compound constituent itself. He provides two synonyms for the compound *vibudha-sakha*: *vidvatsevin* 'one who serves the knowledgeable' and *devendra-sahacara* 'friend of the king of the gods'. He closes his comment on this word by citing *A.* 5.4.91 which provides the compound-final affix *ṭac*.

This passage in Mallinātha's commentary is encoded utilizing the elements and attributes described in §3 as shown in Figure 1. Since the whole passage concerns the single word *vibudhasakha* in the verse, the entire passage is set in an `ab` element with the attribute-value pair `type='pada'` and `corresp` attribute with the value of the `xml:id` of the word in the morphological analysis file. Segments of the passage that 1. provide an analytic paraphrase (*vigraha-vākya*) of the compound or one of its constituents, 2. give synonyms, 3. cite a lexicon, or 4. provide a derivation are each set in `seg` elements with the value of the `type` attribute indicating its function with the respective keyword: 1. `vigraha`, 2. `synonym`, 3. `lexicon`, 4. `derivation`. For example, *vidvatsevī* and *devendrasahacaraḥ* are put in a `seg` element with the `type` attribute with the value `synonym`. Each of these `seg`s has a `corresp` attribute that bears the value of the `xml:id` of the word or morpheme to which it relates, that is, of the word in the case of the analytic paraphrase of the compound *teṣāṁ sakhā*, and of the morpheme in the cases of the analytic paraphrase, synonym, and lexical citation relating to *vibudha*. The citation from the *Viśvakośa* lexicon is set in the `quote` element. This quote and the references to Pāṇini's *Aṣṭādhyāyī* in the derivations are immediately followed by `note` elements supplied with `bibl` elements indicating the text with `corresp` attributes linking to the bibliographic entry in our TEI file of bibliographic entries, `listbiblfile.xml`. Each of these `bibl` elements has a child `biblScope` element giving the precise reference within the text with a `corresp` attribute with the value of the `xml:id` of the passage cited or referenced. In the case of the *Viśvakośa* of which we do not yet have a digital edition, the `biblScope` and the `corresp` values are dummy values. Finally, the compound word in the text being commented upon (*vibudhasakhaḥ*), which finally appears towards the end of the passage, is set in a `w` element with an `xml:id` matching that of the `w` element in the morphological analysis file.

6.2 Repetition of the base text word in the commentary

Generally Mallinātha cites and explains each word in the base text just once in the commentary; however, occasionally he repeats a word, especially the verb, supplying it again to complete the phrase or clause on which he comments. For example, the sixth verse of the first sarga, the first of three verses describing the city Ayodhyā, consists of a relative clause as follows:

निर्माणादद्वस्य समीहितेषु   सीमेव पद्मासनकौशलस्य।
ऊर्ध्वस्फुरद्रत्नगभस्तिमिर्या   स्थितावहस्येव पुरं मघोनः ॥६॥

"Which (Ayodhyā) stood, as if the apex of the skill of the lotus-seated (Brahmā) who was expert in the manifestation of his desired aims, as if laughing at the city of Indra, with the rays from its jewels scintillating upwards."
"(Ayodhyā) which seemed to be the limit of the expertise of the undertaking of the Creator who was careful in (the art of) creating, and stood laughing, as it were, at the city of Indra, with the rays from its jewels scintillating upwards." (Karandikar and Karandikar, 1982, 2)

The past participle *sthitā* serves as the predicate of the long relative clause that comprises the entire verse. The verse contains two similes, one in each line. Mallinātha cites the word twice in his commentary, once with each simile. His commentary on the end of the first line runs as follows:

सीमा परमावधिः इव स्थिता इत्युत्प्रेक्षा।
"'...stood as if the border, i.e. supreme limit,' is a simile."

Figure 1: TEI encoding of Mallinātha's commentary on the word *vibudhasakhaḥ*

```xml
<ab type="pada" corresp="s1.v1.w3" xml:id="s1.v1.ab2">
 <seg type="vigraha" corresp="s1.v1.w3.m1">viSezeRa buDyanta iti vibuDAH</seg>
 <seg type="synonym" corresp="s1.v1.w3.m1">vidvAMso devASca</seg><pc>.</pc>
 <seg type="lexicon" corresp="s1.v1.w3.m1">
  <quote>vibuDO surapaRqitO</quote>
   iti viSvaH..
  <note>
   <bibl corresp="listbiblfile.xml#xxxxx.visvakosa">{v}iSva. <biblScope
       corresp="xxxxx">xxxxx</biblScope></bibl>
  </note>
 </seg>
 <seg type="derivation" corresp="s1.v1.w3">
  <ref target="A3.1.135">igupaDalakzaRaH kapratyayaH..</ref>
  <note>
   <bibl corresp="listbiblfile.xml#panini.astadhyayi">A. <biblScope
       corresp="A3.1.135">3.1.135</biblScope></bibl>
  </note>
 </seg>
 <seg type="vigraha" corresp="s1.v1.w3">tezAM saKA</seg>
 <w n="1" xml:id="s1.v1.w3">vibuDasaKaH</w>
 <seg type="synonym" corresp="s1.v1.w3">vidvatsevI devendrasahacaraH</seg>
 ca
 <s>vijayasahakAritvAttasyeti BAvaH..</s>
 <seg type="derivation" corresp="s1.v1.w3">
  <ref target="A5.4.91">rAjAhaHsaKiByazwac</ref>
  <note>
   <bibl corresp="listbiblfile.xml#panini.astadhyayi">A. <biblScope
       corresp="A5.4.91">5.4.91</biblScope></bibl>
  </note>
  iti wac..</seg>
</ab>
```

