The Pali Text Society's *A Dictionary of Pāli*

Rupert Gethin and Martin Straube

Part 1: A brief history of dictionaries of Pali
Part 2: A dictionary of Pali in a digital age

Part 1: A brief history of dictionaries of Pali

Rupert Gethin
(University of Bristol and President of the Pali Text Society)

The first modern Pali-English dictionary was compiled by Robert Childers and published in two volumes in 1872 and 1875. Its starting-point was the twelfth-century Pali dictionary of Moggallāna, the *Abhidhānappadīpikā* (which itself was based on the Sanskrit *Amarakośa*). Childers' dictionary was based on a very limited number of Pali texts (some only in manuscript).[1] In the decades following the founding of the Pali Text Society by T. W. Rhys Davids in 1881 the publication of Pali texts in European editions greatly increased. The effect of this was that Childers' pioneering dictionary became outdated and was no longer an adequate tool for the study of the available Pali literature. The Pali Text Society's plans for a Pali-English dictionary go back to at least the first decade of the twentieth century.[2] William Stede recounts how the dictionary was first conceived as a collaborative work involving a team of international scholars. The original projected publication date was 1905, but by 1909 only one eighth of the work had been completed. A revised plan for the dictionary was set out by Rhys Davids in full in the "Report of the Pali Text Society for 1909".[3] Four volumes were planned, representing the work of seven scholars. In the event the project met with further delays, while the outbreak of the First World War in 1914 made conditions for international scholarly collaboration unfavourable. In 1916 Rhys Davids handed overall responsibility for the compiling and editing of the dictionary to Stede.[4] *The Pali Text Society's Pali-English Dictionary* was eventually published in eight parts (arranged in four volumes) between 1921 and 1925. The dictionary made use of material from Rhys Davids' personal interleaved and annotated copy of Childers' dictionary, supplemented with material from Hendrik Kern and Edmond Hardy, as well as material on specific letters from Mabel Bode (*b-bh*), Sten Konow (*s*, and *h*), Professor Duroiselle (*k*), and E. J. Thomas (*ñ*). In her "Report of the Pali Text Society for 1926", Mrs Rhys Davids (who had become President of the PTS following the death of her husband in 1922) noted that "the costly task of bringing out the Dictionary, without

---

[1]   Robert Childers, *A Dictionary of the Pali Language* (London: Trübner, 1875), p. xxii.

[2]   "Afterword", in *The Pali Text Society's Pali-English Dictionary*, Part VIII (Vim-H), ed. by T. W. Rhys Davids and William Stede (London: Pali Text Society, 1925), pp. 199-203 (p. 200).

[3]   *Journal of the Pali Text Society* (1909), pp. pp. vi–x.

[4]   Stede, "Afterword", p. 199.

withholding the issue of texts and translations, had been achieved without insolvency, thanks to the generosity of donors, mainly of Japan".[5]

The first edition of *The Pali Text Society's Pali-English Dictionary* was published fifty years after Childers' dictionary and was a quarter of a century in the making. In his foreword to the first part of the dictionary Rhys Davids commented: "This work is essentially preliminary." He went on say that, while he regarded the dictionary as a significant advance in Pali lexicography, it was far from perfect; he thus looked forward to "the eventual issue of a second edition which will come nearer to our ideals of what a Pali Dictionary should be" (p. vii). In effect the Pali Text Society has been trying to realise this second edition ever since.

In her "Report of the Pali Text Society for 1926", Mrs Rhys Davids also noted the publication (in 1925) of the first fascicle (*a-ajja*) of *A Critical Pāli Dictionary* edited by Dines Andersen and Helmer Smith on the basis of the pioneering work in Pali lexicography begun by V. Trenckner, who had made transcripts, with critical notes, of a number of Pali texts from the collection of Pali manuscripts brought back to Copenhagen in the 1820s by Rasmus Rask.[6] In the first decade of the twentieth century it was envisaged that Trenckner's material could be included in Rhys Davids' plans for the PTS dictionary, but when those plans stalled Andersen and Smith continued with the work of redacting Trenckner's materials in Copenhagen and in 1916 planned to produce a complete Pali dictionary, estimating that the task would take "at least fifteen years".[7] But by 1932 only four fascicles had been published covering the words *a-anuṭīkā* (equivalent to about 5% of the Pali lexicon). Clearly the work involved in producing a Pali dictionary had once again been underestimated. In her review of fascicles 2-4, Mrs Rhys Davids, with reference the PTS's dictionary and the Copenhagen dictionary's use of the word "critical" in its title, commented "perhaps the only way, in which the younger sister may deserve her quaint title is by going one better every time, and in the latter feature I think she does".[8] This once again acknowledged the preliminary nature of the PTS dictionary and the need for work to continue in order to produce a better dictionary. However, work on the production of *A Critical Pāli Dictionary* progressed only slowly. By 1948 the first volume was completed covering only the first letter of the Pali alphabet (*a*) and ending with the word *ahosi-kamma*. The deaths of both Andersen (1940) and Smith (1956) meant that new editors were needed and the production of the second volume was conceived as an international project. The first fascicle was published in 1960 and the seventeenth and final fascicle thirty years later in 1990.

The first two volumes covered all words beginning with vowels. The third volume of *A Critical Pāli Dictionary* (covering *k-kāretu-kāma*) was published in eight fascicles between 1992 and 2011 with contributions from various scholars and under the general editorship of Ole Pind and

---

[5]　*Journal of the Pali Text Society* (1924-1927), pp. 15-25 (p. 15).

[6]　Dines Andersen, "A Pioneer in Pāli Lexicography", *A Critical Pāli Dictionary: Volume I*, ed. by Dines Andersen, Helmer Smith and Hans Hendriksen (Copenhagen: The Royal Danish Academy, 1925-1948), pp. iv-viii.

[7]　Dines Andersen and Helmer Smith, "Preface", *A Critical Pāli Dictionary: Volume I*, ed. by Dines Andersen, Helmer Smith and Hans Hendriksen (Copenhagen The Royal Danish Academy, 1925-1948), pp. ix-xi.

