



Vedic Accent and Lexicography

Felix Rau

University of Cologne – Lazarus Project

Vedic Accent and Lexicography
Lazarus Project (Cologne Sanskrit Lexicon) Project Documentation 2

 This work is licensed under the Creative Commons Attribution 4.0 International License.

Lazarus Project (Cologne Sanskrit Lexicon)
University of Cologne
<http://www.cceh.uni-koeln.de/lazarus>
<http://www.sanskrit-lexicon.uni-koeln.de/>

1 Introduction

This paper is a preliminary investigation into the problems the representation of the accents of Vedic Sanskrit poses to Sanskrit lexicography. The purpose is to assess the principles applied in various lexicographic works in the representation of Vedic accents and its relation to the underlying linguistic category as well as traditions of accent marking in different texts. Since the focus is on Sanskrit lexicography, we ignore the complexity of accent marking in manuscripts and the diversity of accent marking across different Indic scripts that were used to write Sanskrit over the ages. We will restrict ourselves to accent marking in Devanagari and Latin script, as these two are the relevant systems for virtually all of modern philological Sanskrit lexicography.

The complexity of accent marking in Vedic Sanskrit derives from several facts. Besides the complexity of the linguistic phenomenon itself [see][among others]Kiparsky1973, the complexity arises from the fact that different textual or editorial traditions employ structurally different systems for marking Vedic accent. The situation is further complicated by the fact that the limited inventory of accent markers can have multiple functions in a marking system or across different systems.

Accent marking in lexicography is particularly complex since accent marking is manuscript, edition, text, and script specific in Vedic Sanskrit, while lexicographic work abstracts away from these particularities and concentrates information about lexicological units in very dense textual units. These lexical entries contain various types of information. This includes phonological information about the prosodic structure of the lexical unit, in particular the location of the main lexical accent, but also examples of usage taken from various textual sources with structurally different systems to mark accents and different typographic devices employed in these systems.

1.1 The Accent of Vedic Sanskrit

Vedic Sanskrit is generally analysed to have a free pitch accent – Kiparsky1973, Lubotsky1988 among many others. In the South Asian grammatical tradition pitch accent or svara is described using the categories udātta, anudātta and svarita. Since traditional as well as modern accent marking systems for Vedic Sanskrit are generally phrased using these three categories as the main reference points, they will be central to our discussion.

The udātta (Skt. ‘elevated, raised’) accent is generally associated with high pitch, while anudātta (Skt. ‘non-elevated, non-raised’) is associated with low or unmarked pitch, while svarita (Monier-Williams gives ‘soundied’) is seen as falling pitch [p. 27]Whitney1869. Kiparsky1973 analyses the accent in more detail and with a focus on diachronic changes. The udātta accent is generally analysed as the actual pitch accent in the western tradition, with the anudātta accents being seen as the unmarked syllables and svarita as a secondary phenomenon occurring in the syllable following the primary accent. However, there is a category of svarita accents, called independent svarita (jātyasvarita or nityasvarita) which occur without a preceding udātta accent. This type of svarita accents arises

when the udātta syllable preceding a svarita syllable is lost, normally when the syllable bearing vowel is reduced to a glide. Independent svaritas cannot be predicted and thus need to be treated similar to udātta accents. The predictable svarita accents are called dependent or enclitic svaritas.

2 Systems of Marking Vedic Accent

There is a large number of systems to mark the pitch accent of Vedic Sanskrit. Only a very restricted subset of systems relevant for the representation of Vedic accent in lexicographic works will be discussed here. Witzel1974 describes the huge variety of systems found in prints and manuscripts of Vedic Sanskrit texts and classifies these systems according to whether they explicitly mark the primary or *udātta* accent or not. Scharf2007 presents a description of a variety of marking systems taken and adapted from MYV1964.

It has to be noted that several of the systems for marking accent in Vedic Sanskrit represent phonetic-prosodic subtleties that go beyond the already very elaborate three accent system of traditional Sanskrit grammar theory. These subtleties seem to derive from particular phrasal or other syntagmatic arrangements in specific passages and are not incorporated in any lexicographic work.

2.1 R̥gveda System

The R̥gveda accent marking system is used in the text tradition of the R̥gveda, the Atharvaveda, and several other text traditions [see][for more information]Scharf2007. It is probably the best known accent marking system in Devanagari script. The system is peculiar from a linguistic point of view in that it explicitly marks all svarita accents – whether independent or not – and some anudātta accents, but not the primary accent, the *udātta*. The diacritic used to mark svarita accent is a vertical stroke over the central glyph of the syllable क.¹ Anudātta is marked by a diacritic underscore क्.

While in principle all svarita accents are marked, only the anudātta accents preceding the first *udātta* or independent svarita of a stanza are always marked. After the first *udātta* or independent svarita only an anudātta immediately preceding an *udātta* is marked. A syllable that follows an *udātta* accent is marked as anudātta, if it is again followed by an *udātta* accent. While it would be marked as svarita accent otherwise.² [p. 449]Macdonell1916 elaborates the peculiarities of this accent marking system and gives examples for its application.

In the following, the first stanza of the R̥gveda (RV 1.1.1) is given in Devanagari with accents marked according to the rules of the R̥gveda accent marking system and with the two diacritics used in this system. The line (1b) gives the same stanza in Roman transliteration, but with accents marked according to the R̥gveda accent marking system and diacritics similar to the ones used in Devanagari. Line (1c) gives the stanza in indological (ISO 15919) transliteration with the *udātta* accents marked.

