

**Drifting in Timeless Polysemy:
Problems of chronology in Sanskrit lexicography**

Ligeia Lugli

King's College London

ligeia.lugli@kcl.ac.uk

[Pre-publication version. To cite please refer to the published version in
Dictionaries: Journal of the Dictionary Society of North America. Vol. 39, Issue 1. 2018.](#)

ABSTRACT

Sanskrit lexicography has a long history. Throughout this history the diachronic development of word-senses has never been satisfactorily addressed. This is due, on one hand, to scarce interest in chronological matters on the part of traditional lexicographers, and, on the other, to the paucity of available information regarding the dates of Sanskrit sources.

I argue here that creation of a “natively digital” Sanskrit dictionary would improve the situation. An electronic dictionary that draws on a database of semantically categorized examples associated with chronological metadata would serve a double purpose. First, it would facilitate keeping diachronic sense-ordering within entries abreast of the latest discoveries on the relative chronology of the sources. Second, the dataset of semantically categorized examples of word in context could serve as the basis for further quantitative analysis, whose findings could, in turn, feed in the lexicographic database in an iterative loop. This iterative workflow holds the potential to generate new knowledge as to the chronology of the sources. Rather than passively relying on

the dating proposed by philologists and historians, lexicography could thus play an active role in dating its own sources.

Keywords: Sanskrit, electronic lexicography, historical lexicography, chronology.

Introduction

Sanskrit literature spans millennia and covers a wide range of domains. Its vocabulary is fantastically rich. Morphological plasticity lends the language great lexical productivity, with virtually infinite potential for word-formation through compounding, prefixation and affixation. It also enjoys great semantic pliability, with many lexical items displaying extraordinary levels of polysemy. Briefly put, Sanskrit is a goldmine for lexicographers. Yet, a fundamental problem has frustrated generations of historically-minded lexicographers of Sanskrit. Most sources cannot be dated with any degree of precision or certainty. A typical “date” for a pre-medieval Sanskrit text comprises a couple centuries (e.g., *Yogasūtra* 4th-5th century CE). The relative chronology of the texts is also remarkably fluid. Whenever new conjectures emerge as to the likely period in which a certain work was composed, its relative position with respect to all other works fluctuates.

Such uncertainty about the chronology of sources complicates—if not entirely precludes—the lexicographic representation of diachronic lexico-semantic development. Most available Sanskrit dictionaries adopt only the coarsest-grained periodization and divide word-senses into those attested in the Vedic period (up to c500 BCE) and those of later attestation. This practice, while understandable given the paucity of chronological data, makes for a confusing representation of polysemy, as it gives the impression that all word-senses are equally likely to be realized at any point in time during the last fifteen centuries. With a history spanning over

three millennia, the scale of this historico-semantic inaccuracy can easily reach enormous proportions.

This article proffers that the creation of a “natively digital” dictionary of Sanskrit can improve the situation. I use the expression “natively digital” to contrast the model I am proposing with the digitized versions of nineteenth century Sanskrit dictionaries currently available online.¹ While innovative in the field of Sanskrit lexicography, the model I propose falls within the standards of rather conservative electronic lexicography.² It consists of a dictionary based on digital corpora, connected to a digital database containing metadata-enriched and semantically categorized examples of words in context, and published in electronic form.

The paper opens with a brief overview of the reasons why the chronology of Sanskrit sources is problematic. It then outlines how the problem of chronology is compounded by the approach to polysemy typical of traditional Sanskrit lexica. Next it moves onto a review of how polysemy and chronology are treated in modern and contemporary Sanskrit lexicography. Finally, it sketches out a proposal for improving on the current representation of polysemy in Sanskrit dictionaries. It closes with the suggestion that switching to a “natively digital” model of Sanskrit lexicography may not only improve the quality of the dictionary, but also shed new light on the chronology of the sources themselves.

¹ A good quality repository of such digitized lexicographic resources is the Cologne Digital Sanskrit Dictionaries site (<http://www.sanskrit-lexicon.uni-koeln.de/>).

² It mostly fits the decade-old guidelines for dictionary making articulated in Atkins and Rundell (2008).

Problems of chronology with Sanskrit sources

The chronology of Sanskrit sources is extremely uncertain. This is especially true for ancient times (before the 8th century CE). Authors (or text-compilers) from this period are often unknown, or only faintly divined through the layers of myth and hagiographic legend that shroud their identities.