[8]　C. A. F. Rhys Davids, "A Critical Pāli Dictionary", *Journal of the Royal Asiatic Society* 65 (1933), pp. 435-37.

Oskar von Hinüber. The three completed volumes cover about one third of the Pali lexicon. By the end of the 1970s it had become clear that the financial resources to fund the continuation of *A Critical Pāli Dictionary* were uncertain: the Carlsberg Foundation, the Danish Research Council and UNESCO were unable to support the project beyond the completion of the second volume. The production of volume three was only achieved with generous support from a variety of sources in Europe and Japan, but the resources necessary to sustain the project into the future and to completion became increasingly uncertain.[9]

Meanwhile *The Pali Text Society's Pali-English Dictionary* had been reprinted four times (1949, 1952, 1959, 1966). By the beginning of the 1970s the slow progress of *A Critical Pāli Dictionary* made it apparent that the only way a complete improved Pali dictionary was likely to be produced in the foreseeable future was by undertaking a revision of the PTS dictionary. In 1972, when the 1966 reprint of the dictionary had sold out, the Council of the PTS decided to reprint for a fifth time to meet demand for the dictionary for the next ten years, while asking K. R. Norman to prepare a revised second edition. Like his predecessors in Pali lexicography, Mr Norman discovered that he had underestimated the work involved and that expecting to be able to complete such a revision in his spare time while carrying the burden of university teaching and administration was unrealistic. Thus when in 1981, Miss I. B. Horner, who had been President of the Pali Text Society since 1959, died and left a portion of her estate to the Society, it was decided to use some of these funds to employ a full-time research assistant to help with the revision of the Dictionary. Dr Margaret Cone was appointed to this position in 1984.[10] The original plan was for her to revise the PTS's 1925 dictionary and produce a second edition, but it soon became apparent that so little of the original dictionary would remain unaltered that what was actually being undertaken was the production of a completely new dictionary. This was to become *A Dictionary of Pāli*.[11]

With the retirement in 1992 of Professor Norman from his post at the University of Cambridge, the Pali Text Society entered into new negotiations with the University and from that date funded the post of an Assistant Director of Research in Pali Lexicography, which was again held by Dr Margaret Cone. Writing in 1995 Professor Norman expressed the hope that Dr Cone's dictionary might appear "early in the next decade".[12] Once again the task involved was underestimated. In the event only the first part of *A Dictionary of Pāli*, covering the letters *a-kh*, was published in 2001. With the withdrawal of the support of the Royal Danish Academy for *A Critical Pāli Dictionary* and further funding no longer forthcoming, this project was brought formally to a close in 2007 nearly ninety years after it was first conceived. If it had not been clear before it was now apparent that all efforts should be focused on making sure that Dr Cone's new *A Dictionary of Pāli* would reach completion.[13]

---

[9] See Oskar von Hinüber, "Concluding Remarks", *A Critical Pāli Dictionary: Volume III*, ed. by Oskar von Hinüber and Ole Holten Pind (Bristol: Pali Text Society, 2011), pp. xxxiii-xxxvi.

[10] Norman, "A Report on Two Pāli Dictionaries", *International Journal of Lexicography*, 8 (1995), 115-125 (pp. 118-19).

[11] Margaret Cone, *A Dictionary of Pāli*, Part I (Oxford: Pali Text Society, 2001), p. viii.

[12] Norman, "Report on Two Pāli Dictionaries", p. 119.

[13] von Hinüber, "Concluding Remarks", p. xxxiv.

Progress on this dictionary has continued steadily. Part two, covering the letters *g-n*, was published in 2010. The two published volumes, together amounting to some 1,500 pages, cover approximately half of the Pali lexicon. Dr Cone retired from her position at the University of Cambridge in 2013, but the Pali Text Society has continued to provide her with financial support while she works on completing part three of the dictionary, covering the letters *p-bh*, which the Society expects to publish in 2019.

Having devoted thirty-five years to Pali lexicography and the production of the new *A Dictionary of Pāli*, Dr Cone will finally retire from the project with the completion of the third volume, leaving about one quarter of the Pāli lexicon still to be covered. With a view to making sure that this dictionary project does not stall as it nears completion the Council of the Pali Text Society made the decision to fund Dr Martin Straube to begin work on the fourth and final part from January 2018. Dr Straube comes to the project with some years' experience working as a research assistant on the *Sanskrit-Wörterbuch der buddhistischen Texte aus den Turfan-Funden und der kanonischen Literatur der Sarvāstivāda-Schule* (*Sanskrit Dictionary of the Buddhist Texts from the Turfan Finds and of the Canonical Literature of the Sarvāstivāda School*) at the Göttingen Academy of Sciences and Humanities (Akademie der Wissenschaften zu Göttingen), headed by Professor Jens-Uwe Hartmann. For much of 2018 Dr Cone and Dr Straube are working on the dictionary in tandem, allowing Dr Straube to learn Dr Cone's methodology as he begins work starting with words beginning with the letter *m*.

With the Pali Text Society's original dictionary approaching its centenary and *A Critical Pāli Dictionary* abandoned, the new *A Dictionary of Pāli* will be an indispensable scholarly resource for the reading of Pali texts for the foreseeable future. The new dictionary incorporates and revises material from its predecessors, but also encompasses the vocabulary of the wide range of Pali texts, especially the commentaries, published since 1925, as well as a century of scholarship that has advanced our understanding of Pali usage.[14] Something of the development of Pali lexicography can be seem from a comparison of the articles for *accharā* in the four dictionaries discussed (fig. 1).

Like *The Pali Text Society's Pali-English Dictionary* and *A Critical Pāli Dictionary* before it, *A Dictionary of Pāli* has benefited from support from Japan. The assistance of Dr Nishi Yasutomo and Professor Ousaka Yumi along with the support of Dr Kawamoto Koichi, Director of Chuo Academic Research Institute, led to the award of a grant from Chuo Academic Research Institute of Rissho Kosei-kai for 2018; this has been crucial in facilitating the recent progress of the project as Dr Cone hands over to Dr Straube. The records of Pali lexicography from the last hundred years are full of over optimistic estimations of the time needed to complete lexicographical projects. I thus hesitate to burden the current project with yet another. Nonetheless it seems not unreasonable to hope that the Pali Text Society's *A Dictionary of Pāli* will come to completion in the next ten to fifteen years.