He comments on the end of the second line as follows:

हसित्वा इव स्थिता इति पूर्वोत्प्रेक्षया सजातीयया संकरः ॥

"'…stood as if laughing' is mixed with the similar previous simile."

Both comments include the word *sthitā* which occurs just once in the verse but is syntactically understood in both phrases. Here Bhaṭṭi has employed the familiar linguistic structure called *gapping* in which the verbal predicate is elided in one of two parallel phrases. The commentator has supplied the verbal predicate in each in his explanation. Because we would like to utilize the commentator's reordering of the words of the verse in the base text to evaluate ordinary prose syntax, we would like to determine which of the two occurrences represents the original word and which is supplied. In English syntax, generally the verb occurs in second position, immediately after the subject, and could be elided in subsequent parallel phrases where it could be understood, and could be supplied in the exposition of the complex sentence. For example, Gillon (2019, 196) gives as his first example of gapping, *Peter saw the movie, and Susan _____ the play.*[1] Hence in our translation, in accordance with English syntax, we have placed *stood* in the normal position of the verb in English and gapped it in the second simile. In contrast to English, in Sanskrit the verb typically appears at the end of its clause and would be gapped in the preceding phrase. Thus Bhaṭṭi places the verbal predicate *sthitā* in the fourth verse quarter in the second simile and gaps it in the first. We therefore determine that the second occurrence of the word *sthitā* represents the original word in the verse, and that Mallinātha has supplied the first occurrence to fill the ellipsis in the first simile.

It is significant to understand that each word of the base text that occurs in the commentary should carry unique identification in our encoding of the commentary. We achieve the unique identification of the commentator's use of each word in the base text by providing the unique `xml:id` attribute of the `w` element that contains the word in the morphological analysis file of the base text as the value of the unique `xml:id` of the `w` element that contains the word in the commentary file. The same value is given to the `corresp` attribute of the `ab` element that contains the commentary on the word. Two occurrences of the same word should not carry same value. Obviously we cannot eliminate the second occurrence from the commentary; for we would then lose the information that the commentator supplied the word to fill the gap in the clause. In the case where the commentator supplies a word a second time, we need to select just one as the representation of the original word and encode the other in a different manner to prevent duplication. We achieve this by using the attribute `copyOf` in the `w` element of the repeated word provided with the value of the `xml:id` of that word in the morphological analysis file of the base text.

In the case of the word *sthitā* in the commentary on verse six, we represent that the second occurrence is the original, and the first occurrence is the copy supplied by the commentator as follows:

```
<ab type="pada" corresp="s1.v6.w4">
 <w n="6" xml:id="s1.v6.w4">iva</w>
 <w copyOf="s1.v6.w8">sTitA</w>
</ab>
…
<ab type="pada" corresp="s1.v6.w8">
 <w n="12" xml:id="s1.v6.w8">sTitA</w>
iti
</ab>
```

The first occurrence of the word is not numbered: it occurs in the `w` element with the attribute `copyOf` without any `n` attribute so that it will not find any place in the reordering of the words of the verse in prose order.

Another instance in which Mallinātha repeats a word that occurs just once in the verse occurs in his commentary on verse eighteen of the first sarga. The verse runs as follows:

---

[1]Comma supplied.

135

ऐषीः पुनर्जन्मजयाय यत्त्वं   रूपादिबोधान्न्यवृतच्च यत्ते ।
तत्त्वान्यबुद्धाः प्रतनूनि येन   ध्यानं नृपस्तच्छिवमित्यवादीत् ॥१८॥

"The king said, 'Your meditation, to which you have resorted for victory over rebirth and which turned back from the experience of (the qualities) color, etc., by which you have known subtle entities, is auspicious?"'

"The king spoke thus: I hope that you are free from hindrances in your meditations intended for the conquest of rebirth which cease after the realization of the Ultimate form, and other things, and through which you grasp the subtlest principles." (Karandikar and Karandikar, 1982, 4)

The word *yena*, instrumental singular of the relative pronoun *yad* 'which', occurs just once at the end of the third verse quarter in a relative clause. In his commentary, Mallinātha takes the word twice in two relative clauses, once in connection with the first verse quarter and again in connection with the third. The pronoun refers to the sage's meditation (*dhyāna*), the term for which occurs in the nominative singular in the fourth verse quarter in the main clause of the sentence. In his commentary on the first verse quarter, Mallinātha, reading *tattvam* for *yattvam*, writes,

ते तव येन ध्यानेन पुनर्जन्मनो जयाय निरासाय अपुनर्भवाय तत्त्वमात्मस्वरूपं त्वमैषीरज्ञासीः ।, ...

