

# New Type Word Indexes for Use in Sanskrit Philology

Boris Oguibénine  
Université Marc Bloch de Strasbourg  
oguibeni@umb.u-strasbg.fr

## Abbreviations and Signs

**Abhis-Dh (Mā-L)** *Abhisamācārikā (Bhikṣuprakīrnaka)* ed. B. Jinānanda, Patna: K.P. Jayaswal Research Institute, 1969 **AiGr** J. Wackernagel, A. Debrunner, *Altindische Grammatik*, I, II, IIa, III, Göttingen: Vandenhoeck & Ruprecht, 1957- **BHSD** F. Edgerton, *Buddhist Hybrid Sanskrit Dictionary*, New Delhi: Motilal Banarsidass, 1972 **BHSG** F. Edgerton, *Buddhist Hybrid Sanskrit Grammar*, New Delhi: Motilal Banarsidass, 1972 **BhīVinR** G. Roth, *Bhikṣuṇī-Vinaya. Manual of Discipline for Buddhist Nuns*. Edited and Annotated for the First Time, Patna: K.P. Jayaswal Research Institute, 1970 **Bhu-P** N.Tatia, *Prātimokṣasūtram*, Patna: K.P. Jayaswal Research Institute, 1975 (Tibetan Sanskrit Works Series, n°16; quotations by pages and line) **Mv** E. Senart, *Le Mahāvastu. Texte sanskrit publié pour la première fois et accompagné d'introductions et d'un commentaire*, T. I-III, Tokyo: Meicho-Fukyū-Kai, 1977 **RV** Die Hymnen des Rīgveda, I-II, Berlin: Akademie-Verlag, 1955  
# pause; >becomes or develops in; < is derived from; vs. versus; (vs) verses

I. This paper deals with the word indexes of the Sanskrit texts belonging to the linguistic and literary tradition of the Mahāsāṃghika-Lokottaravādins that have been published over the last five or six years.<sup>1</sup> The introductory remarks to each of

---

<sup>1</sup>B. Oguibénine, Y. Ousaka et M. Yamazaki, *Prātimokṣasūtram. Word Index and Reverse Word Index*, Tokyo: The Chūō Academic Research Institute, 2001 (Philologica Asiatica, Monograph Series 17); E. Fauré, B. Oguibénine, Y. Ousaka et M. Yamazaki, *Mahāvastu-Avadāna. Word Index and Reverse Word Index*, vol. I, Tokyo: The Chūō Academic Research Institute, 2003 (Philologica Asiatica, Monograph Series 20); E. Fauré, B. Oguibénine, Y. Ousaka, M. Yamazaki, *Mahāvastu-Avadāna. Word Index and Reverse Word Index*, vol. II, Tokyo: The Chūō Academic Research Institute, 2006 (Philologica Asiatica, Monograph Series 22); Y. Ousaka, M. Yamazaki, *Bhikṣuṇī-Vinaya. Word Index and Reverse Word Index*, Tokyo: The Chūō Academic Research Institute, 2002 (Philologica Asiatica, Monograph Series 18). Unfortunately, the last index has many drawbacks. The authors follow slavishly the printed text of BhīVin with regard to the members of

these indexes explain the underlying principles (that I claim to be new). Hopefully, critical observations from those interested in Sanskrit vocabulary and in the issues regarding the juncture as a cross-linguistic phenomenon might stir up further discussions.

Each of the indexes cited in footnote 1 comprise a serial index and a reverse index of the respective Sanskrit texts.

II. Let us compare some of the extant indexes of Sanskrit vocabulary (I mean the indexes that are known to me and supposedly to most Sanskrit scholars). The comparison must permit us to single out the differences in approach that have determined the shape of the indexes under scrutiny. Note that almost all of the indexes mentioned in this paper are focused on restricted corpora of texts.

**Serial indexes** (based on reading according to the “Greek order” from left to right or ordering the words *a fronte* distinct from the order *a tergo*<sup>2</sup>) are usually and primarily lists that record lexical items in the order which, for Sanskrit, is that of the *varṇasamāmnāya* “assemblage or aggregate of letters” or “traditional collection (of words); enumeration from *a* to *ha*”. They are mostly meant for checking the availability of lexical items found in a text. Such indexes make possible several types of research, particularly the research in lexicography or in contents of the texts from any standpoint.