Beside the lack of historical information about the composers of Sanskrit texts, a varied set of adventitious circumstances obfuscates the chronology of the sources. A full review of these circumstances, and of the philological strategies adopted to counter them, lies beyond the scope of this paper. Outlining the problem in very general terms here should suffice.³

To begin with, the texts themselves contain no explicit reference to the date in which they were composed. The extant manuscripts, which can be carbon dated, are typically late and offer no clue as to the date of composition of the texts they contain. Ancient texts were orally transmitted for centuries and the earliest manuscripts on which they were, eventually, written down have been lost to the humid climate of South Asia (see Katre (1954, Ch. 2). Moreover, the history of the Indian subcontinent before medieval times is not known in much detail. So, references to kings and socio-political events are often not as illuminating as they would be in contexts where detailed chronicles and annals are available.

Internal criteria for dating are also problematic. Morphology and spelling are not reliable indicators of date of composition, for two reasons. First, many texts are likely to have adopted

³ For a brief introduction to the problem and its relation to issues of textual transmission in South Asia, see Cola (1999); for a short but insightful problematization of the very concept of textual chronology in the ancient Sanskrit context, see Fosse (1997, 142).

archaistic morphological and phonetic features to enhance their authority and prestige.⁴ Second, these features are likely to have undergone changes in the long history of textual transmission.⁵ The lack of textual stability during transmission has undermined many attempts to extract chronological information through statistical methods, and has left indologists to rely mostly on qualitative analysis for dating sources (see Jamison [1999] and Fosse [1997]). On the basis of intertextual connections and reference to conceptual or technical developments whose diachronic progression can be conjectured or inferred, a tentative relative chronology of texts has emerged. This chronology is constantly contested and revised.⁶ It occasionally benefits from the little certainty that can be harvested from dated translations of Sanskrit texts in other languages. Many Buddhist sources, for example, were translated into Chinese and Tibetan in the course of the first millennium CE. Those translations are often clearly dated and serve as *terminus ante quem*, the date by which a text must have been in circulation. However, this tells little about the actual date of composition of the text being translated. It could have been composed days or centuries before its translation. Add to the aforementioned complications that arise when the date of the

⁴ See Fosse (1997, 277); Houben (1976, 178); and Silk (2009). Cf. Norman (1997, 74) for an occasion where morphology does neatly succeed in discriminating relative chronology.

⁵ For an overview of typical changes in textual transmission, see Katre (1954, Ch. 5).

⁶ For a glimpse into the precarious fluidity of relative chronology of Sanskrit texts, see Deleanu (2006, 186-195).

translation is itself doubtful, or when the dated translation does not correspond to the same recension of the Sanskrit text that is available to us today.⁷

All in all, it comes as no surprise that despite over a century of scholarly effort, the chronology of Sanskrit literature remains largely tentative.

Chronology and polysemy in traditional Sanskrit lexicography

Uncertainty about the chronology of Sanskrit sources has obvious repercussions on historical lexicography. It obfuscates the diachronic lexico-semantic evolution of Sanskrit. Traditional South Asian works on the Sanskrit language are of little assistance in this regard. Ancient Indian grammarians and lexicographers were extremely careful in recording many aspects of their language. Yet, they paid no attention to matters of chronology. They were well aware of lexical, semantic, phonetic, and morphological variation in Sanskrit but did not record diachrony as a determining factor in such variation. Scholars have labeled this approach to Sanskrit as embedding a “panchronistic” view of the language, whereby diachronic and synchronic dimensions of the study of the language collapse into one another, or as betraying a “timeless” conception of language, where matters of chronology simply do not play any role in the description of the language (see Kahrs (1998) and Deshpande (1996, 401); cf. Akhujkar (1996, 74n23). This feature has sometimes been connected to the traditional Hindu characterization of Sanskrit as an “eternal language” (see Cardona 1990 and Deshpande 1985)

⁷ On the unreliability of some date Chinese translations from Sanskrit, see, for example, Nattier (2008, 8–9).