"By which meditation of you you have gone to, i.e. you have known, the reality, i.e. the nature of the self, ..."

and in his commentary on the third verse quarter he writes,

येन ध्यानेन प्रतनूनि सूद्ममाणि तत्त्वानि महदादीन्यबुद्धा अज्ञासीः ।, ...

"By which meditation you have become aware of, you have known, the fine, i.e. subtle, elements, i.e. the great (*mahat*) etc., ...

While the base text in the edition of Trivedi (1898, 14) reads the neuter accusative singular of the relative pronoun in the first quarter ('to which you have resorted'), the commentary in the same edition lacks a relative pronoun there where Mallinātha reads *tattvam*.[2] Due to the absence of a relative pronoun where needed, Mallinātha borrows the instrumental relative pronoun from the third quarter in his paraphrase of the first quarter ('by which meditation ...you ...have known'). The occurrence in the third quarter is clearly the original, so we determine that Mallinātha's first use is a copy and his second use takes the original. We represent these two as follows as described above.

```
<ab type="pada" corresp="s1.v18.w3">
 <w copyOf="s1.v18.w13">yena</w>
 <seg type="antecedent">DyAnena</seg>
</ab>
…
<ab type="pada" corresp="s1.v18.w13">
 <w n="9" xml:id="s1.v18.w13">yena</w>
 <seg type="antecedent">DyAnena</seg>
</ab>
```

The word that is a copy is put in a `w` element with the `copyOf` attribute and without an `n` attribute. The word that takes the original is put in a `w` element with an `xml:id` and `n` attribute. The former being unnumbered will not occur in our extraction of the commentators reordering of the words of the verse in the syntactic order of ordinary prose; the latter will.

6.3 Intertwined commentary on more than one word

Generally Mallinātha comments on one word at a time and explains aspects of the word in a consistent order. First he provides an analytic phrase, then a synonym, and then a derivation

---

[2]We discuss other cases of variant readings below in §6.5 on page 117.

if all are present, though not all of these three aspects of explanation are always present. He usually provides a synonym (for $251 = 61\%$ of 409 words), a derivation half as often (126/409 $= 31\%$), and an analytic phrase less than a quarter as often ($54/409 = 13\%$). If he provides a synonym based on a thesaurus, the synonym is immediately followed by a citation from the thesaurus. While discussing the derivation of the word in question, the significant rules in the *Aṣṭādhyāyī* are referenced.

However, sometimes he deviates from this procedure and comments on a phrase by providing these aspects of explanation serially for more than one word in the phrase at a time, thereby intertwining the individual words and their explanations. For example, he does so in his commentary on the first two words of the first quarter of verse twenty-five. The verse and its translation are as follows:

प्रयास्यतः पुण्यवनाय जिष्णो   रामस्य रोचिष्णुमुखस्य धृष्णुः ।
त्रैमातुरः कृत्स्नजितास्त्रशस्त्रः   सध्र्यङ्कृतः श्रेयसि लक्ष्मणो ऽभूत् ॥२५॥

"Devoted to (Rāma's) good, the daring Lakṣmaṇa, the son of three mothers, who completely mastered missiles and blades, became the fellow-traveller of the victorious Rāma who had a beaming face and was about to set out for the holy forest."
"Dedicated to (Rāma's) well-being, the daring Lakṣmaṇa, who had three mothers and had acquired all the missiles and weapons, became the fellow-traveller of the victorious Rāma who had a beaming face and had set out for the holy forest." (Karandikar and Karandikar, 1982, 6)

Mallinātha comments on the first two words of the first quarter in reverse order: *puṇyavanāya prayāsyataḥ*, as follows:

ततो ऽनन्तरं पुण्यवनाय तपोवनाय प्रयास्यतः तपोवनं गमिष्यतः॥ गत्यर्थकर्मणि इत्यादिना कर्मणि चतुर्थी॥ लृटः सद्वा इति शत्रादेशः॥

"After that, of the one about to set out for the holy-forest (sacred-forest), (of the one about to go to the sacred-forest). A fourth-triplet nominal termination occurs in the sense of the direct object by *A.* 2.3.12 *gatyarthakarmaṇi dvitīyācaturthyau ceṣṭāyāmanadhvani*. The affix *śatf* occurs by *A.* 3.3.14 *lṛṭaḥ sadvā* (*bhaviṣyati* 3)."

Mallinātha first provides a synonym for each word, glossing *puṇyavanāya* 'holy forest' as *tapovanāya* 'sacred-forest', and *prayāsyataḥ* 'of the one about to set out' as *gamiṣyataḥ 'of the one about to go'*. Then he discusses the derivation of each word. First he explains why the first of these words, *puṇyavana*, is in the dative case and provides the relevant *sūtra* that provides the fourth-triplet nominal termination to denote the direct object (*karman*) of a verb of locomotion. Then he explains that the second word, *prayāsyat*, is a future active particple by citing the rule that replaces the future indicative affix *lṛṭ* with the participle suffix *śatr*. In such cases where the commentary on words is intertwined, we put the commentary on the whole phrase in a single `ab` element and provide its `corresp` attribute with two values matching the values of the `xml:id`s of the two words as shown in Figure 2. The synonym and derivation, and where present the analytic paraphrase, of each of the words is linked to the particular word to which it relates by the correspondence of their `seg` elements' `corresp` attributes with the `xml:id` of the word to which they relate.