---

[14]　See Margaret Cone, "The I. B. Horner Lecture 1995: Lexicography, Pāli and Pāli Lexicography", *Journal of the Pali Text Society*, 22 (1996), 1-34; Margaret Cone, "Caveat Lector", *Journal of the Pali Text Society*, 29 (2007), 95-106.

1. Childers, *A Dictionary of the Pali Language* (1875)

ACCHARĀ (*f.*), A moment, the snapping of a finger, the twinkling of an eye [चपर]. Ab. 66.

2. Rhys Davids and Stede, *The Pali Text Society's Pali-English Dictionary* (1921)

Accharā[1] (f.) [etym. uncertain, but certainly dialectical; Trenckner connects it with ācchurita (Notes 76); Childers compares Sk. akṣara (see akkhara); there may be a connection with akkhaṇa in akkhaṇa-vedhin (cp. BSk. acchaṭā Divy 555), or possibly a relation to ā + **tsar**, thus meaning "stealthily", although the primary meaning is "snapping, a quick sound"] the snapping of the fingers, the bringing together of the finger-tips: 1. (lit.) **accharaṃ paharati** to snap the fingers J II.447; III.191; IV.124, 126; V.314; VI.366; DhA I.38, 424. — As measure: as much as one may hold with the finger-tips, a pinch J V.385; DhA II.273 (°gahaṇamattaṃ); cp. ekacchara-matta DhA II.274. — 2. (fig.) a finger's snap, i. e. a short moment, in ek'acchara-kkhaṇe in one moment Miln 102, and in def. of acchariya (q. v) at DA I.43; VvA 329.
  -saṅghāta the snapping of the fingers as signifying a short duration of time, a moment, °*matta* momentary, only for one moment (cp. BSk. acchaṭāsaṅghāta Divy 142) A I.10, 34, 38; IV.396; Th I, 405; 2, 67 (expld. at ThA 76 as ghaṭikāmattaṃ pi khaṇaṃ aṅgulipoṭhanamattaṃ pi kālaṃ). **-sadda** the sound of the snapping of a finger J III.127.

3. Andersen and Smith, *A Critical Pāli Dictionary* (1924)

[1]acchara, *f.* (*comp. also* acchara-; *Amg* acchara; acchaṭā Vyu 138,42; *etymol. unknown*; Tr PM 76 *compared sa.* ācchurita, *n.* = nakhavādya, *noting v. l.* acchurā- Thī 67; *the* Aṭṭhakathā *derived* accariya *from this, see* accharāyogga), **1.** *a snapping of the fingers* (*like to sa.* pucchaṭī, mukuṭī, mucuṭī), Abh 66; *in the phrase* ~aṃ paharati, (**a**) *as signal of command*: Ja IV 336,3 (*to a peacock*); IV 438,5 (*to dogs*); Ps III 153,6 (*to a horse*); — (**b**) *expressive of reprimand*: Ja II 447,28; IV 124,20; Dhp-a I 38,4; III 8,22; 414,6; — (**c**) *do. of dismissal or refuse*: Ja III 191,21; V 314,14; VI 542,7; Dhp-a I 424,2; — (**d**) *do. of disregard or contempt*: Ps II 524,5; — (**e**) *do. of satisfaction or joy*: Ja VI 336,25. — **2.** *the two or three fingers by which a pinch is taken*, ~āya gaṇhitvā (gahetvā), Dhp-a III 19,10-14 (*cf. ib.* 18,9: tīhi aṅgulīhi gahetvā). *Cf.* accharā-gahaṇa. — **3.** *a pinch*, ~aṃ sakkharāya (*of sugar*), Ja V 385,19. — *Ifc. v.* ekacchara-kkhaṇa, ekacchara-matta.

4. Cone, *A Dictionary of Pāli* (2001)

accharā[1], *f.* [*cf AMg* accharā, *BHS* acchaṭā; *Trenckner (Notes 76) connects with* ācchurita], **1.** *snapping of the fingers* (~aṃ paharati; *as a gesture of command; of annoyance, refusal, contempt; also as a gesture of pleasure*); Ja II 447,28 (bodhisatto ~aṃ paharitvā); IV 124,20 (rājā ~aṃ paharitvā nassa vasali ... ti tajjesi); 336,5 (~aṃ paharitvā moriṃ vassāpesi); Sv 43,16 (°-yoggan ti acchariyaṃ, ~aṃ paharituṃ yuttan ti attho); Ps II 389,17 (kiṃ tvam etthā ti ~aṃ pahari, so ṭhātuṃ asakkonto tatth' eva antaradhāyi); III 161,8 (daharo ... ~aṃ pahari, asso āgantvā ... bhattaṃ bhuñji); Spk I 293,32 (ekā pi gāyi ekā pi nacci ekā pi ~aṃ pahari); Cp-a 213,1 (tvaṃ ito aññattha yāhī ti tassa ~aṃ pahari); — **2.** *bringing together the finger-tips (to take a pinch)*; Spk I 99,16 (~āya gahitaṃ pi nakhena phāletvā parassa datvā); Dhp-a III 19,10 (asukena nāma ~āya gaṇhitvā taṇḍulādīni dinnāni); — **3.** (*as measure*) *a pinch*; Ja V 385,19 (~aṃ sakkharāya ... dehi); — **4.** *a finger's snap, a short moment*; Abh 66 (khaṇo das' ~ā); Paṭis-a 317,34 (°-mattaṃ vā dasaccharāmattaṃ vā khaṇaṃ samāpattiṃ ṭhapetuṃ); — °-**kkhaṇa,** *m.* [accharā + khaṇa[1]], *the time of a finger's snap, a moment*; Spk II 99,30 (ekasmiṃ hi ~e anekāni cittakoṭisatasahassāni uppajjanti); — *ifc see* ek'-; — °(a)-**ggahaṇa,** *n.,* *a pinch*; Dhp-a II 273,14 (~-mattaṃ siddhatthakaṃ laddhuṃ vaṭṭati); — °-**saṅghāta,** *m.,* *a finger-snap, a moment*; A I 10,19 (~-mattaṃ pi ce ... bhikkhu mettacittaṃ āsevati; Mp I 70,17: accharāpaharaṇamattaṃ, dve aṅguliyo paharitvā saddakaraṇamattan ti attho); Th 405 (~-mattaṃ pi cetosantiṃ anajjhagaṃ) ≠ Thī 67; Mil 142,6 (~-mattaṃ pi).