¹We are following the convention of presenting the Devanagari diacritics in relation to the syllable क ka.

²[p. 449]Macdonell1916 illustrates this with the passage RV 1.1.6 तवेत्सूत्यम् tavet tat s̄atyam (tāvēt tāt satyām), in which the syllable स sa would be marked as svarita, if the following syllable would not be an *udātta* accent. However according to [p. 449]Macdonell1916 the enclitic svarita accent “is ousted by the Anudātta, which is required to indicate that the following syllable tyam has the *Udātta*.”

अग्निमीळे पुरोहितं यज्ञस्य देवमूत्तिविजम् । होतारं रत्नधातमम् ॥ agnim īle purohitam yajñasya de-
vam ḥtvijām । hotāram ratnadhātāmam ॥ agním īle puróhitam yajñásya devám ḥtvíjam ।
hótāram ratnadháttamam ॥

2.2 Sāmaveda System

The Sāmaveda accent marking system is used in the Sāmaveda. It is not discussed by Witzel1974 on the grounds that it is “based on musical reproduction of the texts” [p. 473]Witzel1974. The system itself is quite complex, but the crucial part for our purposes here is, that it explicitly marks udātta, anudātta, and svarita accent with small Devanagari numerals above the respective syllables.

१ ५ ३ ६ २ ३ १ ५ ३ १ ५
तेषां वो अग्निनुन्नानामिंद्रो हंतु वर्वरम् ॥ २ ॥

Figure 2.1: Accent marking in the Sāmaveda

The basic system is straightforward, but the whole system is too elaborate to be discussed here in detail. [p. 19]Scharf2007 describes the system in more detail. This marking system is based on the following principles: The Devanagari numeral one १ indicates udātta accent क, the numeral two २ marks svarita accents क, and a numeral three ३ above a syllable indicates anudātta accent क. Figure 2.1 gives a passage from the Sāmaveda Saṃhitā Kauthuma recension (SV 4.9.3.8.2) with the accents marked according to the Sāmaveda accent marking system. The text is repeated in (2a) and the same text is given in Roman transliteration in (2b), but with accents marked according to the Sāmaveda accent marking system and with acute accent for udātta and grave accent for all svarita accents (dependent). Example (2c) gives the same text with udātta accents according to ISO 15919.

तेषां वो अग्निनुन्नानामिंद्रो हंतु वर्वराम् téṣāṁ vo agnínūnnānāmíndrō hantu váràṁvaram téṣāṁ
vo agnínunnānāmíndro hantu váramvaram

[p. 228-229]Howard1986³ describes the Sāmaveda system as follows:

Numbers 1 [१] and 3 [३] always represent udātta and anudātta, respectively. Number 2 [२] indicates svarita, but it denotes also an udātta syllable followed by anudātta. When two or more udātta syllables appear in succession, only the first is marked with 1 [१], but the sign 2r [२२] is placed above the following svarita. If, however, an anudātta follows, 2u [२३] is placed above the first udātta syllable and the rest are left undesignated. In a series of anudātta syllables at the beginning of the line, only the first is marked with 3 [३]. An independent svarita has the sign 2r [२२], and the preceding anudātta is marked

³Taken from [p. 3]EversonScharf2007.

3k [३क]. Pracaya⁴ syllables have no markings. [Addition of Devanagari characters mine F.R.]

2.3 Śatapatha Brāhmaṇa System

The accent marking system of the Śatapatha Brāhmaṇa is very minimalistic and exclusively utilises the diacritical underscore ्. Despite the minimal inventory of accent markers, researchers have interpreted the system in different ways. [p. 88]Whitney1889 dismisses it as “scanty and imperfect” and does not discuss it at all. [p. 451]Macdonell1916 interprets the diacritic underscore as a marker for udātta accent in the syllable it is placed under. In case of more than one successive udātta accent only the last syllable is marked. In this interpretation, the Śatapatha Brāhmaṇa system marks exclusively udātta accent and an independent svarita is indicated by marking the preceding syllable as udātta accent. Hoffmann1956, [English translation in][p. 475, n17]Witzel1974, reinterprets the system and considers the diacritical underscore ् a marker for svarita accents, which is placed on the syllable preceding any svarita accent, enclitic or independent. Hoffmann’s interpretation of the Śatapatha Brāhmaṇa system is currently the generally accepted view. The following passage from the Śatapatha Brāhmaṇa Weber1849 demonstrates the marking system.⁵

Figure 2.2 and ?? give two passages from the Śatapatha Brāhmaṇa. The highlighted passages can also be found in part as quotations in pwg and are repeated in example (3) and (4). They are given in Devanagari with the accents marked according to the Śatapatha Brāhmaṇa system in (3a) and (4a), a transliteration mimicking the Śatapatha Brāhmaṇa system in placing the diacritical underscore in the positions found in the Devanagari original (3b) and (4b), and in (3c) and (4c) transcriptions according to ISO 15919 are given.