6.4 Missing words

Mallinātha comments on almost all of the words in each verse. Of the 409 words in the twenty-seven verses in first sarga, he comments on 369 (90%) by mentioning the word in the commentary explicitly. He comments on another 36 words (9%) by analyzing them without repeating the word, in which case the word may be inferred. On just 4 words (less than 1%), he neglects to comment at all.

Almost all of the inferrable words are compounds where he comments on one of the constituents and repeats the other, or on both of the constituents separately. For example, com-

Figure 2: TEI encoding of Mallinātha's intertwined commentary on words in the phrase *prayā-syataḥ puṇyavanāya*

```xml
<ab type="pada" corresp="s1.v25.w2 s1.v25.w1" xml:id="s1.v25.ab2">tato
    'nantaraM
 <w n="1" xml:id="s1.v25.w2">puRyavanAya</w>
 <seg type="synonym" corresp="s1.v25.w2">tapovanAya</seg>
 <w n="2" xml:id="s1.v25.w1">prayAsyataH</w>
  tapovanaM
 <seg type="synonym" corresp="s1.v25.w1">gamizyataH</seg><pc>..</pc>
 <seg type="derivation" corresp="s1.v25.w2">
  <ref target="A2.3.12">gatyarTakarmaRi</ref>
  <note>
   <bibl corresp="listbiblfile.xml#panini.astadhyayi">A. <biblScope
       corresp="A2.3.12">2.3.12</biblScope></bibl>
  </note>
  ityAdinA karmaRi caturTI..
  <ref target="A3.3.14">lfwaH sadvA</ref>
  <note>
   <bibl corresp="listbiblfile.xml#panini.astadhyayi">A. <biblScope
       corresp="A3.3.14">3.3.14</biblScope></bibl>
  </note>
  iti SatrAdeSaH..</seg>
</ab>
```

menting on the word *śrutānvitaḥ* 'endowed with Vedic knowledge', an adjective describing king Daśaratha in the second quarter of the first verse, Mallinātha writes,

श्रुतेन वेदशास्त्रश्रवणेन अन्वितः स्वयं च विद्वानित्यर्थः।

"Endowed with learning (learning of the science of Veda)."

Here, Mallinātha does not mention the compound *śrutānvitaḥ*. Yet he provides an analytic paraphrase of the compound, with the words *śrutena* 'with that which is heard' ...*anvitaḥ* 'endowed', which shows that it is a trtīyā-tatpuruṣa compound. Immediately after the initial compound constituent, he provides the synonym *vedaśāstraśravaṇena* 'with learning (literally 'hearing') of the science of Veda'. As mentioned in §3 on page 107, in cases where the word can be inferred from his analysis, we supply the word in a `w` element with a `type` attribute with the value `inferred`.

There are several instances where the word in the base text can be inferred by syntactic criteria. For example, the first line of verse twenty-two contains a relative clause in the second quarter coordinated with the main clause in the first quarter by the use of the instrumental singular of coordinated masculine singular demonstrative and relative pronouns *tena* ...*yena*. The line reads as follows:

घानिष्यते तेन महान् विपक्षः   स्थायिष्यते येन रणे पुरस्तात्।

"A great adversary will be slain by him, by whom standing will be done in front in battle."

"By him, who will stand at the vanguard in the battlefield, (even a) great adversary will be killed." (Karandikar and Karandikar, 1982, 5)

Mallinātha omits the relative pronoun `yena` in his commentary on the second quarter. He writes:

रामेण रणे पुरस्तादग्रे स्थायिष्यते स्थास्यते

"By Rāma standing will be done in front (at the fore) in battle."

The appropriate position of the relative pronoun *yena* can be inferred in the commentary, despite Mallinātha's omitting it, by his use of its antecedent *rāmeṇa*. We place the pronoun immediately preceding the antecedent in a `w` element with the attribute-value pair `type="inferred"`.

The few words on which Mallinātha neglects to comment entirely are mostly pronouns or particles, for example, the relative pronoun *yā* at the end of the third quarter of verse six on page 111, and the conjunction *ca* in the second quarter of verse eighteen on page 113. Unusually, he neglects to include the word *rataḥ* 'dedicated' in the fourth quarter of verse twenty-five on page 115. The fact that it is so unusual for Mallinātha to neglect to comment on a word is contributory evidence for understanding his omission of the relative pronoun *yat* 'which' and the second-person singular pronoun *tvam* 'you', which occur at the end of the first quarter of verse eighteen, as due to reading instead *tattvam* 'reality' as remarked in §6.2 on page 114. In order to extract from the commentary the words of the verse transformed into prose order in the commentary, it is necessary to represent every word that occurs in the base text because the transformed syntax would be incomplete if it lacked even a single word that appears in the base text. In this case, we supply the word at the proper place. As mentioned in §3 on page 108, where Mallinātha neglects to comment entirely on a word, we supply the word in a `w` element with a `type` attribute with the value `supplied`. For example, for the pronoun *yā* in verse 6:

```
<w type="supplied" n="1" xml:id="s1.v6.w7">yA</w>.
```

We place this supplied `w` element at the beginning of its clause because Mallinātha has similarly placed other pronominals at the beginning of their clauses in other verses.