Fig. 1 Comparison of articles for *accharā*

Part 2: A dictionary of Pali in a digital age

Martin Straube

(Research Fellow in Pali Lexicography, Pali Text Society

Associate research fellow, Indologie und Tibetologie, Philipps-Universität Marburg)

Today, dictionaries are accessed in different ways. While printed books still remain important, the number of online versions is increasing rapidly, as is the quality of their content in terms of faithfulness to their sources. Most scholars in the field of Indian and Buddhist studies are used to consulting the main Sanskrit, Pali and other dictionaries via searchable online versions. However, the great majority of these online versions is based on books that have been printed long before the use of computers and the rise of the internet. To convert the content of a printed dictionary into a format that can be accessed by a data base is a time consuming and expensive task. And, what is more, the conversion, which in most cases is a multi-stage process, involves changes to the data contained in the source, so that the output is fraught with flaws and defects. At present there exist very few online versions of important dictionaries pertaining to Indian studies that are faithful in the sense that they could be relied upon or quoted without having recourse to the printed original.

When I was entrusted with the continuation of *A Dictionary of Pāli* (DOP) at the beginning of 2018, one of the key questions that arose was that of an appropriate format for the files in which the lexicographical data will be saved. As a matter of course, DOP will be published in printed form, but an online version appears highly desirable too. Thus, the file format should be suited to produce both a printable version and an output that can be used for an online version; moreover, it should be in a common, simple and open standard. The most natural solution for this is obviously XML (Extensible Markup Language). As is well known, XML has been designed to store and share textual data (mainly on the internet) by defining a set of rules for its encoding. As such XML makes no specifications on the types of texts to be stored, nor does it provide any scheme of how a specific type of text may be structured. In this regard it is completely open which allows it to be adapted to many different needs. However, a widespread and generally accepted standard for textual encoding that conforms to XML exists with the Text Encoding Initiative (TEI).[15] The TEI makes "recommendations about suitable ways of representing those features of textual resources which need to be identified explicitly in order to facilitate processing by computer programs. In particular, it specifies a set of markers (or tags) which may be inserted in the electronic representation of the text in order to mark the text structure and other features of interest."[16] To this end encoding schemes for different types of texts are provided, among them also a rather detailed one for dictionaries (TEI Guidelines, chapter 9). Here TEI gives recommendations for probably most of the existing types of dictionaries and their different structural elements. However, in many cases it allows two or more different solutions for the same problem. The reason for this is, as stated in the introductory part of chapter 9 of the Guidelines, that "[b]oth typographically and structurally,

---

[15]　See *www.tei-c.org*. The current version is TEI P5. All references are to the Guidelines of this versions. The Guidelines also contain a chapter (iv.2) on the "Historical Background."

[16]　TEI Guidelines, chapter iv.

print dictionaries are extremely complex" which makes a flexible encoding scheme necessary.

The encoding for DOP presented here conforms to the recommendations set out in the TEI Guidelines. Furthermore, it has been designed in consultation with the Cologne Center for eHumanities (CCeH) at the University of Cologne. Among the CCeH data files are those of *A Critical Pāli Dictionary* (CPD) that are in the charge of the Pali Text Society (PTS); these are presently edited along the TEI Guidelines, with the goal to create a reliable online version of this important dictionary.[17] Consultations with the persons in charge of CPD at the staff of the CCeH proved to be useful in many ways. As CPD a future online version of DOP could also eventually be maintained by the CCeH, thus care has been taken to adjust fundamental structural principles of the scheme for DOP so that they match those used for CPD. This pertains first of all to the way subentries are nested within the main article (see below). But beyond specific adjustments between the schemes for CPD and DOP, the expertise and experience of the staff of the CCeH in questions related to TEI encoding has been and still is of immense help. Since DOP has its own specific structure it needs its own encoding scheme in many minor points. However, it has been possible to adjust the main and fundamental principles to that of CPD scheme. At present, the encoding scheme for DOP is fairly complete for articles on nouns. Articles on verbs are not fundamentally different, but have some peculiarities. When working on the scheme I have tried to keep one eye on articles on verbs. But, since I have not yet written an article on a verb, it may be necessary to adjust or add some minor points later in order to finally have a scheme that matches all kinds of articles in DOP. In what follows I will try to briefly expound how the encoding scheme for DOP works at present.

To map DOP onto a scheme that conforms to the TEI Guidelines it is necessary first to see how DOP is structured. An article in DOP can be divided into seven parts. Not all parts may be present in each article, and some parts may recur within the same article:[18]

1. The headword with possible orthographic variants.

2. Grammatical information: part of speech, gender of nouns, declension, conjugation and the like.

3. An explanation of the form of the word: in most cases parallels in Sanskrit or a Middle Indic dialect, references to scholarly literature if necessary.

4. Meaning and usage of the headword. This can be divided in two parts:
   4.1 An English definition that gives the meaning.
   4.2 Quotations in the original Pali that illustrate the given meaning.

5. Cross references to related articles outside the article under discussion.

---

[17] See *http://cpd.uni-koeln.de* for a first yet not perfect version.

[18] For the following compare Cone, "Lexicography, Pāli and Pāli Lexicography", pp. 26 ff.; Cone elaborates on the lexicographical aspects, which is not my main concern here. If the structuring given by me differs from Cone's, this is only due to my task of mapping the structure of DOP onto a TEI scheme. However, this does not involve any substantial changes to the dictionary's design.