नऽनाम तुतो विमुच्चामेति यतो क्षेव युज्जन्ति तुतो विमुच्चन्ति ॥८॥ तस्य वा उत्स्यानसः । अग्निदग्धमिवैषां ब्रह्मवृहं भवत्यथ यज्ञाधनेन कस्तम्भीं प्रातङ्गं वेदिरेवास्य सा नीडु एव हर्विर्धानम् ॥९॥

Figure 2.2: Accent marking in the Śatapatha Brāhmaṇa

अग्निदग्धमिवैषां ब्रह्मवृहं भवति agnīdagdhamivaiṣāṁ vāham bhavati agnídagdhamivaiṣām váham bhavati

तानक्षण्या सुन्तुन्दन्ति युद्यक्षण्या न शक्नुयादुपि समीचः tānakṣṇayā samṝndanti yādyakṣṇayā nā śaknuyādāpi samīcaḥ tānakṣṇayā sámṝndanti yādyakṣṇayā nā śaknuyādāpi samīcaḥ

⁴The name *pracaya* is given to a sequence of anudātta syllables which are described as having udātta-like high pitch. See Whitney1869 for a description of *pracaya* accent.

⁵The diacritical double underscores found in Weber1849 were introduced by him, to distinguish the underscore he interpreted as independent svarita markers from the underscores he interpreted as udātta markers [] [p. xii-xiii]Weber1849. The double underscores indicate independent svaritas and are not part of the traditional Śatapatha Brāhmaṇa system, in which these syllables would also feature a simple underscore.

त्विरत्युत्कृत्यां किरामीत्यत उद्धवति तत्कृत्यामुत्किरति ॥१२॥ तान्बादुमात्रा-
 न्खनेत् । अतो वाऽश्वोऽत्तेनैवितत्कृत्यां मोहयति तानक्षण्या संतृन्दत्ति ध्य-
 क्षण्या न शक्त्यादपि समीचस्तस्मादिमे प्राणाः परः संतृष्टाः ॥१३॥ तान्यथाखा-

Figure 2.3: Accent marking in the Śatapatha Brāhmaṇa

2.4 Böhtlingk & Roth System

The Böhtlingk & Roth system is used in the große Petersburger Sanskrit-Wörterbuch pwg and the Sanskrit-Wörterbuch in kürzerer Fassung pw. This system is also used by Whitney1889. Originally developed for pwg, Böhtlingk and Roth follow indological conventions to mark udātta and independent svarita. They represent Sanskrit in Devanagari and there is no traditional system marking accent according to these principles that is used in printing Devanagari and as far as we know also not in Devanagari manuscripts. To mark Vedic accent according to this system, they introduce a new device, a small, raised ॐ (Devanagari short u) above the syllable as a diacritic to unambiguously mark udātta accent syllables ॐ. Figure 2.4 gives an example of this udātta accent marking diacritic as typeset in pwg.

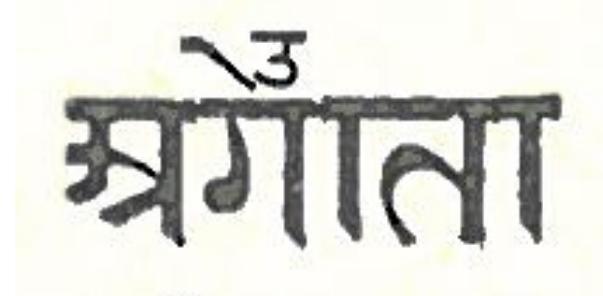


Figure 2.4: Udātta accent marking in Böhtlingk & Roth

Marking udātta accent by a raised ॐ is an innovation by Böhtlingk, according to [p. 30]Whitney1889. However, a raised ॐ occurs in the Sāmaveda accent marking system, but it is not a general marker of udātta accent in this system. [p. 485]Witzel1974 mentions a system used in South Indian Grantha manuscripts. In these manuscripts, udātta accent is marked by “a superscribed u” – so probably a raised ൗ (Grantha short u).⁶ Whether Böhtlingk was aware of this South Indian system or whether this striking similarity is accidental is unknown.

⁶The Grantha Unicode block (code points U+11300 to U+1137F) contains diacritical glyphs for the Sāmaveda accent marking system, parallel to the ones defined in the Devanagari extended block. In contrast to the Devanagari block however, the Grantha block does not define a diacritical letter u.

Independent svarita accent is marked by a vertical stroke above the syllable क्. This diacritic is identical to the one used to mark enclitic and independent svarita accent in the ऋग्वेदा accent marking system. However, while it marks independent svarita in the Böhtlingk & Roth system, it marks all svarita accents in the ऋग्वेदा system unless the requirement to mark the syllable preceding an udātta as anudātta supersedes it.

2.5 ISO 15919 System

The ISO standard 15919 defines a system for the transliteration of Indic scripts in Latin script. As part of this system, it specifies a system to mark Vedic accent in Latin script. The system follows the Western philological tradition to mark udātta and independent svarita. Rule 13 (page 17) of the standard defines the placement and form of accent marks.

Rule 13 The Vedic accent Udatta shall be transliterated as an acute accent over the transliterated vowel, and the independent Svarita as a grave accent over the transliterated vowel. In the case of the digraphs ai, au, the accent shall be attached to the second vowel.

This rule specifies that udātta accent is obligatorily marked as á or aú and independent svarita is marked as à or aù. Other svarita accents should not be marked with this diacritic or any other method. However in the recommendation section, ISO 15919 (p. 18) specifies an optional way to mark anudātta accent:

Where it is desired to show the Vedic accent Anudatta, it should be transliterated as an underscore. In the case of the digraphs ai, au, both Latin vowels should be underscored.