6.5 Recording variants

Editions of classical Sanskrit texts vary, critical editions record variant readings, and commentators often record variants known to them. The text of Mallinātha's commentary on the *Rāvaṇavadha* occasionally reveals that he had a version of the base text that differs from the base text in the same edition (Trivedi, 1898). Moreover both occasionally differ from the base text as in the edition Das used for the morphological analysis (Śāstrin, 1912). In addition, Mallinātha records variant readings known to him. As noted by Ajotikar et al. (2018), the variants used and mentioned in commentary are crucial information from the point of view of text-transmission because they usually predate the oldest manuscripts. Hence we take care to record these variants. To do so, we use the attribute-value pair `type="variant"` to identify each variant present in Mallinātha's commentary. Most of these variants occur where Mallinātha simply uses a different word from the word in the base text. For example, in his commentary on the third verse, he uses a different verb from the one that occurs at the end of the first quarter of the verse in both the above-mentioned editions (Trivedi, 1898; Śāstrin, 1912) and in the edition with text and translation by Karandikar and Karandikar (1982). The editions read:

वसूनि तोयं घनवद्व्यकारीत्
He scattered riches just as a cloud does water.

The first quarter of the verse in all three editions has the word *vyakārīt*, aorist third-person singular of the verbal root *kṛ* with the preverb *vi* meaning scatter. In contrast, Mallinātha records the reading of the word *vyatārīt* which is the same form with the same preverb but with a different root, i.e. *tṛ*, meaning bestow. We encode this variant by putting the attribute-value pair `type="variant"` in the `w` element as follows:

```
<ab type="pada" coresp="s1.v3.w4">
  <w n="5" type="variant" xml:id="s1.v3.w4">vyatArIt</w>
</ab>
```

Besides simple cases such as the above where the commentator has a different reading from the editions of the base text, there are cases where the commentator remarks on a variant along with the reading given in the edition of the base text. For example, in the last quarter of verse eleven, the edition with Mallinātha's commentary edited by Trivedi (1898, 9) reads

स कर्मठः कर्म सुतानुबन्धि

"Expert in his work, he (...performed) the sacrifice that would result in sons."
(Karandikar and Karandikar, 1982, 3)

After commenting on the last word, *sutānubandhi*, in the commentary in the same edition, Mallinātha mentions and comments on the variant *sutānubandham*, which is the reading adopted in the base text both by Śāstrin (1912, 9) and by Karandikar and Karandikar (1982, 3), Mallinātha writes:

सुतानन्बध्नाति सन्तनोतीति सुतानुबन्धमिति वा पाठान्तरम्। कर्मण्यण्।

In this case, after Mallinātha mentions the reading *sutānubandhi* in the base text and provides a derivation for it, including a synonym for the morpheme *sutānubandha*, he proceeds to comment on the variant (*pāṭhāntara*) *sutānubandham*, providing it with the analytic paraphrase *sutān anubadhnāti*, a synonym *santanoti*, and a derivation *karmaṇyaṇ*. The situation complicates the encoding because instead of there being segments of the type synonym, vigraha, and derivation associated with one word, there are two sets of segments of these kinds associated with two words: the adopted reading and the variant, both associated with a single word in the base text. To encode this situation, we incorporate the entire section in a single ab associated with the word in the base text by its corresp attribute with the value of the xml:id of the word in the morphological analysis file as usual, but we segregate all of the information about the variant in a single subordinate seg element with the attribute-value pair type="variant". The variant word is put in a w element with the attribute-value pair type="variant" and an xml:id whose value is the xml:id of the word in the base text with the suffix .var added to it. The segments that provide an analytic paraphrase, synonym, and derivation of the variant reading are encoded in subordinate seg elements with the appropriate type attributes, and with corresp attributes with the value of the variant word's xml:id. The w element containing the variant and its synonym segment, vigraha segment and derivation segment are all child nodes of the variant segment as shown in Figure 3.