1. Index of words following the publication of a manuscript of the *Bhikṣuṇī-Karmavācanā* by M. Schmidt.<sup>3</sup>

---

the compounds. These may appear separately (for example, the compound *BhīVin* § 28 *deśita-śikṣā* ‘[a young girl] who has been given the precepts’ is quoted under two entries, citing each member; but this compound as such, like many others of the same ilk, is precisely characteristic of the language of the *Mahāsāṃghika-Lokottaravādins*). Conversely, the words written joined in the manuscript and so reproduced in the printed text are not disjoined as is required for lexicographical purposes. Thus the sentence *BhīVin* § 5 *gaccha āvusa ānanda* ‘go, brother Ānanda!’ appears as one long sequence *gacchāvusānanda* so that it is impossible to locate the rarely occurring *āvusa* (to be added to the list of its occurrences in *BHSD* p. 108).

<sup>2</sup>See, for example, *Laterculi vocum latinarum. Voces latinas et a fronte et a tergo ordinandas curavit Otto Gradewitz*, Leipzig, 1904.

<sup>3</sup>M. Schmidt, ‘*Bhikṣuṇī-Karmavācanā*. Die Handschrift Sansk.c.25 (R) der Bodleian Library Oxford’ in *Studien zur Indologie und Buddhismuskunde für*

2. *Word Index to the Bhagavadgītā* by F.X. D'Sa (Pune: Institute for the Study of Religion, 1985).

In these two indexes, the words are quoted in their stem forms. The only peculiar feature in D'Sa's index is that the nouns are quoted in the stem forms, whereas the pronouns and the verbs are quoted in the inflected or conjugated forms in which they appear in the texts: thus s.v. *etad-* (the demonstrative pronoun), we find the case forms *etat-*, *etayoh*, *etasya*, and under the verbal root *han-* 'to smite, slay, kill, etc.' we find the finite forms *ghātayati*, *hanti*, etc. Even if it is empirically useful, one cannot but wonder at the inconsistency of such an indexation, a flaw probably due to the neglect of the linguistic perspective.

3. 'KWIC (Key Word in Context) Index to the Vākyapadīya, Kāṇḍa 1' by T. Unebe, *Nagoya Studies in Indian Culture and Buddhism: Saṃbhāṣā* 22, 2002, pp. 1-237, is fairly exceptional. First, it gives a list of all the words occurring in the text with a short section of the context preceding and following each occurrence of the word. Such references are undoubtedly useful for the semantic and lexicographic research. But, second, and that point is highly relevant for the theoretical assessment of indexing of Sanskrit texts, the author has effected two modifications in the words: one is to restore the anusvāra not occurring in word-endings to homorganic nasals, for example *saṃkhyā* > *saṅkhyā*, *saṃjñā* > *saṅjñā* (no reason is apparent and no explanation is given for this change); another modification consists in separating the words from each other and to cancel sandhi in the case of word initials. For example, when the text has *pradeśo 'pi*, the index will quote this sequence as *pradeśo ^ api*; *yac chabdān* will be quoted as *yac ^ śabdān*, etc. Thus is intelligibility achieved with regard to word initials modified by sandhi.
4. L. Renou, 'Index védique' in *Vedic Studies*. A

*Collection of the Research Papers of Prof. Dr. Raghu Vira* (Śata-Piṭaka Series, Indo-Asian Literatures, vol. 272), New Delhi, 1981, p.538-699.

5. W. Rau, 'Vaidika-Padānukrama-Koṣa und Veda-Lexikographie', *Kratylos*, 28, 1984, pp. 1-25, collects an impressive mass of information on the word indexes contained in Vishva Bandhu's *Vaidika-Padānukrama-Koṣa* (five volumes between 1942 and 1965). Part A lists the indexes of Vedic texts, while part B provides an extensive bibliography of Vedic indexes published in India and elsewhere in the world. Rau's summary of Vishva Bandhu's work provides a useful glance at the points that are material for grammatical studies and that are to be kept in mind when using these indexes: the inconsistent use of ordering of the lemmas (partly alphabetic, partly etymologic); the listing of the compound verbs under the verbal prefixes and not under the simple roots; the lists of the non-prior members of the nominal compounds combining with the prior members. Rau also draws user's attention to the extended denotation that Vishva Bandhu imprinted on the concept of the Vedic language (p. 5): his indexes include the Epic-Puranic Sanskrit vocabulary as well as that of Tantric and Modern Sanskrit.
6. Among the indexes listed in part B (86), only a few account specifically for the grammatical characteristics of the words.