Thus, anudātta accent should be marked as a or au⁷ if required. However, linguistically the obligatory part of the standard – the marking of udātta and independent svarita accent – is sufficient to comprehensively mark the accents of Vedic Sanskrit.

agním īle puróhitam yajñásya devám ḡtvíjam | hótāram ratnadhátamam ||

Transferring the anudātta marks from the Devanagari editions of the ऋग्वेदा into the ISO 15919 system results in the following romanisation:

agním īle puróhitam yajñásya devám ḡtvíjam | hótāram ratnadhátamam ||

However, it should be noted that the recommendations for marking anudātta are less specific than the rules applying to anudātta accents in Indic textual traditions. In the ऋग्वेदा accent marking system as described by [p. 449]Macdonell1916, all unaccented syllables following the start of a half-line are marked as anudātta accent. However, anudātta accents following a svarita accent remain unmarked unless the anudātta immediately precedes an udātta accent. In contrast, ISO 15919 recommendations do not distinguish anudātta accents by the relative position in the stanzas or in relation to other accents. ISO 15919 allows transferring the anudātta accent from the ऋग्वेदा accent marking system into the Latin transliteration; as in example (2.5) above. However, it also allows to mark all anudātta accents as is done in (2.5) or any subset.

⁷ISO 15919 is not explicit which character should be used to encode this combining underscore. The best candidates in the Unicode character set are the code points U+0332 COMBINING LOW LINE and U+0331 COMBINING MACRON BELOW. In this document, we have opted for U+0332 COMBINING LOW LINE as this character is explicitly defined as an underscore in the unicode standard.

agním īle puróhitām yañásya dēvám ṛtvíjam | hótāram r̥atnādhátmam ||
The status of anudātta accent marking is clearly peripheral in the ISO 15919 system.

3 Encoding

Two encoding systems are relevant for our discussion of pitch accent markers in Vedic Sanskrit and their representation in digital lexicography. The first is the general encoding standard Unicode. The second is the Sanskrit specific system Sanskrit Library Phonetic Basic encoding scheme (SLP1).

3.1 Unicode

The Unicode Standard maintained by the Unicode Consortium¹ is the most important system to encode characters in digital media. It is crucial for any digital approach to lexicography, especially web-based ones. At the time of writing, the current version is Unicode 7.0.² The Unicode standard defines over 110,000 characters and assigns them to code points. The Devanagari block of the Unicode standard (code points U+0900 to U+097F) is well supported by fonts found in all major operational systems and even the rendering of the particularly complex ligatures found in Sanskrit is well implemented in these fonts. However, even though it is the largest encoding standard in existence, it is not possible to represent the different traditions of Vedic accent in a straightforward manner.

Let us first examine the diacritics used in the Latin alphabet, as these are well represented in the Unicode standard. The acute and grave accent as well as the combining underscore are included in the Unicode block Combining Diacritical Marks. The relevant glyphs with their respective code points and name are given below.³

- ⋮ U+0301 COMBINING ACUTE ACCENT
- ⋮ U+0300 COMBINING GRAVE ACCENT
- ⋮ U+0332 COMBINING LOW LINE

Figure 3.1: Relevant characters from the Combining Diacritical Marks range

The representation of Vedic accents in the Latin alphabet and its encoding in the Unicode standard are unproblematic. Furthermore, the glyphs for these characters are included in nearly every standard font making the use of these characters straightforward. Representing the marking of Vedic accents as found in Devanagari script is far less straightforward. Two of the diacritics used in Devanagari to indicate Vedic accents are represented by characters of the Unicode block Devanagari. The two diacritics are listed in Figure 3.2. The naming of these characters is highly unfortunate as the vertical stroke ᳚ is used to mark svarita accent in most marking systems, but is called DEVANAGARI STRESS

¹<http://unicode.org/>

²<http://www.unicode.org/versions/Unicode7.0.0/>

³Note that additionally several combined characters consisting of vowel glyphs with the acute and grave accent diacritics have their own code points e.g. á U+00E1 LATIN SMALL LETTER A WITH ACUTE.

SIGN UDATTA in the standard. The underscore ् is called DEVANAGARI STRESS SIGN ANUDATTA. While it is dominantly used to mark anudātta accent, it has other functions in some accent marking systems. This holds in particular for in the Śatapatha Brāhmaṇa, but also combined with numerals in the R̥gveda and several other textual traditions [p. 159]ScharfHyman2011. The Unicode nomenclature udātta for the vertical stroke ॑ is probably derived from its function to mark udātta accent in the Maitrāyaṇī Saṃhitā and the Kaṭhaka Saṃhitā [p. 450]Macdonell1916.

- U+0951 DEVANAGARI STRESS SIGN UDATTA
- U+0952 DEVANAGARI STRESS SIGN ANUDATTA

Figure 3.2: Relevant characters from the Devanagari range

Additionally, the Unicode standard contains a block of 48 code points (ranging from U+1CD0 to U+1CFF) called Vedic Extension. This block contains several diacritical characters organised by the textual traditions. The traditions explicitly named are Sāmaveda, Yajurveda, Śatapatha Brāhmaṇa, R̥gveda, and Atharvaveda. Not all of these are accent marking characters and in fact only one of them, a small J-shaped diacritical hook placed below the main character is used in lexicographic works. This character is given in Figure 3.3. Unfortunately, no widespread font seems to contain a glyph for this character, making it problematic for use in web-based applications.