6. Any other remarks: In most cases misprints or wrong readings related to the headword that occur in text editions are noted here, but any other information can be given.

7. Related entries. If the headword is a noun: entries for compounds with the headword as first member, as well as forms with *a-*, *su-*, *du(r)-*, *ni(r)-*. If it is a verb: entries for participles, absolutives, and secondary verbal forms, e.g. causatives.

The dictionary scheme of TEI provides adequately defined elements for each of these parts. To illustrate this, I will take a rather short article on Pali *makula* to demonstrate how it can be mapped onto TEI:

The outermost structural level of an entry is provided by the <entry>[19] element. This encloses all the information related to one single article of DOP. In the case of homonyms an attribute in the form of a number is added to the <entry> element in order to distinguish it from other entries with a headword of the same form. In our case the number is "1":

> <entry n='*1*'>
> ...
> </entry>

1. The <form>[20] elements groups the information on the forms of the headword. The TEI allows information on the written form as well as on pronunciation, hyphenation, and the like. Since only the written form of the headword is relevant to DOP the content of <form> is confined to instances of <orth>[21] elements. Many entries will contain only one <orth> element, but, since orthographical variants of the same word frequently occur in the Pali texts, more than one <orth> element can be given as headword:

> <form>
>   <orth>**makula**</orth>
>   <orth>**makuḷa**</orth>
> </form>

---

[19] "<entry> contains a single structured entry in any kind of lexical resource, such as a dictionary or lexicon." (TEI Guidelines)

[20] "<form> (form information group) groups all the information on the written and spoken forms of one headword." (TEI Guidelines)

[21] "<orth> (orthographic form) gives the orthographic form of a dictionary headword." (TEI Guidelines)

2. Next follows the <gramGrp>[22] element. Any grammatical information relating to the headword is given here:

> <gramGrp><abbr>**n**</abbr> **and** <abbr>**iic**</abbr></gramGrp>

The grammatical abbreviations "n" (neuter) and "iic" (*in initio compositi*, "at the beginning of a compound") are each enclosed in an <abbr>[23] tag. This is done (a) to define them as abbreviations of technical terms, and (b) to be able to give them any desired shape in the output. The printed output will be "n." and "iic", according to the list of abbreviations given in Part I of DOP, but any other shape may be chosen for another output.

3. The <etym>[24] element contains explanation of the form of the headword; in our case it is simply stated that the parallel for Pali *makula* in Sanskrit is *mukula*:

> <etym><abbr>**Skrt**</abbr> <w xml:lang='*sa*'>**mukula**</w></etym>

As in the case of grammatical abbreviations "Skrt" is enclosed by <abbr> tags; the printed output will be "S." The quoted word *mukula* is enclosed by <w>[25] that here serves as a kind of placeholder for the TEI attribute xml:lang. This attribute defines the language of the given word or phrase within a standard system of language assignments which in our case is "sa" for Sanskrit.

4. Each meaning of the headword is grouped by one <sense>[26] element. As in the case of <entry>, multiple meanings are differentiated by assigning numeral attributes.

4.1. The English definition is enclosed by <def>.[27] Since our headword has different grammatical characteristics in its several meanings a <gramGrp> element recurs here containing the information that *makula* is n. when meaning "a bud":

> <sense n='*1*'>
> <gramGrp><abbr>**n**</abbr></gramGrp>
> <def>**a bud**</def>
> …
> </sense>

4.2. The definition is followed by quotations that illustrate the given meaning. Each quotation

---

[22] "<gramGrp> (grammatical information group) groups morpho-syntactic information about a lexical item [...]" (TEI Guidelines)

[23] "<abbr> (abbreviation) contains an abbreviation of any sort." (TEI Guidelines)

[24] "<etym> (etymology) encloses the etymological information in a dictionary entry." (TEI Guidelines)

[25] "<w> (word) represents a grammatical (not necessarily orthographic) word." (TEI Guidelines)

[26] "<sense> groups together all information relating to one word sense in a dictionary entry, for example definitions, examples, and translation equivalents." (TEI Guidelines)

[27] "<def> (definition) contains definition text in a dictionary entry." (TEI Guidelines)

is enclosed in a <cit>[28] element which is made up of a <bibl>[29] element containing the bibliographic information, and a <quote>[30] element for the actual quotation:

```
<cit>
  <bibl type='text'><abbr>Thi</abbr> 260</bibl>
  <quote xml:lang='pi'>pattali-<oRef rend='tilde'/>-vaṇṇasadisā sobhate su dantā</quote>
  <note>read <w xml:lang='pi'>sattali-?</w> cf <bibl type='lit'>K.R. Norman, 2007, II, ad loc.</bibl></note>
 </cit>
```

<bibl> has the attribute type='text' to indicate that it contains a reference to a text in Pali, Sanskrit, etc. The abbreviation "Thi" will be "Thī" (= Therīgāthā) in the output. <quote> has the language assignment xml:lang='pi' for Pali. The element <oRef/>[31] provides a reference to the headword. Its attribute rend='tilde' defines that it will be realised as ~ in the printed output. The quotation is followed by a <note>[32] element which is optional and may contain any kind of remark or comment relating to the quotation. In our case it is a reference to a note in K.R. Norman's translation of the *Therīgāthā*[33] that makes a suggestion on how to read and understand the difficult word *pattali*. The actual bibliographic reference is again enclosed by <bibl>, in this case with the attribute type='*lit*' to define it as a reference to secondary, scholarly literature.