**Reverse indexes** (based on the *a tergo* reading) are, in principle, meant to enable the users to find out the grammatical characteristics of words, given that in Sanskrit these characteristics gather mostly (but not exclusively) in the final portions of words, i.e., those morphologically relevant portions which carry the grammatical information. I briefly characterize below those reverse word indexes of Sanskrit of which I am aware (including the reverse indexes mentioned by Rau).

7. H. Grassmann's index, appended to his *Wörterbuch zum Rigveda* (6.Auflage von M. Koziánka, Wiesbaden: O. Harrassowitz, 1996, pp. 1686-1740), is divided into two parts: one contains the declinable words i.e., the nouns and the verbal roots (termed by Grassmann *biegsame Wörter*), the other comprises the undeclinable words (*unbiegsame Wörter*). In the nouns and pronouns cited in stem forms, special attention is given to the word final suffixes separated by a hyphen, e.g., *devāvā-ta* 'desired by gods'. Words with no suffix are followed by the hyphen, e.g., *pratidhā-* 'a portion drunk at once' (RV.8.77.4). This part of the index is utterly incomplete: e.g., if *pratidhā-* is quoted, one finds no *pratidhī* 'a cross-piece on the pole of a carriage'.<sup>4</sup> The prefatory note has a strange remark that "Die Pronomen, sofern sie pronomielle Deklination zeigen, sind durch † [...] bezeichnet" (the pronouns cannot but follow the pronominal declension!). In fact, not only are the pronouns mostly cited as such, but so also are the pronominal adjectives, e.g., *anyā*, *para-mā*, *ta-kā*, etc., in other words, the suffixed formations. Another oddity of this index is the assumption of the suffix *-ithi* in such nouns and personal names as *ātithi* 'host', *médhātithi*, *nīpātithi* (p. 1721), etc., whereas the standard view finds there the suffix *-thi* [although AiGr II,2, p. 328 mentions the suffix *-ithi*, all the respective formations are treated under *-thi*, p. 722]. The undeclinable words (*unbiegsame Wörter*) include the adverbs, verbal prepositions, and particles. One can see that Grassmann's intention was not only to provide

---

<sup>4</sup>Grassmann's technique, although said to be "[an] ingenious device and management of a hyphen, final or inserted between primitive and added ending" is rejected by W.D. Whitney, 'Index Verborum to the Published Text of the Atharva-Veda', *Journal of the American Oriental Society*, vol. 12, 1881, p. 343. Most interesting, for the reasons that will appear below, is Whitney's statement that one of the drawbacks to the successful use of Grassmann's reverse index is "the separation of stems with a phonetically altered ending from others of the same formation".

the word lists, but also to supply concise grammatical information, although scarcely enough. Of course, the main bulk of Grassmann's dictionary provides very substantial grammatical material. Summing up, although his reverse index is relatively short of data on the grammatical characteristics of words, he should be reckoned among the keenest Indologists of his time; he clearly states in his prefatory note to the reverse indexes that these are "besonders für grammatische Fragen unentbehrlich" (p. 1686).

8. 'Index by Finals', a chapter (pp. 343-372) of W.D. Whitney, 'Index Verborum to the Published Text of the Atharva-Veda' (see footnote 4) presents "the great body of Atharvan simple declinable stems" as well as the indeclinables and the verbal roots of the Śaunaka recension of the Atharva-Veda. There is an attempt to give special consideration to the grammatical characteristics of words. Whitney explicitly points out that "all comparative and superlative stems are duly entered [...] in their alphabetic place". The reason for the emphasis on these formations, although not commented explicitly, is probably to draw attention to the distribution of the comparative and superlative suffixes in the Ṛgveda and the Atharvaveda as it is analyzed in Whitney's *Sanskrit Grammar* (§ 471c). Furthermore, the feminines in *ā*, *ī* and *ū* are given separately from the main bulk of the index (pp.372-377). That is justified by the particular morphological features of certain feminine formations with regard to the masculines (e.g., see such cases as *āsiknī* f. to *āsita* m. 'black, dark-coloured'; *śyēnī* f. to *śyetā* m. 'reddish white, white', etc). Finally, a list of "tense-, conjugation-, and mode-stems" (pp. 377-382) is a useful grammatical addendum.
9. W.D. Whitney's, *Roots, Verb Forms, and Primary Derivatives of the Sanskrit Language* (Leipzig:

Breitkopf and Härtel, 1885) contains an “Index of Roots arranged in reversed alphabetical order” (pp. 243-247), which lists “more than 800 roots”. The special feature of this index is the chronological classification of the roots occurring in “the earlier” or “the later” Sanskrit texts. Whitney explicitly states that “it seemed better not to classify by periods”.