1CD7 ◦ VEDIC TONE YAJURVEDIC KATHAKA
INDEPENDENT SVARITA
= vaidika svarita adho vakra rekhaa

Figure 3.3: The relevant character from the Vedic Extension range

Some means of accent marking found in textual traditions and in lexicographical works have been included into the Unicode standard. This block is called Devanagari extended and contains 38 code points (ranging from U+A8E0 to U+A8FF). It has been included following the proposal by EversonScharf2007. This extension adds the diacritical numerals of the Sāmaveda tradition, and the diacritical raised letters, including the lexicographically highly important diacritical raised ॒ which is used as an udātta marker in the system invented by Böhtlingk and Roth for their dictionary. The characters relevant for Sanskrit lexicography are the following four code points:

- ॑ U+A8E1 COMBINING DEVANAGARI DIGIT ONE
- ॒ U+A8E2 COMBINING DEVANAGARI DIGIT TWO
- ॓ U+A8E3 COMBINING DEVANAGARI DIGIT THREE
- ॔ U+A8EB COMBINING DEVANAGARI LETTER U

Figure 3.4: Relevant characters from the Devanagari Extended range

Some problems with Unicode support in fonts remain. The rendering of some of the existing characters defined in the Unicode standard is problematic. The clearest case is the lack of glyphs for the VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA mentioned above. Another problematic case is the combination of the vertical stroke $\hat{}$ and the underscore $\underline{}$ with the Devanagari numerals – especially ॑ and ॒ – as it occurs in the R̄gveda and other textual traditions. Furthermore but less important for lexicographical works, the combined diacritics ॒॑, ॒॒, and ॒॒॑ of the Sāmaveda system are not supported in any widespread font.

In summary, it can be said, that the Unicode standard is a reasonably comprehensive and well supported system to encode Devanagari texts. However, the support in fonts to represent of Vedic accent is still problematic. However, from the perspective of a lexicographer, Unicode and its support by fonts is sufficient to represent Devanagari in web-based applications.

3.2 SLP1

The Sanskrit Library Phonetic Basic encoding scheme (SLP1) as defined in ScharfHyman2011 is an elaborate and well-constructed encoding standard using a subset of (mostly Latin alphabet) unicode characters to encode Sanskrit. The crucial aspect of SLP1 is, that it encodes phonological and even phonetic categories of Sanskrit and not glyphs or characters. However, it only represents categories found in the texts and the traditional analysis of these texts as found in the traditional grammatical descriptions of Sanskrit.

The inventory for the description of intonational (or accent related) aspects of Sanskrit phonology consists of eight basic elements [p. 154]ScharfHyman2011:

/	high pitch
\	low pitch
^	circumflex
6	extra low tone
7	low tone
8	high tone
9	extra high tone
+	sharpness

Figure 3.5: Accent related characters of SLP1

The signs /, \, and ^ represent udātta, anudātta, and svarita respectively. However, these characters listed above can be combined in several ways to represent particular intonational phenomena of Sanskrit. The following four examples are the most relevant from a lexicographical point of view [p. 155]ScharfHyman2011:

This also means that SLP1 characters have multiple correspondences in the different textual traditions. As an example, SLP1 a/ [p. 162; also SLP2 002]ScharfHyman2011 can

/8	high tone (udātta)
\7	low tone (anudātta)
^98	dependent and unaggravated svarita
^97	aggravated independent svarita

Figure 3.6: Accent marking in SLP1

encode अ if the character occurs in the Śākala Saṃhitā of the Ṛgveda, the Vājasaneyi Saṃhitā of the Śukla Yajurveda, the Taittirīya Saṃhitā Kṛṣṇa Yajurveda, or the Śaunakī Saṃhitā of the Atharvaveda, औ in the Maitrāyaṇī Saṃhitā of the Kṛṣṇa Yajurveda, Ṛgveda Khilāni, and Kahmiri manuscripts of the Vājasaneyi Saṃhitā of the Śukla Yajurveda, the Kāṭhaka Saṃhitā Kṛṣṇa Yajurveda, and Paippalāda Saṃhitā of the Atharvaveda, it represents अ in the Sāmaveda Saṃhitā of the Kauthuma Śakhā [p. 162]ScharfHyman2011.

A full lists of SLP1 characters with their correspondences in the different text traditions can be found in [p. 162-203; also SLP2 002]ScharfHyman2011.

4 Accent marking in Sanskrit Dictionaries

Dictionaries differ significantly in how they treat and mark the pitch accent of Vedic Sanskrit. Several dictionaries disregard accent completely, while others mark it in the headwords. Dictionaries that indicate Vedic accent are pwg, gra, pw, ccs,cae, as well as mw. However, only pwg mark Vedic accent in the headwords as well as in quotations from Vedic and Brāhmaṇa texts. Beyond the extent of accent marking and the location in the structure of the entry, the individual dictionaries differ in the script they use to represent Sanskrit and in details of the marking systems employed. In the following the different systems are discussed.

4.1 Böhtlingk & Roth Sanskrit-Wörterbuch (1855-1875)

pwg – also known as *großes Petersburger Sanskrit-Wörterbuch* – is the largest Sanskrit dictionary. The *große Petersburger* is a bilingual Sanskrit-German dictionary and represents Sanskrit exclusively in Devanagari script. Böhtlingk und Roth employ three different systems to mark the accent of Vedic Sanskrit in the dictionary. Which system is used is determined from the position of a Sanskrit word in the entry. Headwords are treated differently from quotations from text. For quotations, the source of the Sanskrit text influences the choice of accent marking. However, the choice seems to be not determined by it, as will be discussed below.