Thus, the <sense> element discussed so far looks like:

```
<sense n='1'>
  <gramGrp><abbr>n</abbr></gramGrp>
  <def>a bud</def>
  <cit>
    <bibl type='text'><abbr>Thi</abbr> 260</bibl>
    <quote xml:lang='pi'>pattali-<oRef rend='tilde'/>-vaṇṇasadisā sobhate su dantā</quote>
    <note>read <w xml:lang='pi'>sattali-?</w> cf <bibl type='lit'>K.R. Norman, 2007, II, ad loc.</bibl></note>
```

---

[28]　"<cit> (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example." (TEI Guidelines)

[29]　"<bibl> (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged." (TEI Guidelines)

[30]　"<quote> (quotation) contains a phrase or passage attributed by the narrator or author to some agency external to the text." (TEI Guidelines)

[31]　"<oRef> (orthographic-form reference) in a dictionary example, indicates a reference to the orthographic form(s) of the headword." (TEI Guidelines)

[32]　"<note> contains a note or annotation." (TEI Guidelines)

[33]　K. R. Norman, *The Elders' Verses. II: Therīgāthā*. 2nd edn (Lancaster: Pali Text Society, 2007).

</cit>
</sense>

For the purpose of our sample only one quotation is given here. In fact, several quotations are given for each meaning if the word is attested more than once in Pali texts. In the same way other meanings of the headword *makula* are defined. I skip the second meaning and give here only the third, again with only one quotation:

<sense n='*3*'>
 <gramGrp><abbr>**iic**</abbr></gramGrp>
 <def>"**[being in the state of] a bud", ie not yet unfolded, developed**</def>
 <cit>
  <bibl type='*text*'><abbr>**Vism**</abbr> **245,**<hi rend='small'>**26**</hi>
<abbr>**foll**</abbr></bibl>
   <quote xml:lang='*pi*'>**makkaṭo** <gap rend='*dots*'/> <oRef rend='*tilde*'/>-
tālapaṇṇasūciṃ daḷhaṃ gahetvā</quote>
 </cit>
 ...
</sense>

The reference "Vism" is to a prose text (the *Visuddhimagga*) which requires specification of page and line of the quoted edition. The line number is marked with <hi>[34] and an attribute rend='*small*' that indicate that it will be printed in a smaller type.

5. Cross references to articles on compound words that contain the headword **makula** as a second, third or even later member occur at their appropriate place, i.e. nested in the proper <sense> element. For the sake of illustration it suffices to give here cross references for meaning 1 with **makula** as a second member (ifc). The element used to group those references is <xr>:[35]

  <xr>**ifc see**
   <ref type='*full*'>**kandalamakula**</ref>
   <ref>**kandala-**</ref>,
   <ref type='*full*'>**sumanamakula**</ref>
   <ref>**sumana-**</ref>
  </xr>

The cross references are each marked with <ref>[36]. In DOP cross references are printed as

---

[34] "<hi> (highlighted) marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made." (TEI Guidelines)

[35] "<xr> (cross-reference phrase) contains a phrase, sentence, or icon referring the reader to some other location in this or another text." (TEI Guidelines)

[36] "<ref> (reference) defines a reference to another location, possibly modified by additional text or comment." (TEI Guidelines)

kandala- with the hyphen representing the headword **makula**. This suits the requirements of saving space in a printed dictionary, since the information implied by the hyphen can be supplied by the reader. However, to make an online version fully searchable it is necessary to give the full form of the reference too. Therefore, an additional <ref> tag with the attribute type='*full*' provides the full form **kandalamakula** that can be ignored if desired; the print output will simply skip it.

6. A <note> states that *makulam* printed in the European (i.e. PTS) edition of the *Sammohavinodanī* (*Vibhaṅga-aṭṭhakathā*, Vibh-a) on p. 239, line 31, is a wrong reading for *makulitam*. A cross-reference is given to the relevant article where the reader will find a quotation of the passage and further information:

> <note><w xml:lang='*pi*'>**makulam**</w> **in** <abbr>**Ee**</abbr> **at** <bibl type='*text*'><abbr>**Vibha**</abbr> **239,**<hi rend='small'>**31**</hi></bibl> **is wr for** <w xml:lang='*pi*'>**makulitam**</w>**, see** <ref>**makulita**</ref></note>

7. Next follows a compound entry. DOP does not assign compounds with the headword as its first member to the different meanings of the headword. Instead, compound entries simply follow after the last given meaning in alphabetical order according to the second member. The element used for compound entries is <re>.[37] It can, and, in many cases, actually does contain all the elements of an <entry>; however, it appears not at the outermost structural level, but inside <entry> or even inside another <re>. DOP not only provides compound entries related to the headword (first level), but also to other compound entries (second level). I decided to explicitly mark the level at which a <re> element occurs with an attribute, although structurally this is not necessary, since the level is defined by the nested structure. However, it helps to keep track of the relevant level and to control the output:

```
<entry>
...
 <re type='1'>
 ...
   <re type='2'>
   ...
   </re>
 </re>
<entry>
```

In our case there is only one compound entry on the first level that looks as follows:

```
<re type='1'>
 <form>
   <orth>
```

```
    <choice>
      <w type='full'>makuladāṭha</w>
      <w type='print'><oRef rend='grade'/>dāṭha</w>
    </choice>
   </orth>
  </form>
  <gramGrp><abbr>mfn</abbr></gramGrp>
  <sense>
  <def>having not (yet) developed teeth</def>
    <cit>
      <bibl type='text'><abbr>Ja</abbr> IV 345,<hi rend='small'>21</hi></bibl>
      <quote xml:lang='pi'>ṭhapesi <gap rend='dots'/> tesaṃ anantarā <oRef
   rend='tilde'/>e taruṇasūkare, tesaṃ anantarā mahādāṭhe</quote>
     </cit>

     ...
   </sense>
 </re>
```

Again, attention is given to providing full forms of compounds even though they will not be part of the printed version. DOP gives compounds in the form °-**dāṭha** with the sign ° referring to the headword **makula**. As in the case of cross-references a fully searchable online version requires the statement of the full form of the compound **makuladāṭha**. Therefore, the <orth> element contains a further element called <choice> that "groups a number of alternative encodings for the same point in a text."[38] The <choice> elements in turn contain the full as well as the abbreviated, i.e. printed form of the compound. For the two forms I use the <w> element with appropriate attributes that define the respective type. The print type again contains the <oRef/> element, here with the attribute rend='*grade*' to indicate that it should be rendered with the grade sign °.