10. The V<sup>th</sup> section of Vishva Bandhu’s Vaidika-Padānukrama-Koṣa (see above, under **Serial indexes**, n.5) is an index Ab Ultimo (628 pp.) [*non vidī*].
11. The most complete, and therefore very useful, *Rückläufiges Wörterbuch des Altindischen* by W. Schwarz and O. E. Pfeiffer (Wiesbaden: O. Harrassowitz, 1974-1978) presents Sanskrit words in stem forms with no specification of their grammatical categories, although the verbal roots are kept apart under the heading *Verba* (pp. 905-924).

It clearly appears that the extent of the grammatical information provided is not only highly unequal in the reverse indexes,<sup>5</sup> but also that the grammatical information is recorded randomly and with no special emphasis on its relevance for linguistic research.

III. Our reverse indexes have been conceived to address the problems lying on the border of grammar and lexicography from a linguistic and grammatical perspective. Among such problems are those of word-compounding and the occurrence

---

<sup>5</sup>Several review articles by I. Mel’čuk point to the importance of the references to the grammatical information in the reverse indexes. See especially *Voprosy jazykoznanija*, 6, 1958 (on the *Dičionar invers*, București, 1957); (with R.V. Bakhturina) *Voprosy jazykoznanija*, 5, 1965 (on M.L. Alinei, *Dizionario inverso italiano*, The Hague, 1962); *Voprosy jazykoznanija*, 6, 1966 (on M. Lejeune, *Index inverse du grec mycénien*, Paris, 1964, E. Mater, *Rückläufiges Wörterbuch der deutschen Gegenwartssprache*, Leipzig, 1965 and A. Juilland, *Dictionnaire inverse de la langue française*, The Hague-Paris, 1965) and *Izvestija Akademii Nauk, Serija Literatury i jazyka*, XXVII, 5, 1968 [on *Inverzionnyj slovar’ gruzinskogo jazyka pod rukovodstvom B. Počkhuva*, 1967; *Obraten rečnik na makedonskiot jazik*, sost. Vl. Militič, Skopje, 1967, and *Normal Reverse English Word List* (compiled under the direction of A.F. Brown), Philadelphia, 1963].



of morphological marks in word-inflection whose solution depends on questions like the following: “Does a compound with the final member  $x$  exist? and what  $y$  and  $z$  members precede it?” or “By how many members may a given final member  $x$  be preceded?”, “What and how frequent are the inflectional endings of such and such declined or conjugated form?”, and so on.

But what has incited me to a new conception of the indexes are the issues regarding the morphophonemic rules (*external* sandhi rules; further on I say “sandhi”). Indeed, sandhi in Buddhist Sanskrit is applied with the result of a rather bewildering mass of diverse forms conveying the same grammatical meaning. The seemingly random application of sandhi produces the impression that the words in contact are not submitted to any observable constraints [some cases of sandhi in Buddhist Sanskrit texts not occurring in standard (classical) Sanskrit are the subject of a short chapter in BHSG, pp. 32-37; see a detailed treatment of sandhi in the texts of the Mahāsāṃghika-Lokottaravādins in my forthcoming monograph<sup>6</sup>].

It therefore appeared advisable to organize the serial indexes as repertoires of lexical items in their basic morphophonemic shape, i.e. *disregarding (cancelling) the sandhi effects and showing the words in their form in pausa*.

The main reason for this choice was the distribution of the grammatical information in the words undergoing changes due to the application of sandhi. Only final portions of words (morphemes or submorphemic sequences of sounds) most frequently show alternants due to sandhi, except for such relatively rare occurrences as Mv.1.1.9 *bhagavān cchākyamunir*, Mv.1.2.7 *bhagavāṃ cchākyamuniḥ* where not only the final of the word *bhagavān*, but also the initial of the word *sākyamuni* is affected by sandhi [see, however, the non-application of sandhi in Abhis-Dh (Mā-L) 33.5 *bhagavān śrāvastyām viharati* vs. its partial application in BhīVin § 170 *bhagavāṃ cchrāvastyām* and Mv.1.182.16 (vs) *tato amantraye śāstā śrāvakāṃ* (AccPl) *chāsane* (< *sāsane*) *ratāṃ* 'Then the Master addressed the disciples delighted in his teaching']. The serial index recording the basic forms of words accounts for the state of the text open