From a lexicographical point of view, the treatment of Vedic accent in headwords is particularly important. Since all traditions of accent marking are text specific, there is no established method to mark accent in words outside of a textual context. The textual abstraction of lemmata in lexical entries requires an unambiguous method to mark Vedic accent. This is the rationale behind the innovation of the Böhtlingk & Roth System described above. This system is used in the headwords of entries and marks udātta accent by a diacritical raised [ঃ] above the syllable ^ক and independent svarita by the vertical stroke [ু]. Figure 4.1 as well as 4.2 and 4.3 are examples of entries with a headword containing an udātta accent. Figure 4.2 contains two entries – śapathyā and śápana – the first with an independent svarita and one with an udātta accent. Figure 4.5 shows that this system of accent marking is employed not only in the initial form of the headword, but throughout the form part of the entry, such as the adverbial instrumental case form akṣṇayā in the entry akṣṇa.

Accent in quotations from texts which indicate Vedic accent is marked by other means. Quotations taken from the R̥gveda and Atharvaveda are marked according to the textual and editorial tradition of these texts. This can be seen in the entry agótā in Figure 4.1. The accent marks in quotations from the Sāmaveda are transferred from the Sāmaveda system into the R̥gveda system. This is demonstrated by the quotation in the entry agníunna in

अर्गोता (von अगो) f. *Mangel an Kühen*: मा नौ अग्ने इमेतये मावीरेतयै
रीरंधः । मागोतायै सहस्रस्पुत्रं मा निदे इप द्वेषांस्या कृधि ॥ RV.3,16,5.
कुधामारं तृज्ञामारम् गोतामनपृथ्यतोम् । अपामार्गं बयो वृपं सर्वं तदपे मृमहे ॥
AV.4,17,6.

Figure 4.1: Böhtlingk & Roth entry agótā

शपथ्य (wie eben) adj. auf *Fluch beruhend* RV. 10,97,16.
शपन (von शप्) n. = शपथ AK. 1, 1, 5, 10. H. 262. *Fluch Trik.*

Figure 4.2: Böhtlingk & Roth entries śapathyā and śápana

अग्निनुन् (अग्नि + नुन् part. praet. pass. von नुद्) adj. *von Agni, vom Blitzstrahl getroffen*: तेषां वा अग्निनुन्नानुमिन्द्रो हतु वरं वरम् SV.II,9,
3,8,2. — Vgl. अग्निनूद्.

Figure 4.3: Böhtlingk & Roth entry agninunna

अग्निदग्ध (wie eben) adj. *vom Feuer verbrannt*: अग्निदग्धमिवेषं वृहं
भवति ÇAT. Br. 1,1,2,9.

Figure 4.4: Böhtlingk & Roth entry agnídagdha

Figure 4.3.

Quotations from the Śatapatha Brāhmaṇa are problematic. Some passages such as Ś.Br. 1,1,2,9 in Figure 4.4 are given with accents marked in the Śatapatha Brāhmaṇa system. In other cases, accent seems to be marked in the Ṛgveda system. The Śatapatha Brāhmaṇa (Ś.Br. 3,5,4,13) in the entry *aksṇa* in Figure 4.5 is an instance of this form. It is unclear what determines the choice of accent marking system. According to the preface of pwg, only Weber's edition of the Śatapatha Brāhmaṇa is used Weber1849. This text edition consistently uses the Śatapatha Brāhmaṇa system to mark accent. The letters in [Brückner et al.(2007)Brückner, Zeller, and Stache-Weiske] do not seem to contain any information that could help understanding what determines the choice of marking system for the individual passages.

1. अद्वा adj. *in die Quere gehend*. Nur der adverbiale Instr. अद्वाया ist zu belegen: 1) *in die Quere*: तानद्वाया सं तृन्दति पद्वद्वाया न शङ्कुयादपि समीचः ÇAT. Br. 3, 5, 4, 13. (Sch. = वक्रमार्गेण, Sch. zu KÄTJ. 8, 5, 11 = कौटिल्येन) अथ पद्वद्वायावद्यति । सव्यस्य च देष्टो द्विनिषायांश्च श्रोणेर्द्विनिषास्य च देष्टोः सव्यापाश्च श्रोणेस्तस्माद् यं पश्चुर्द्वाया पदो द्वृत्यथ यत्सम्यग्वद्येत्समीचौ हृवायं पशुः पदो द्वृते 3, 8, 2, 27. अद्वाया द्विनिषां इसे श्रोणां श्रोणामसे KÄTJ. 5, 4, 14. (VS. 5, 12). — 2) *in verkehrter, sündhafter Weise*: स पद्वनेन किंचिद्वद्वाया कृतं (vielleicht als Compositum अद्वायाकृतं zu lesen; ÇĀMKĀR. sieht ein अकृतं darin) भवति तस्मादेनं सर्वस्मात्पुत्रो मुद्धति Br. ÄB. UP. 1, 5, 17. — Von अद्वा.

2. अद्वा n. U. 3, 17. Zeit CKDR. (अद्वा).

Figure 4.5: Böhtlingk & Roth entry akṣṇa

The case of Śatapatha Brāhmaṇa quotations is problematic for two separate reasons. Except for some passages from the Śatapatha Brāhmaṇa, all quotations in pwg seem to use the Ṛgveda system to mark accent. This includes not only quotes from the Ṛgveda and the Atharvaveda, but also quoted passages from the Sāmaveda, which are transferred into the Ṛgveda system as SV II.9,3,8,2 in Figure 4.3 shows. If the Ṛgveda system were used consistently for accent marking throughout the dictionary, the diacritical underscore ₖ would have a single function in this work. With the introduction of the Śatapatha Brāhmaṇa, the diacritical underscore ₖ becomes multi-functional marking pre-udātta anudātta accents in one system and udātta – or in another interpretation pre-svarita syllables – in the other. The situation is further complicated by the fact that not all passages from the Śatapatha Brāhmaṇa are given in the same system.