A perhaps more useful way of approaching lack of diachronic interest in traditional literature on Sanskrit is to consider the ‘learned’ status of this language.⁸ In ancient India Sanskrit was largely learned through memorization of texts that were transmitted from earlier times but were still very much part of a living tradition. In this way, different diachronic layers of the language co-existed, to some extent, synchronically due to the Sanskrit speakers’ familiarity with traditional quotations from earlier sources. The most eminent of such sources was the Veda, whose antiquity was perceived more in terms of authoritativeness and prestige than in terms of chronological precedence.⁹

The ancient authors of Sanskrit grammatical treatises and glossaries, ever attentive to matters of linguistic prestige, did record the difference between Vedic and learned words on the one hand, and words used in daily transactions on the other (cf. Houben 1996, 165–71). They did not, in contrast, provide any details as to which items of the vocabulary sounded obsolete or which had only recently entered the vocabulary. Their focus lay in establishing semantic relations among words. The more the relations the better; this seems to have been the implicit standard for good traditional lexicology. Alas, this enthusiasm for multiplying semantic relations, as we shall soon see, affected Sanskrit lexicology well beyond premodern South Asia, and had detrimental repercussions on Western Sanskrit lexicography as well—especially as far as an historically informed representation of polysemy is concerned.

⁸ On the status of Sanskrit in pre-medieval India from a socio-linguistic point of view, see Houben (1996).

⁹ On the importance of considering alternative to chronology when studying the conceptualisations of historical matters in South Asia, see Thapar (2013, Ch.1).

Traditional Sanskrit lexicography used three primary methods to draw semantic relations between words.¹⁰ The first is the so called “etymological (*nirukti/nirvacana*) method, whereby a word is said to derive its meaning from another on the basis of either sharing its stem, or simply displaying some assonance with it.¹¹ This method serves to explain the meaning of words by relating them to other words and is used primarily for text interpretation. Vedic glossaries (*nighaṇṭu*), the earliest of which is often considered as marking the beginning of India’s lexicographic tradition, are the foremost representations of this method. The etymological method is also applied outside of the Vedic tradition and provides the ground for much of the hermeneutics in the commentarial literature of virtually all South-Asian religio-philosophical schools, where the words of foundational religious texts are subjected to interpretive twists to

¹⁰ For a brief overview of the rich Sanskrit lexicographic tradition, see Boisson et al. (1991, 180–181); for a longer study, see Vogel (2015).

¹¹ An example of this technique as applied in a Buddhist treatise: “*nirukti* is an etymological explanation (*nirvacanaṃ*), for example: *rūpyate tasmād rūpam* (it is physical matter because it can be crushed), *vijānātīti vijñānam*, (it is consciousness because it knows or distinguishes), *cinoṭīti cittam* (it is mind because it accumulates” (*Abhidharmakośabhāṣya*, Pruden (tr.), 1154). See also Houben (1996, 419): “Yāska says that commonality of even one vowel or consonant may be sufficient to make a connection, and that an etymologist can never say no to a word. He must attempt an etymology (cf. *avidyamāne sāmānye 'py akṣaravarṇasāmānyān nirbrūyāt, na tv eva na nirbrūyāt, Nirukta 2.1*).” For an introduction to the history of and techniques used in traditional *nirvacana*, see Kahrs (1998, Ch. 1–2).

accommodate ever new ideas.¹² The practice complicates the study of Sanskrit historical lexicology: it results in an artificial multiplication of meanings and overlays old word-uses with new senses, thus creating semantic anachronisms that are not always obvious.¹³

The second method consists in supplying lists of “synonyms.” This method constitutes the foundation of the most widely represented genre in Sanskrit lexicography, the thesaurus (*kośa*). While *nighaṇṭus* were primarily used for text interpretation, *kośas* aided text production. They catered especially to poets (Vogel 1979, 304n5; Bronner 2010, 128–32). To keep up with the increasingly more flamboyant aesthetic of Sanskrit *belles lettres* and the resulting needs of poets, lexicographers—who were often poets themselves—strove to find “synonyms” that would fit different meters, that were amenable to puns, and resonated with metaphors (Bronner 2010, 128–32). This led to a proliferation of strings of synonyms that had but the most tenuous semantic relations with one another.¹⁴ Such artificial multiplication of synonyms is accompanied by an inflation of polysemy.

¹² For use of this method within the Śaiva tradition, see Khars (1998); for uses in Jainism, see Dundas (1996); for applications in Buddhism, see Apple (2009).

¹³ The flipside of this exegetical practice is that sometimes the inaccuracy is blatant and reveals something about the lexical competence or level of comprehension of older texts in later periods. For one such case, as well as in general for the use of *nirukti* in South Indian exegesis, see Norman (1984, 80-81).