<sup>6</sup>A *Descriptive Grammar of Buddhist Sanskrit. I. The language of the Mahāsāṃghika-Lokottaravādins Textual Tradition. General Introduction. Sound Patterns. Sandhi Patterns.*



to sandhi alternations, which are not identical to those of standard Sanskrit and are, as stated above, rather randomly (optionally) applied. That means that it is impossible to derive the attested alternants by automatically applying the sandhi rules to the lexical items. In standard Sanskrit the morphophonemic shape of words is predictable, whereas the predictability is much lower in our texts.

To bypass this inconvenience, I decided to provide a full account of the occurring morphophonemic alternants of the words in the reverse index. This distinction matches what is understood when W.S. Allen says that “selecting one of the alternants as “basic” involves that [...] the others can be unambiguously derived from it by the application of appropriate rules”.<sup>7</sup>

Consequently, the user of the indexes has to approach them according to the kind of information he requires. The vocabulary is presented either from the lexicological perspective – which necessarily includes the morphological information (the serial index) – or from the morphophonemic perspective, which shows how the finals of words vary according to sandhi alternations (reverse index). The amount of grammatical information found in the latter is more comprehensive.

The following example illustrates the different forms of the same word recorded in our two indexes.

Mv.1.1.9 reads *bhagavāñ cchākyamunir mamāpi śākyamunir iti* ‘[May I be called like this] Blessed One Śākyamuni, [may] my [name] be Śākyamuni’. The serial index records the items as *bhagavān, śākyamuniḥ, mama, api, śākyamuniḥ, iti*, whereas the reverse index will show them as *bhagavāñ, cchākyamunir, mama, api, śākyamunir, iti*. Repetitions in both indexes that exhaustively record all lexical items in the occurring forms are evidently unavoidable, as may be seen from the quoted sequence.

Since the sandhi effects are cancelled in the serial index, the vowels are disjoined in the cases of vowels contracted at the word boundaries, and the words are shown as independent lexical items, e.g., *mamāpi* in the quoted example: both indexes will include *mama* and *api*. If a vowel is elided, as in

<sup>7</sup>W.S. Allen, *Sandhi. The Theoretical, Phonetic, and Historical Bases of Word-Junction in Sanskrit*, ‘s-Gravenhage: Mouton, 1962, p.19.

Mv.1.68.8 (vs) *h'imau*, the serial index will quote *hi* and *imau*, but the reverse index will record these two items as *h'* and *imau*,<sup>8</sup> if Mv.1.72.1 (vs) reads *hi 'naikabhavakalpaśatehi* 'for after many a hundred kalpas of existence [...]', the reverse index shows these words as quoted, but the serial index has them as *hi* and *anaikabhavakalpaśatehi*.

As noted in BHSD 2.64, "there is much confusion in writing the anusvāra sign and both *m* and *n*, especially final". When this applies to Mv, given that the anusvāra and stop nasals frequently appear in the same phonetic environment and the same syntactic positions, it is difficult to determine clearly enough what are operating constraints as it is difficult to determine whether we have to deal here with phonetics or with orthography. Indeed, on the one hand, Mv.1.200.14 (vs) reads *adbhutaṃ tubhyaṃ* and Mv.1.206.11 (vs) has *adbhutaṃ idaṃ*; on the other hand, such sequences with the anusvāra as Mv.1.13.10 (vs) *kālasūtraṃ idaṃ*, Mv.1.14.14 *rauravaṃ iti akhyāsi*, Mv.1.17.4 *yūyaṃ imehi*, etc., are most frequent. It would be wrong to conclude that the anusvāra is replaced by the stop nasal before a vowel as it is done in standard Sanskrit.

It has, therefore, been found useful not to arbitrarily emend the words in the indexes, but rather to record both variants of a given item in the serial index, i.e., in the quoted cases, *adbhutaṃ* and *adbhutam*.<sup>9</sup> Interestingly, Whitney seems to have anticipated the necessity of such listing the words which differ by their finals in accordance with their phonetic environment, i.e. in accordance with the sandhi rules (see his criticism of Grassmann's reverse index reported in footnote 4).