There is a second problem with the interpretation of the accent marks in passages from the Śatapatha Brāhmaṇa in pwg. The interpretation of the accent marking system of the Śatapatha Brāhmaṇa is controversial. The currently preferred interpretation is generally attributed to Hoffmann1956. The previously preferred interpretation can be found in [p. 451]Macdonell1916, but seems to be much older. Differences between the two interpretations of the system could lead to different results when the accents marked in the Śatapatha Brāhmaṇa system are translated into the Ṛgveda system. pwg contains over 10000 references to the Śatapatha Brāhmaṇa. Although not all involve a quotation from the text, it is not possible to predict how accent is marked in all these cases and how the marking has to be interpreted.

The Sanskrit-Wörterbuch in kürzerer Fassung pw employs the same system to mark Vedic accent in the headwords as the große Petersburger Sanskrit-Wörterbuch pwg. Since pw does not contain quotations from Sanskrit texts, no other system of accent marking

occurs in it.

4.2 Monier-Williams Sanskrit-English Dictionary (1872/1899)

Monier-Williams introduced the representation of Vedic accent in his new edition of the Sanskrit-English Dictionary mw. Accent is only marked in the Latin script rendering of headwords and is virtually identical with the system described in ISO 15919. Udātta accent is marked by an acute accent á and independent svarita is marked by grave accent à. On diphthongs, the accent is marked on the second vowel character aú. Monier-Williams describes the system in a very succinct manner in the introduction of the dictionary [] [p. xviii]mw. The original edition of 1872 mw72 did not mark accent on the headwords.

4.3 Grassmann Wörterbuch zum Rig Veda (1873)

Grassmann in his dictionary of the Ṛgveda gra records accents. The dictionary is typeset in Latin script exclusively and the accent marking is solely based on udātta accents.

Udātta accent on a short vowel is marked by an acute accent á. On diphthongs, the first vowel character aú carries the accent mark. If an udātta accent occurs on a long vowel, the diacritical acute accent and the vowel length indicating macron fuse into a diacritical circumflex â. Independent svarita accents are marked as udātta on the glide preceding the svarita accent vowel i.e. ýa. Thus Grassmann gives his entry for शपथ्य śapathyà as çapathýa (with ç as the transliteration of श in Grassmann's system).

4.4 Cappeller Sanskrit Wörterbuch (1887) & Sanskrit-English Dictionary (1891)

Cappeller in his Sanskrit Wörterbuch ccs and his Sanskrit-English Dictionary cae follows pwg in setting Sanskrit words exclusively in Devanagari and indicating udātta and independent svarita accents. However, the system used in ccs,cae is different in the diacritical characters it employs. The diacritical inventory of these two dictionaries is derived from the Kāṭhaka Saṃhitā Kṛṣṇa Yajurveda. It uses the vertical stroke over the central glyph of the syllable क् to indicate udātta accent and a lying J-shaped diacritic below the syllable to indicate independent svarita. This diacritical accent marker can be seen in the entry śapathy from cae in Figure 4.6. The system can be encoded in Unicode with the following code points – U+0951 DEVANAGARI STRESS SIGN UDATTĀ and U+1CD7 VEDIC TONE YA-JURVEDIC KATHAKA INDEPENDENT SVARITA – but the latter character is not included in any of the widespread Devanagari fonts.

4.5 Macdonell Sanskrit-English Dictionary (1893)

md represents Sanskrit headwords in Devanagari and Latin script. While he disregards accent in Devanagari, he marks udātta accent in the Latin script form. The accent is marked

शपथ् a. consisting in a curse.
 शपन् n. cursing.
 शप्त् a. cursed, conjured; n. curse, oath.

Figure 4.6: The entry śapathya in cae

by an acute accent á. Vowel length is marked by a combining circumflex accent above the vowel and a long vowel with udātta accent is ā or â'. As can be seen from the syllables pā and pū in Figure 4.7, the relation between the two diacritics is not entirely clear in the print. Macdonell uses a particularly idiosyncratic and analytic system to represent Sanskrit words in Latin script. So, he transliterates ए as aj joint by a bow below. This might also be the reason that we could not verify an instance of an independent svarita in md.

अग्रेपा agre-pā, a. drinking first; -पु, a. *id.*
 अग्रेसर् agre-sara, a. (i) going before.

Figure 4.7: The entry agrepā in md

4.6 Schmidt Nachträge zum Sanskrit-Wörterbuch (1928)

Schmidt in his Nachträge zum Sanskrit-Wörterbuch sch writes Sanskrit in Latin transliteration and marks udātta as acute accent. Since he uses a diacritical macron above vowels to indicate length, the resulting forms are á and â. we could not verify an instance of an independent svarita in sch.

4.7 Summary

Seven dictionaries record Vedic accent in the headwords. One single work – pwg – features extensive quotations in Devanagari script with accent marks. All dictionaries base their accent marking systems in the headword form on the udātta accent. pwg, mw, and ccs,cae mark independent svarita with a different device than the udātta, while gra basically reinterprets independent svarita as an udātta on the glide preceding the svarita bearing vowel and marks it accordingly with the same diacritical acute accent as the udātta. Two dictionaries md,sch seem to lack instances of independent svarita in headwords. The following table summarises the systems of accent marking found in the dictionaries.