Figure 3: TEI encoding of Mallinātha's commentary on the variant *sutānubandham*

```
<seg type="variant">
 <seg type="vigraha">sutAn anubaDnAti
  <seg type="synonym" corresp="s1.v11.w16.var">santanoti</seg>
   iti
  <w type="variant" n="14" xml:id="s1.v11.w16.var">sutAnubanDam</w>
   iti vA pAWAntaram..
  <seg type="derivation" corresp="s1.v11.w16.var">
   <ref target="A3.2.1">karmaRyaR..</ref>
   <note>
    <bibl corresp="listbiblfile.xml#panini.astadhyayi">A. <biblScope
       corresp="A3.2.1">3.2.1</biblScope></bibl>
   </note>
  </seg>
 </seg>
</seg>
```

7 Interpolation

Trivedi's (1898, 17) edition of Mallinātha's commentary on verse twenty-three includes a portion which cannot possibly have been composed by Mallinātha. The portion reads as follows:

कौमुद्यां तु ततो ऽनुजझ इति पाठमनुसृत्य अनुपसर्गाज्झः इति शास्त्रप्रातिकूल्यमुद्भाव्य नृप इत्यत्र नृपेणोति विपरिणामं कृत्वा दोषपरिहारः कृतः।

> "In the kaumudī, after adopting the reading *anujajñe*, and stating that this form violates *A.* 1.3.76 *anupasargājjñaḥ*, the fault is rectified by changing the reading *nrpaḥ* to *nrpeṇa*."

The passage refers to Bhaṭṭoji Dhīkṣita's citation of the second half of verse twenty-three of the *Rāvaṇavadha*, as it appears in the editions of Trivedi (1898, 17), Śāstrin (1912, 16) and Karandikar and Karandikar (1982, 5), in his comment on *A.* 1.3.76. The comment including the citation runs as follows:

कथं तर्हि भट्टिः
इत्थं नृपः पूर्वमवालुलोचे ततो ऽनुजज्ञे गमनं सुतस्य॥

इति। कर्मणि लिट्। नृपेणेति विपरिणामः॥ Then how does Bhaṭṭi write,
"The king previously thought thus, then he approved his son's going."
"Thus did the king reflect first and then permit the departure of his son." (Karandikar and Karandikar, 1982, 5)
The verbal termination of the perfect (textit**liṭ**) occurs to denote the direct object.
There is alteration of the nominal termination 'by the king (*nrpeṇa*).

A. 1.3.76 provides that an ātmanepada verbal termination occurs after the verb *jñā* 'know' without a preverb if the result of the action accrues to the agent. The rule accounts for the use of *jñā* in the middle voice only without a preverb not with the preverb *anu* as it occurs in the fourth quarter of verse 23. Bhaṭṭoji Dhīkṣita justifies the form by claiming that the form *anujajñe* is passive rather than middle and that its agent, which appears in the nominative (*nrpaḥ*) must be understood to be an alteration of the instrumental (*nrpeṇa*). The passage in the edition of Mallinātha's commentary explains Bhaṭṭoji Dhīkṣita's justification.

While the passage makes a correct observation, it must be a later interpolation. For it is impossible that Mallinātha would quote Bhaṭṭojī Dīkṣita's *Siddhāntakaumudī* because Bhaṭṭojī Dīkṣita (1547–1633 CE Dash (2007, 111b)) is dated a two centuries after Mallinātha (14th century CE (Lienhard, 1984, 39)). Trivedi (1898, Notes 20) leaves the reader in doubt on this point when he comments in his note on verse 23, "But the true reading seems to be the one adopted in the text. Malli. notices it and shews how it is discussed in the *Siddhāntakaumudī*," and allows the reader to be misled by including it in the text of his edition. However, after discussing the identity and date of Mallinātha in detail in his introduction, he (1898, xxviii) concludes regarding this and one other citation from the *siddhāntakaumudī*, "they must be considered as interpolations. They may have been originally marginal notes of some copyist and afterwards inserted into the commentary." Further down the page he notes that the other passage that cites the *Siddhāntakaumudī* is omitted in a manuscript he obtained from the Theosophical Society of Madras, and continues regarding the citation under discussion, "As regards the second reference to the kaumudī (C. I. St. 23) I have secured, ..., a copy of the commentary of Mallinātha from another manuscript at Udamalpet, in which the reference to the kaumudī is omitted." Finally, he (1898, xxviii–xxix) concludes his introduction writing, "I therefore conclude that ...he flourished before Bhaṭṭojī Dīkṣita and that the two references to the Siddhāntakaumudī in the commentary are interpolations." To mark the text as an interpolation, we tag this passage with the attribute-value pair `type="interpolation"`.

8 Utility of the TEI-tagged data

The analysis and tagging of major Sanskrit poetic works and their commentaries in the manner described above leads to a number of useful applications. Several such applications are described below.

- **Dynamic lemmatized index.** In the morphological analysis file described in §3 on page 107, the root of each verb and nominal base of each nominal is provided, as are the word's lexical and morphological identifiers. From this information dynamic indices

can be created that can provide access to words and their contexts based on a number of criteria. These indices include an alphabetical list of words (*pada*s) and an alphabetical list of lemmas (roots and nominal bases). These lists could be trimmed according to the selection of additional criteria such as various lexical and morphological categories, and text extents.

- **Dictionary of synonyms (*paryāyakośa*).** The commentary provides the synonym of each word in the base text as described in §2 on page 107, and the TEI-tagged commentary file tags these synonyms as described in §3 on page 107. From this information one can programmatically generate a dictionary of synonyms.