The sequences alternating the final sibilants with the visarga, in which both are replaced with a voiceless palatal in appropriate combinations, are treated differently in our texts. See, for example, Mv.1.89.13 *kīrtiślokaparāś ca bhavanti* 'and they set value on fame and renown'. In *pausa*, the final syllable of the first word would appear as *-āḥ*. In order to show how the

<sup>8</sup>Cf. among numerous similar occurrences the sequence combining the particle *hi* and an initial *ī* with both vowels preserved: Mv.2.327.5 (vs) *hi imam*. Both words belong to the light words that most frequently, although not always, fuse their vowels.

<sup>9</sup>In Buddhist Sanskrit the position in *pausa* is one determining the alternation of the nasal stop *-m* which appears then, rather regularly, as anusvāra. By this token this alternation is to be considered as due to sandhi (*-m + #*). As may be seen from the examples quoted in the preceding paragraph, it is not applied automatically before vowels.

sandhi constraints are applied, the serial index records the first word as *kīrtiślokaparāḥ*, while the reverse index quotes it as it appears before the particle *ca*. This evidence is even more remarkable since the alternation *-āḥ > -ās* occurs chiefly before the conjunctive *ca* (a light word), while the sandhi between the full-fledged words are extremely rare: see, for example, Mv.1.125.2 (vs) *tuṣitebhyaś cyavantānām* ‘[the number of] those who pass away from Tuṣita’; Mv.1.329.21 *evarūpaś cetaso* or Mv.1.330.19 *evamrūpaś cetaso* ‘the mental reflection of such a kind’. On the contrary, the absence of sandhi between the full-fledged words as, for example, in Mv.1.59.4 *catasraḥ caturaśītkoṭivarṣasahasrāṇi* ‘four times eighty four thousand *koṭis* of years’ or in Mv.1.107.9 *avaivartikāḥ cakravartirājyāni* ‘[the Bodhisattvas] who do not lapse [and] exercise the sway of universal kings’ is rather common. It is, therefore, important to record both shapes of words in the respective indexes: comparing the same word in both indexes immediately shows the transformations that they have undergone. In the given examples, one can compare the entries and easily see in what position the final visarga undergoes sandhi operations. If both indexes were completely identical, their usefulness for understanding the morphophonemic processes in Buddhist Sanskrit would be reduced since these processes are not automatized to the same extent as in standard Sanskrit.

IV. Thus the indexes presented here include the grammatical information in two ways. Since in the serial indexes the word-final positions appear with no effect of sandhi, the morphological characteristics of these words are shown as they usually are in the lexical lists, but with markers of grammatical categories as they appear in the texts, e.g., Bhu-P 6.2 *devamanuṣyeṣu* (LocPl marked by *-su* becoming *-ṣu* by internal sandhi effect). The reverse indexes quote the words with the word-finals modified in accordance with the requirements of adjustment of juxtaposed words within the utterances (sandhi or asandhi, respectively).

In consequence, since our indexes are not meant as mere word lists, they account for the distributional facts of primary importance for the phonetics of Buddhist Sanskrit of the indexed texts. More precisely, the reverse indexes show the morphophonemic changes in words as well as their

morphological characteristics, given that the phonetic adjustments of words are strongly dependent on the latter characteristics (This topic is more fully developed in the forthcoming monograph, see footnote 6).

It is worth emphasizing that this language has elaborated a particular treatment of the pause and of sandhi. The linguistic meaning of the pause in Sanskrit is that it is a correlative of sandhi, which is to be considered as a strategy opposite to, or the inverse of, the pause.<sup>10</sup> In standard Sanskrit, the pause separates the words within utterances from each other to signify “either the final place in the sentence or line” (Allen, p. 99, on the Sanskrit expression *avasāne* translated as ‘in terminal position’ or, more technically, as ‘before pause’; other cases where the pause is required are described in Renou, *Grammaire sanscrite*, p. 30). But, as Renou puts it (ibid., p. 31), “dans le cours normal de la phrase, les finales sont soumises au traitement plus complexe du sandhi”.

In Buddhist Sanskrit of the texts we indexed, the pause fulfils the function of marking the word boundaries with no less efficiency than the sandhi rules, but the mutual involvement of word-finals and the grammatical information they carry is particular. If the textual function of sandhi is “to signal the text cohesion”,<sup>11</sup> that is certainly true in a general cross-linguistic view, and this applies likewise to standard Sanskrit. In our language, however, even if sandhi achieves the text cohesion on the *phonetic level*, the grammatical characteristics of the conjoining words are frequently distributed in a specific way.