If we put all the elements discussed so far together we get the following outline of the article:

```
    <entry n='1'>
     <form>
      <orth>makula</orth>
      <orth>makuḷa</orth>
     </form>
     <gramGrp><abbr>n</abbr> and <abbr>mfn</abbr>
    <abbr>iic</abbr></gramGrp>
     <etym><abbr>Skrt</abbr> <w xml:lang='sa'>mukula</w></etym>
     <sense n='1'>
      <gramGrp><abbr>n</abbr></gramGrp>
      <def>a bud</def>
      <cit>
        <bibl type='text'><abbr>Thi</abbr> 260</bibl>
```

```
    <quote xml:lang='pi'>pattali-<oRef rend='tilde'/>-vaṇṇasadisā sobhate su
dantā</quote>
    <note>read <w xml:lang='pi'>sattali-?</w> cf <bibl type='lit'>K.R. Norman,
2007, II, ad loc.</bibl></note>
   </cit>

  ...
  <xr>ifc see
   <ref type='full'>kandalamakula</ref>
   <ref>kandala-</ref>,
   <ref type='full'>sumanamakula</ref>
   <ref>sumana-</ref>
  </xr>
 </sense>

 ...
 <sense n='3'>
  <gramGrp><abbr>mfn</abbr> <abbr>iic</abbr></gramGrp>
  <def>"being [in the state of] a bud", ie not yet unfolded, developed</def>
  <cit>
   <bibl type='text'><abbr>Vism</abbr> 245,<hi rend='small'>26</hi>
<abbr>foll </abbr></bibl>
   <quote xml:lang='pi'>makkaṭo <gap rend='dots'/> <oRef rend='tilde'/>-
tālapaṇṇasūciṃ daḷhaṃ gahetvā</quote>
  </cit>

  ...
 </sense>
 <note><w xml:lang='pi'>makulam</w> in <abbr>Ee</abbr> at <bibl
type='text'><abbr>Vibha</abbr> 239,<hi rend='small'>31</hi></bibl> is wr for <w
xml:lang='pi'>makulitam</w>, see <ref>makulita</ref></note>
 <re type='1'>
  <form>
   <orth>
    <choice>
     <w type='full'>makuladāṭha</w>
     <w type='print'><oRef rend='grade'/>dāṭha</w>
    </choice>
   </orth>
  </form>
  <gramGrp><abbr>mfn</abbr></gramGrp>
  <sense>
  <def>having not (yet) developed teeth</def>
   <cit>
    <bibl type='text'><abbr>Ja</abbr> IV 345,<hi rend='small'>21</hi></bibl>
    <quote xml:lang='pi'>ṭhapesi <gap rend='dots'/> tesaṃ anantarā <oRef
rend='tilde'/>e taruṇasūkare, tesaṃ anantarā mahādāṭhe</quote>
   </cit>
  </sense>
```

```
    </re>
  </entry>
```

To get the desired output the XML file has to be transformed in a suitable processor with a help of an XSL (Extensible Stylesheet Language) file. This file defines the general kind of output (html or plain text) and gives a set of rules that control the structure of the resulting document. The XSL transformation does not change the original file; rather, it creates a new file based on the content of the existing one. This is important—no information contained in the original file can be lost. To get a printable output of DOP I designed a plain text XSL transformation that creates a LaTeX file. This can immediately be processed to a pdf, or, if desired, may also be edited to adapt some typographical details. **Figure 2** shows the complete article on *makula* without further editing. For the sake of illustration the full forms of the compound entry and the cross references have been included and highlighted in grey.

makula[1] (*also written* makuḷa), *n. and iic* [*S.* mukula], **1.** (*n.*) *a bud;* Thī 260 (pattali-~-vaṇṇasadisā sobhate su dantā, *read* sattali-? *cf* K.R. Norman, 2007, II, *ad loc.*); Ja III 298,13 (kakkaṭako … kattariyā uppala-~aṃ viya aḷehi ubhinnam pi sīsaṃ kappetvā); Ja IV 333,10 (aṇḍaṃ kaṇikāra-~aṃ viya suvaṇṇavaṇṇaṃ hutvā); Ap 575,17 (~-khārakākārā kalasā va payodharā); Vism 250,6 (dāṭhā … sumana-~-sadisā); Sp 703,6 (kunda-~-rāsisadisaṃ … [odanaṃ] pacitvā) ≠ Dhp-a III 371,17 *foll.* (sumana-~-sadisabhattassa pūrā); Mp II 36,14 (uppala-~aṃ viya Sinerupabbato); Ap-a 215,19 (padumā … ~-pupphādayo gaṇhanti); Bv-a 38,3 (kamala-~-sadisaṃ añjaliṃ sirasi katvā); Vibh-a 239,29 (hadayan ti … saṇṭhānato bāhirapattāni apanetvā adho-mukhaṭṭhapitapaduma-~-saṇṭhānaṃ); Mhv 30:27 (jāti-~-kumbhe so mahābodhaṅgaṇe tayo ṭhapāpetvā); Abh 544 (~aṃ vā kuṭumalo khārako tu ca jālakaṃ); Abh 811 (~e … kosam); — *ifc see* kandalamakula kandala-, sumanamakula sumana-; — *see also* avikasitaketakīmakulasaṇṭhāna *(sv* vikasati[1]*),* ahicchattaka- *(sv Addenda III,* ahi*)* ; — **2.** (*n.*) *a bud as decorative motif;* Ja I 31,24 (ratana-~aṃ sata-sahassagghaṇakaṃ suvaṇṇapātiṃ sappissa pūrāpetvā) ≠ Bv-a 144,2 (ratanamatta-~aṃ …); Dhp-a I 136,10-11 (bhattaṃ nāma ratana-~āya suvaṇṇapātiyaṃ uṭṭhahatī ti) ≠ Dhp-a IV 127,1 (ratanubbedha-~āya mahāsuvaṇṇa-pātiyaṃ bhattaṃ uṭṭhātī ti); — *ifc see* kañcanapātimakula kañcanapāti-, pātimakula pāti-; — **3.** (*iic*) *"[being in the state of] a bud", ie not yet unfolded, developed;* Vism 245,26 *foll.* (makkaṭo … ~-tālapaṇṇasūciṃ daḷhaṃ gahetvā) ≠ Vibh-a 259,30-31 (… ~-paṇṇasūciṃ …, *cf* Pj I 73,7-8: makkaṭo … ekaṃ eva ghanatālapaṇṇaparivethitaṃ tāla-sūciṃ nissāya tiṭṭhati); Mp III 319,4 (attano ~-buddhakāle diṭṭhe pañca mahāsupine vitthāretuṃ, *Ee is wr in taking* Makuḷa- *as a proper name;* ~-buddhakāle *paraphrases* A III 240,17-18: pubb' eva sambodhā anabhisambuddhassa bodhisattass' eva sato); — makulam *in Ee at* Vibh-a 239,31 *is wr for* makulitaṃ, *see* makulita; — **makuladāṭha** °-daṭha, *mfn., having not (yet) developed teeth;* Ja IV 345,21 (ṭhapesi … tesaṃ anantarā ~e taruṇasūkare, tesaṃ anantarā mahā-dāṭhe); Ps II 199,5 (kaṇerukā ca dantānaṃ kaṇerutāya. tā kira ~ā honti).