- **Index of quotations.** The commentary cites a number of different texts in a variety of disciplines as discussed in §2 on page 107. The highest quality scholarly editions of Sanskrit works enrich their editions with lists of quotations through time-consuming and tedious labor. Many editions lack such useful lists. The TEI-tagged commentary file tags quotations as described in §3 on page 108. From this information one can programmatically generate an alphabetical or searchable index of quotations very easily.

- **Index of derivations.** As described in §2 on page 107, commentaries provide the Pāṇinian derivation of numerous words. This information is significant for Sanskrit readers. Yet it would also be extremely valuable for linguists if systematized. The TEI-tagged commentary file tags such derivations as described in §3 on page 108. One can programmatically generate an index of derivations that provides access to them by the word derived, roots and stems, affixes, and the numbers of sūtras referenced.

- **Index of meters.** Although commentaries identify the metrical patterns commonly found in each chapter of poetic texts as described in §2 on page 107 and these will be tagged in our analysis as described in §3 on page 108, they do not provide the exact sub-type or description of the metrical pattern of each verse. However, TEITagger checks the metrical pattern of each verse in the process of creating the TEI-tagged XML file from the original text file of the base text as mentioned in §5 on page 110 and inserts the specific meter name in the `type` attribute of the `lg` element which houses each verse. TEITagger also identifies the class and general type of the meter, and the metrical pattern of the verse as described by Melnad et al. (2015a), Melnad et al. (2015b) and Scharf (2018), and discovers aberrant cases. From this information, one can programmatically generate an alphabetical or searchable index of meters used in the text.

- **Index of figures of speech.** Commentaries identify figures of speech used in each verse as described in §2 on page 107 and these will be tagged as described in §3 on page 108. From this information, one can programmatically generate an alphabetical or searchable index of figures of speech used in the text.

- **Preparation of a tree-bank of Sanskrit prose.** The file containing the prose sentences in the commentator's reordering of the verse should prove to be extremely useful to those studying Sanskrit syntax. Kulkarni et al. (2015) recently explored the limitations of free word order in Sanskrit, and Scharf et al. (2015) examined differences in Sanskrit prose versus verse word order. The prose word order file links every word with the morphological analysis file where every word is supplied with its lemma, and lexical and inflectional identifiers. Adding relational information to this information could form the basis of a tree-bank of Sanskrit prose.

Scharf wrote extensible stylesheet language transformations (XSLT) that prepare a dictionary of synonyms and derivations, a word index, an index of citations, and indices of passages in which Mallinātha discusses meters and figures of speech. The resultant XML files conform to the TEI

guidelines for dictionaries and indices and can easily be utilized to create Web-interfaces for each. Figure 4 shows an entry for the verb *vyatārīt* which occurs in Mallinātha's commentary on the first quarter of verse three as discussed above in §6.5 on page 117. The `etym` element houses the derivation provided by the commentator, and the `def` element, as a child of the `sense` element houses the synomym he provides. The additional `bibl` elements not copied from references in the commentary itself provide links back to the text of the commentary.

Figure 4: TEI encoding of the dictionary entry for the verb *vyatārīt*

```
<entry>
 <form>vyatArIt</form>
 <etym>taraterluNi
  <xr corresp="A8.2.28">iwa Iwi</xr>
  <note>
   <bibl corresp="listbiblfile.xml#panini.astadhyayi">A.<biblScope
       corresp="A8.2.28">8.2.28</biblScope></bibl>
  </note>
   iti sijlopaH.
  <bibl>BhRV.<biblScope corresp="s1.v3.w4"/></bibl>
 </etym>
 <sense>
  <def>vitIrRavAn
   <bibl>BhRV.<biblScope corresp="s1.v3.w4"/></bibl>
  </def>
 </sense>
</entry>
```

9 Syntactic research

Higher-level encoded text can be used to resolve pertinent issues of research. While the encoding of text-structure and document-format in digital text is useful for creating a variety of print and HTML publications from a single document source, higher-level encoding can efficiently be used to investigate various research problems. We plan to use our encoding of Mallinātha's commentary on the first sarga of the *Rāvaṇavadha* to explore his reordering of the words of the verse as evidence of the unmarked syntax of ordinary prose. To this end, Scharf wrote a pair of XSLT transformations that extract each word of each verse in the order in which Mallinātha comments on it and constructs prose renderings of each verse. One version of the transformation preserves the numbering of the order of the words in the original verse and information on whether we inferred or supplied the word; the other omits this information providing a readable prose rendering of the words of each verse. We plan to subject the original verse and Mallinātha's prose to the sort of syntactic analysis conducted by Scharf et al. (2015).

At present the twenty-seven verses of the first sarga of Bhaṭṭi's *Rāvaṇavadha* does not comprise a significant data-set. Tripathi's project described in §5 on page 110 will extend the data-set by a factor of ten. Moreover, Scharf and Tripathi have begun a two-year project entitled, "Digitization and markup of the masterpieces of Sanskrit kāvya and their most essential commentaries." The project will produce TEI editions of the base text and commentaries of five major and minor poetic works. We hope to present some interesting results bearing on Sanskrit syntax in a couple of years.