To illustrate this point by adding to examples provided in the introductory remarks to the published indexes (footnote 1), recall that sandhi is applied mostly between the full-fledged words and the light words, such as various enclitics, pronouns, postpositions, adverbs, etc. Consider, for example, BhīVin § 5 *bhagavāms tena* and BhīVin § 8 *bhagavān tena*. In standard Sanskrit the final nasal *-n* before the voiceless *t-* is

<sup>10</sup>It suffices to read L. Renou, *Grammaire sanscrite*, Paris: Adrien Maisonneuve, 1975, pp. 30-31. For a recent theoretical assessment of the sandhi and the pause in Sanskrit, see E.O. Selkirk, ‘Prosodic domains in phonology: Sanskrit revisited’ in *Juncture*, ed. by M. Aronoff and M.-L. Kean, Saratoga, CA: Anma Libri, 1980, pp. 107-129 (in spite of its dogmatic generativist inspiration, this article contains some useful observations).

<sup>11</sup>H.Andersen, *Sandhi Phenomena in the Languages of Europe*, Berlin-New York-Amsterdam: Mouton de Gruyter, 1986, p. 5.

replaced by anusvāra plus *s*, but in Buddhist Sanskrit this applies (if it applies) only to isolated occurrences. In the first occurrence, indeed, the function of anusvāra plus the intrusive *s* is to point to the neighboring word boundary followed by *t* and, at the same time, to help achieve the morphological identification of the word *bhagavān* as the NomSgm following the sandhi rules, identical to those in standard Sanskrit. But in the second occurrence, this word is in the form not affected by sandhi (in its form *in pausa*) and is equal to its lexical form. Its morphological identification is to be performed by considering that the pause between the two words is the corollary of non-application of sandhi.

This, by the way, brings about a momentary functional equivalence of the word in the NomSgm and its lexical form where, by definition, no case morpheme may occur.

In BhīVin § 5 *addasāsi khu anyataro bhikṣur mahāprajāpatim* [...] *rudamānīn tiṣṭhantīm pādāṅguṣṭhena bhūmiṃ vilikkentīm* ‘A monk saw Mahāprajāpati [who was] standing, weeping [and] scratching the earth with her great toe’, both the present middle participle *rudamānīn* and the present active participle *tiṣṭhantīm* stand in the AccSgf. The final syllable *-in*<sup>12</sup> of *rudamānīn* results from the final nasal *-m* (or, better, *-ṃ* as in Buddhist Sanskrit the forms *in pausa* end commonly in anusvāra) becoming *-n* before *t*. BhīVin § 163 *tvam ekākinī nadīm tarasi* ‘[when] you cross the river singly’ is so read by G. Canevascini, the editor of the electronic text of BhīVin, but G. Roth’s edition reads *nadīn*, and that reading fits in the set of similar collocations exhibiting the AccSgf *nadīm* before *t*, viz., BhīVin § 163 *nadīn taritum*, *nadīn tare*, *nadīn tareya* (G. Canevascini reads similarly in all these cases). Although the phonetic cohesion is achieved by sandhi, the grammatical information carried by the terminal morpheme *-iM*<sup>13</sup>>*-in* is obscured in all quoted occurrences. *-in* appears indeed, on the one hand, in the AccSgf *nadīm* > *nadīn* (as in *rudamānīm* > *rudamānīn*), and, on the other hand, it marks the AccPlf in BhīVin § 102 *bhikṣuṇī-y-ovādako* [...] *rathyāyaṃ bhikṣuṇīn paśyati* ‘the nun’s advisor sees the nuns on a road’. The only possibility of recognizing the AccSgf in both participles in BhīVin § 5

<sup>12</sup> Within this syllable, the long vowel *-ī* marks the feminine gender and the nasal *-n* marks the AccSgf.

<sup>13</sup> I use *M* as a conventional symbol for the terminal *m/ṃ* as suggested by Allen, *ibid.*, p. 80.

*rudamānīn tiṣṭhantīm* is to disambiguate the morpheme *-in* and to consider the syntactic closeness of both words. These happen, as it were, to have one common inflectional morpheme at the end of the second word that provides the key for the grammatical identification of the whole sequence. In other words, the phonetic cohesion by sandhi makes the participle *rudamānīn* contain only one signifying and unambiguously informative morph (or segmental sign<sup>14</sup>), the long *-ī* marking the feminine gender, whereas its final nasal *-n* is a morph common to at least two inflectional endings.