Fig. 2 *A Dictionary of Pāli* article for *makula*

The reader will notice that the XML encoding discussed above contains neither punctuation marks that separate the headword from the grammatical information, the definition from the quotations, one quotation from another and so on, nor brackets that enclose the etymological information or the quotations. All this is added during the XSL transformation by specific rules that define what kind of punctuation marks or brackets have to be added to which kind of content under which conditions (preceding and following tags, parent or child tags and so on); the same applies to typographical features such as font styles (italics and bold) and so on, as well as to some formalized remarks. In other words, with very few exceptions, the XML file contains only the lexicographical data together with information on its structural function, while the XSL transformation provides information on how this data will be made visible. Thus, content and form are kept separate as far as possible. This makes it possible to use the same data file for differently designed outputs, without the need to change it. Once encoded in the XML file the data is safely stored and ready to be accessed in any desired way. As is well known, XML is especially suited to be transformed into HTML. In fact, this is both easier and more common than a transformation into LaTeX that has been designed to meet the needs of a printable output. Therefore, it will easily be possible to create another transformation for an online edition that will give an output with the same content as the printed edition, but in a design suitable to be read on screen.

Compared to the printable output the XML encoding may appear a bit complicated and confusing. However, using an appropriate XML editor immensely helps customising material to the special structure of an XML file. And, once the scheme for the different types of entries is well established, the lexicographer only needs to know which kind of information has to be filled into which tag, without care for the typographical output. Thus, he can concentrate on his core business which is—lexicography.

BIBLIOGRAPHY

1. Pali Dictionaries

A Critical Pāli Dictionary, Volume I, ed. by Dines Andersen, Helmer Smith and Hans Hendriksen (Copenhagen: Royal Danish Academy, 1925-1948)

A Critical Pāli Dictionary, Volume II, ed. by Ludwig Alsdorf and others (Copenhagen: Royal Danish Academy, 1960-1980)

A Critical Pāli Dictionary, Volume III, fascicles I-VIII, ed. by Oskar von Hinüber and Ole Pind and (Copenhagen: Royal Danish Academy, 1992-2011)

Cone, Margaret, *A Dictionary of Pāli: Part I a-kh* (Oxford: Pali Text Society, 2001).

———, *A Dictionary of Pāli: Part II g-n* (Bristol: Pali Text Society, 2010).

Childers, Robert Caesar, *A Dictionary of the Pali Language* (London: Trübner, 1875).

Rhys Davids, T. W. and William Stede, eds, *The Pali Text Society's Pali-English Dictionary* (London: Pali Text Society, 1921-1925)

2. Other Works

Andersen, Dines, "A Pioneer in Pāli Lexicography", in *A Critical Pāli Dictionary*, Volume I, ed. by Dines Andersen, Helmer Smith and Hans Hendriksen (Copenhagen: Royal Danish Academy, 1925-1948), pp. iv-viii.

Andersen, Dines and Helmer Smith, "Preface", in *A Critical Pāli Dictionary, Volume I*, ed. by Dines Andersen, Helmer Smith and Hans Hendriksen (Copenhagen: Pali Text Society, 1925-1948), pp. ix-xi.

Cone, Margaret, "The I. B. Horner Lecture 1995: Lexicography, Pāli and Pāli Lexicography", *Journal of the Pali Text Society*, 22 (1996), 1-34

———, "Caveat Lector", *Journal of the Pali Text Society*, 29 (2007), 95-106.

von Hinüber, Oskar, "Concluding Remarks", in *A Critical Pāli Dictionary: Volume Iii*, ed. by Oskar von Hinüber and Ole Holten Pind (Bristol: Pali Text Society, 2011), pp. xxxiii-xxxvi.

Norman, K. R., "A Report on Two Pāli Dictionaries", *International Journal of Lexicography*, 8 (1995), 115-125.

———, *The Elders' Verses. II: Therīgāthā.* 2nd edn (Lancaster: Pali Text Society, 2007).

Rhys Davids, T. W. "Report of the Pali Text Society for 1909", *Journal of the Pali Text Society* (1909), pp. vi-x

———, "Foreword", in *The Pali Text Society's Pali-English Dictionary*, Part I (A), ed. by T. W. Rhys Davids and William Stede (London: Pali Text Society, 1921), pp. v-viii.

Rhys Davids, C. A. F, "Report of the Pali Text Society for 1926", *Journal of the Pali Text Society* (1924-1927), pp. 15-25

———, "A Critical Pāli Dictionary", *Journal of the Royal Asiatic Society* 65 (1933), pp. 435-37.

Stede, William, "Afterword", in *The Pali Text Society''s Pali-English Dictionary*, Part VIII (Vim-H), ed. by T. W. Rhys Davids and William Stede (London: Pali Text Society, 1925), pp. 199-203, reprinted as pp. 734-738.

"TEI: P5 Guidelines" <www.tei-c.org/guidelines/p5/>