10 Conclusion

The digital medium provides methods to find information quickly, and contemporary scholars expect to be able to locate the information they seek using these methods. Developments in

digital humanities have made it possible to ease the process of making very specific information embedded in texts available in the way that contemporary scholars expect. To make this information available requires careful analysis of the structure and content of the concerned texts. We have analyzed the structure and content of Sanskrit commentarial literature on poetry which provides rich information in the form of references, explanation and derivation of specialized vocabulary, explanation of metrical patterns, syntax, figures of speech, etc. as described above. We have devised templates in accordance with the TEI guidelines to represent all of the relevant information in poetic texts and their commentaries, have written XSLT transformations that prepare dictionaries of and indices to this information, and have outlined procedures to build a library of digital Sanskrit commentarial literature. The employment of these procedures and construction of this library should prove to be invaluable to linguists, philologists, and other humanities scholars.

References

Tanuja P. Ajotikar and Peter M. Scharf. 2020. Development of a tei standard for digital sanskrit texts containing commentaries. In *Proceedings of the international conference on the tradition of commentary and the stream of knowledge (bhāṣyaparamparā jñānapravāhaśca)*. Somnath Sanskrit University.

Tanuja P. Ajotikar, Anuja P. Ajotikar, and Peter M. Scharf. 2018. Enriching the digital edition of the *Kāśikāvrtti* by adding variants from the *Nyāsa* and *Padamañjarī*. In Huet and Kulkarni (hue, 2018), pages 277–294.

Siniruddha Dash, editor. 2007. *New catalogus catalogorum*, volume 16. University of Madras.

Brendan S. Gillon. 2019. *Natural language semantics*. MIT Press.

Pawan Goyal, Gérard Huet, Amba Kulkarni, Peter Scharf, and Ralph Bunker. 2012. A distributed platform for sanskrit processing. In *Proceedings of COLING 2012*, pages 1011–1028. The COLING 2012 Organizing Committee.

2018. Computational sanskrit and digital humanities.

Gérard Huet and Idir Lankri. 2018. Preliminary design of a sanskrit corpus manager. In Huet and Kulkarni (hue, 2018), pages 259–276.

Maheshvara Ananta Karandikar and Shailaja Karandikar. 1982. *Bhaṭṭikāvyam*. Motilal Banarasidas.

Sriram Krishnan and Amba Kulkarni. 2019. Sanskrit segmentation revisited. In *ICON 2019*. NLP Association of India.

Amba Kulkarni and Devanand Shukl. 2009. Sanskrit morphological analyser. volume 70, pages 169–177.

Amba Kulkarni, Preeti Shukla, Pavankumar Satuluri, and Devanand Shukl. 2015. How free is 'free' word order in sanskrit? In Scharf (sch, 2015), pages 269–304.

Siegfried Lienhard. 1984. *A history of classical poetry*. Number III.1 in A history of Indian literature. Otto Harrasssowitz.

Keshav Melnad, Pawan Goyal, and Peter M. Scharf. 2015a. Identification of meter in sanskrit verse. In Scharf (sch, 2015), pages 325–346.

Keshav Melnad, Pawan Goyal, and Peter M. Scharf. 2015b. Updating meter identifying tool (mit). Paper presented at the 16th World Sanskrit Conference, Bankok.

Yonghui Qiu, Bangwei Wang, George Cardona, Hong Luo, and Peter M. Scharf, editors. 2020. *Saṁskr-tādhvā*.

Vināyaka Śāstrin, editor. 1912. *The Bhaṭṭikāvya of Bhaṭṭi*. Nirṇaya-sāgara Press, 4 edition.

2015. Sanskrit syntax.

Peter M. Scharf and Malcolm D. Hyman. 2011. *Linguistic issues in encoding Sanskrit*. The Sanskrit Library.

Peter M. Scharf, Anuja Ajotikar, Sampada Savardekar, and Pawan Goyal. 2015. Distinctive features of poetic syntax. In Scharf (sch, 2015), pages 305–324.

Peter M. Scharf. 2014. Linguistic issues and intelligent technological solutions in encoding sanskrit. 16(3):15–29.

Peter M. Scharf. 2015. Providing access to manuscripts in the digital age. In Justin Thomas McDaniel and Lynn Ransom, editors, *From Mulberry leaves to silk scrolls*, number 1 in The Lawrence J. Schoenberg Studies in Manuscript Culture, pages 231–271. University of Pennsylvania Libraries.

Peter M. Scharf. 2017. Accessing manuscripts in the digital age. volume 6, pages 131–173. National Mission for Manuscripts.

Peter M. Scharf. 2018. Teitagger. In Huet and Kulkarni (hue, 2018), pages 169–191.

Peter M. Scharf. 2020. Issues in digital sanskrit philology and computational linguistics. In Qiu et al. (Qiu et al., 2020), page xxxxx.

Kamalāśaṅkara Prāṇaśaṅkara Trivedi, editor. 1898. *Rāvaṇavadha*. Government Central Book Depot.