The application of sandhi seems to depend on the word categories and sometimes on the lexical categories. In the verbs, for example, the 3SgOpt in *-yāt* will be affected by sandhi in two ways. On the one hand, Bhu-P 25.7 reads *dadyāt* #, and Bhu-P 21.20 reads *dadyād anyatra* (from the root *dā-* ‘to give’). The former form ends in the voiceless dental before pause, the latter ends in the voiced dental before a vowel.<sup>15</sup> But their final consonant is dropped before the voiced consonant Mv.1.80.6 *yo dadyā jambudvīpaṃ* ‘he who would present [the Daśabalas] with Jambudvīpa’, and it is kept before *r-* Mv.1.80.8 *yaś ca [...] dvīpāṃ dadyāt ratnācitām* ‘he who should give the [four] continents with their heaps of jewels’. Likewise, the 3SgOpt from *as-* ‘to be’ is Mv.1.199.9 (vs) *syā [nārīye]*, Mv.1.286.13 *syāt [khalu]*, BhīVin § 71 [*tāsām ...*] *syād [brahmacaryāntarāyaḥ #]* ‘[there could be] a hindrance to their chastity’. Further, the 3SgImpf from *vad-* ‘to speak’ is Mv.1.305.9 (vs) *avadad [rājā ]* ‘so spoke the king’ or Bhu-P *avadat* #; the 3SgAor from *vac-* ‘to speak’ is either Mv.1.35.10 *avocat* # or Mv.1.61.8 [*idam*] *avocad [bhagavān]* ‘so spoke the Blessed One’ or Mv.1.306.19 (vs), 2.37.12 (vs) *avoca [nivedaye]* ‘he said “Let know!”’. The difference between these two lexically distinct verbs of speaking is enhanced by their variable sensitivity to sandhi along with the tense category: the preterital forms from *vad-* do not drop the final consonant in whatsoever environment, while the aorist forms from *vac-* may drop it.

The point I wish to emphasize is that, if compared to the

<sup>14</sup>On morphs as elementary segmental signs, see I.A.Mel’čuk, *Towards a Language of Linguistics*, München: Fink, 1982, p. 63.

<sup>15</sup>In the serial index the entry is *dadyāt* Bhu-P 21.20, 25.7, but in the reverse index, *dadyāt* 25.7 and *dadyād* 21.20 appear separately. For other quoted forms, the reader is invited to glance at the appropriate pages of the index to Bhu-P.

sandhi strategy in standard Sanskrit, Buddhist Sanskrit sandhi functions as a morphophonemic variation of the words, while its “obstructive” force<sup>16</sup> lies less in the domain of pure phonetics than—more often—in the domain of the morphological ambiguity (although this is not demonstrated in the verb forms quoted above in order to show the variability in sandhi). Indeed, in BhīVin § 5 *bhagavāṃs tena*, the intrusive consonant *-s* has no communicative value other than delimitative with regard to the word boundary. Also, should the sequence BhīVin § 5 *rudamānīn tiṣṭhantīm* take on a standard Sanskrit shape, its first word would end in anusvāra, while its second word would end in *-m* (if before #) or anusvāra in accordance with the initial *p-* of the following word. The grammatical information would be shown then in both words individually and redundantly. Andersen’s cross-linguistic remark is that “sandhi rules may be understood as a way of utilizing redundancies in lexical representations for the production of subsidiary signs of cohesion” (Andersen, *ibid.*, p.7). It should be added that Buddhist Sanskrit sandhi, even randomly applied and achieving the phonetic cohesion, helps remove the redundancy at the price of special distribution of the morphological marks. The morphological and syntactic patterns can be interpreted only by retrieving the properties of word sequences from the overall properties of the utterances in which they appear, as well as from the overall semantic information in texts.

Under the views presented here, the technique employed for indexation in our indexes is meant to enhance their descriptive potential and, hopefully, to make them useful for lexicographical, philological and linguistic research in Buddhist Sanskrit.

---

<sup>16</sup>I mean that sandhi rules are sometimes perceived as a “peculiar encumbrance which obscures the underlying representations of morphemes, [...] counterproductive from a communicative point of view” (Andersen, *ibid.*, p.7).