¹⁴ Bronner (2010, 14-16) points to the words for ‘king’ and ‘moon,’ which were probably never considered interchangeable in daily communication but were listed as synonyms in thesauri because of a traditional simile by which royalty is compared to the moon.

The identification of polysemy is the third and last method traditional Sanskrit lexicology uses to draw lexico-semantic relations. It is the scope of the so called *anekārthakośa*, dictionaries of homonyms and polysemous words. These lexica resemble lists of word senses, with each entry containing a lemma and a short description of one of its meanings. They tend to be heterogenous, due to the conflation of homonymy and polysemy.¹⁵ They also grow longer over time.¹⁶ Their length is partially a genuine reflection of scale of polysemy in Sanskrit. After all, the multifarious specialized uses and metaphorical meanings that a vocabulary tends to acquire when is used over a body of texts as vast as Sanskrit literature, make for wide semantic spectra. Partly, however, the representation of polysemy in these works is just a lexicographic artefact.¹⁷

Traditional Sanskrit lexica often combine homonymical and synonymical arrangements. Typically, they divide into sections listing words that differ in form but share the same meaning (synonyms) and words that share the same form but differ in meaning (homonyms or polysemic lemmata). As a result of this parallel arrangement, the number of words a lemma is given as

¹⁵ On the distinction between homonymy and polysemy in the classical Sanskrit tradition, see Aussant (2014).

¹⁶ See Squarcini (2015, xlii-xliii) for an interesting example of the acquisition of new meaning on the part of the lemma *yoga* in successive stages of traditional Sanskrit lexicography.

¹⁷ Dictionary makers the world over have often indulged in multiplying the polysemy of their headwords. Contrary to the tendency in Western lexicography, Sanskrit lexica are not prone to over-subdividing senses to unnecessary levels of granularity (e.g., differentiating ‘woman’s milk’ from ‘cow’s milk,’ like lexicographers of Middle English; see Lewis 2002, 150), they rather tend to multiply metaphorical extensions.

synonym with often guides the distinction of the different senses of that lemma. In other words, the number of synonyms of a word is seen as a function of the number of different senses it expresses (see Ghatage 1973, 31). The process over-represents polysemy and sheds no light on its historical development.

Chronology in modern Sanskrit Lexicography in the West

Sanskrit Lexicography in the West has improved remarkably little on the situation outlined above. The first widely circulating Sanskrit-English dictionary, Wilson 1832, was heavily based on traditional *kośas* and reproduced the timeless polysemy typical of this literature. Wilson provided no examples or references to attestations of the word senses in context, thus letting word-senses that may have never actually been used in real text slip into his work. Subsequent dictionaries tried to rectify this by including examples and references. Still, they did little to uncover diachronic semantic development. Some thirty years after Wilson, Burnouf and Leupol address the problem directly in their dictionary preface and conclude that a chronological ordering of the senses is simply not achievable for Sanskrit.¹⁸ Around the same period, Boehtlingk and Roth (1868) published an extremely well curated Sanskrit-German dictionary and did not even mention the possibility of attempting a full a chronology of attested senses. This is not to say that Boehtlingk and Roth disregarded questions of diachronic development

¹⁸ Burnouf- and Leupol (1866, vi): “dans chaque article on peut présenter les significations selon leur ordre de succession chronologique, et faire de la sorte un dictionnaire historique de la langue. Dans l'état présent des études orientales, nous croyons qu'un travail de ce genre est à peu près impossible pour le sanscrit.”

altogether. Simply, they focused their attention on a single facet of the problem: whether a word-sense was attested in the Vedas or not.

The early lexicographers of Sanskrit all focused on Vedic literature because the Vedas correspond to the oldest stratum of the language, and it is therefore historically important to record whether a word-sense is attested there. Thus, the traditional distinction between Vedic and “daily” uses of Sanskrit words, beloved to ancient grammarians, is preserved in Western lexicography; but recast in historical terms.¹⁹ In probably the most widely used Sanskrit–English dictionary to date, Monier-Williams (1899) reiterated and elaborated the same ideas. In the preface, he remarks, “Sanskrit literature comprises two distinct periods: Vedic and post-Vedic. [...] Vedic literature begins with the R̥gVeda (probably dating from about 1200-1300 BC) [...] post-Vedic literature begins with the code of Manu (probably dating in its earlier form from 500BC)” (Monier-Williams 1899, xxi).