	अग्नि agní	अग्रजा agra᷍jā	शपथ्य śapathyà
pwg	अग्ने᷍	अग्रजा᷍	शपथ्यं
mw	अग्नि agní	अग्रजा agra᷍jā	शपथ्य śapathyà
gra	agní	agra᷍jâ	çapathyá
ccs,cae	अग्ने᷍	अग्रेपा᷍	शपथ्य [lying J below य]
md	अग्नि agní	अग्रेपा᷍ agrepâ'	—
sch	agní	parapátam	—

Figure 4.8: Accent marking in Sanskrit Dictionaries

5 Principles applied in the TEI CSL

The Cologne Sanskrit Lexicon (CSL) marking Monier, KappMalten1997 – also known as Cologne Digital Sanskrit Lexicon – is the largest lexicographical resource for Sanskrit. The creation of a TEI version of the Cologne Sanskrit Lexicon is part of the Lazarus Project¹ and aims for long-time preservation of the data. It is based on the original digitisations and mark-up versions of the CSL and uses the TEI Guidelines, especially the dictionary module. The objective of the TEI Cologne Sanskrit Lexicon is to preserve all information contained in the original prints, as far as it was preserved in the digitisation process [] [as described in] KappMalten1997, while using a well documented and standardised XML. The second objective is to display the information as consistent and faithfully as possible to the original prints, while allowing the user to choose the writing system in which the Sanskrit words are displayed.

Internally, the CSL uses a slightly modified version of SLP1 to encode Sanskrit. The crucial difference to SLP1 as defined in ScharfHyman2011 is that the CSL encodes accent diacritical marks and not the accents themselves. That enables the system to be agnostic to the system employed and to the interpretation of the system. This is particularly crucial in the case of the passages from the Śatapatha Brāhmaṇa in pwg. In particular, this means that the vertical stroke के is internally represented as ^ regardless of whether के represents dependent or independent svarita as in the R̥gveda accent marking system found in quotations in pwg or independent svarita as in the headwords in pwg. In pwg, underscore के is internally represented as \ regardless of whether it represents anudātta accent as in the R̥gveda accent marking system found in quotations from the R̥gveda, Atharvaveda, Sāmaveda, and some passages of the Śatapatha Brāhmaṇa in pwg or whether it is the only accent mark of the Śatapatha Brāhmaṇa system with disputed function.

The deviation from the SLP1 system can best be seen in case of the R̥gveda aggravated independent svarita marking ऽ in which the numeral ? features the vertical stroke के as well as underscore के. Figure 5.1 gives an instance from the entry जन्मन् janman from pwg. The aggravated independent svarita marker should be encoded as ^97 following ScharfHyman2011, but is represented as 1\^ in the internal transcription of the CSL. The CSL treats the aggravated independent svarita not as a simplex marker ऽ, but as a composite entity consisting of the numeral ? as its base to which two diacritics are added. In this case, the internal encoding of the CSL represents typographical facts and can remain agnostic to specific annotations of these markers.

At the time of writing, when the Devanagari script accent marks are converted into ISO 15919 romanisation, the diacritical ऽ udātta के is represented by an acute accent, underscore के is generally discarded, while vertical stroke के is discarded in quotations, but represented as grave accent in headwords. This results in the correspondences for pwg and potentially pw, given in 5.2.

However, this means that the accent marking is lost when quoted passages as found in pwg are converted into Latin script. An algorithm that allows to convert the anudātta and

¹<http://www.cceh.uni-koeln.de/lazarus>

Figure 5.1: Aggravated independent svarita in pwg

Display form	Print	SLP1	Unicode
Devanagari	०	/	U+A8EB COMBINING DEVANAGARI LETTER U
	॑	^	U+0951 DEVANAGARI STRESS SIGN UDATTĀ
	॒	\	U+0952 DEVANAGARI STRESS SIGN ANUDATTĀ
ISO 15919	०	/	U+0301 COMBINING ACUTE ACCENT
	॑	[DISCARDED in examples]	
	॒	^	U+0300 COMBINING GRAVE ACCENT in headwords
	॒	\	[DISCARDED]

Figure 5.2: Technical treatment of accent in the CSL

svarita of the R̥gveda marking system into the udātta and independent svarita marking in diological Romanisation is not trivial. The fact that some (but not all) quotations from the Śatapatha Brāhmaṇa are in the Śatapatha Brāhmaṇa system while others were converted into the R̥gveda marking system by pwg makes an automatic conversion virtually impossible.

Unfortunately, in the case of the two Cappeller dictionaries ccs,cae, the vertical stroke क (representing udātta accent) is encoded as / in the internal representation of the CSL and the same character has been used to encode the lying J-shaped line below (representing independent svarita). This issue is so far unresolved.

	क	क	क	[Lying J below]
Böhlingk & Roth	/	^	\	
Capeller		/		/

Figure 5.3: Problems with the technical treatment of accent in the CSL

Bibliography

[Brückner et al.(2007)Brückner, Zeller, and Stache-Weiske] Brückner, Heidrun, Gabriele Zeller, and Agnes Stache-Weiske, ed. 2007. *Otto Böhtingk an Rudolf Roth : Briefe zum Petersburger Wörterbuch 1852-1885*. Wiesbaden: Harrassowitz.