Needless to say, over the seven centuries of Vedic literature and the one and half millennia of post-Vedic literature the vocabulary is likely to have changed significantly. Even if we accept that words from the early strata of Sanskrit remained in use to some extent in later periods, we cannot assume that the lexical repertoire remained fixed through the centuries. New expressions were lexicalized in the course of time, pre-existent words acquired new meaning, and the old word uses declined.²⁰ It seems also naive to suppose that the entire vocabulary followed a single

²⁰ The word *dharma* constitutes one of the few well studied cases of semantic change in Sanskrit. The reader may turn to Olivelle (2004) for a study of changes in Vedic literature, and Cox (2004) for an analysis of changes in Buddhist literature (which would fall in Monier-Williams’ “post-

trajectory of lexico-semantic development (or lack thereof) across the all textual traditions. Some word-meanings may have become obsolete in some segments of Sanskrit literature and remained current in others. Similarly, terminological specialization likely evolved at different paces and in different directions depending on the tradition. Obvious as these issues may appear, they received little attention in Western Sanskrit dictionaries.

After Wilson, the debate on Sanskrit historical lexicography centered on the importance of including carefully referenced examples of actual word-uses from the sources. Useful in many respects, such examples provide poor guidance to the historical development of a lemma, because they indicate titles of texts without reference to their dating. It is up to the user to know where to place the text in the history of Sanskrit literature. Apart from the inherent uncertainty about the chronology of Sanskrit sources, the sheer vastness of Sanskrit literature makes it unlikely that a user will be familiar with scholarly conjectures about the dating of texts that fall outside of his or her area of expertise (e.g., an expert in Gupta Buddhist literature may well not be aware on debates on the chronology of medieval medical literature). Most dictionary users will have to consult further resources to learn the proposed dating of texts referred to in the dictionary. In brief, to reconstruct diachronic information from lexicographic references is impractical, especially so given that such references only provide the blandest information about semantic history of a word.

Examples included in Sanskrit dictionaries point to the most illustrative uses of a word-sense in context. They do not aim to offer a comprehensive range of sources where that word-sense is

Vedic” period). For other instances of semantic change in Sanskrit, with varying quality of analysis and sources, see Kamboj (1986).

attested. Dictionary users can learn from an entry that a lemma is used in a certain sense in a certain text, but they cannot draw any inference as to the position of this text in the semantic history of that lemma. Is the text quoted in the dictionary entry among the first ones to use that certain word in that certain sense? Is an example representative of a certain type of literature or of all Sanskrit literature?²¹

Questions like these cannot be answered by consulting a Western Sanskrit Dictionary. If this is disappointing to the historical linguist, it can prove utterly misleading for the inexperienced student, who, not having yet developed an intuition for the vocabulary used in a certain period or tradition, easily assumes (despite the repeated advice of the instructor) that all the meanings listed for a lemma are equally likely to be expressed in the text she or he is trying to read. Faced with a lemma like *kalpa*, for which Monier-Williams (1899, s.v. *kalpa*) lists over twenty senses ranging from 'law,' to 'alternative,' 'research,' 'medical treatment,' 'cosmological division of

²¹ Historical lexicographers of most Western languages since Franz Passow (see Podhajecka, this issue, pages xx) do attempt to identify first-known uses, and because the quotations proposed as first very often aren't, some lexical researchers focus on antedating (see Shapiro, this issue, as well as Taylor and Christensen, this issue). Earliest quotations are a principle concern of the *Oxford English Dictionary* (see Sheidlower 2011, 200–203), but textual diversity is not (Sheidlower 2011, 205–206). Other historical dictionaries, such as the *Middle English Dictionary*, represent the variety of text types that participate in a word's history; if all quotations come from one type, the implication is that the word in question is restricted to that type (see Adams 2013, 10).

time,' and 'tree of paradise,' students can be excused if they feel overwhelmed and more than a little perplexed.

Chronology in the 'Poona' dictionary

Fortunately, there is an exception to this confusing lexicographic representation of Sanskrit, the Sanskrit-English dictionary that is being produced in Poona, India. Published under the title of *An Encyclopaedic Dictionary of Sanskrit on Historical Principles* and commonly referred to in the field as "the Poona dictionary," it is by far the most ambitious Sanskrit-English dictionary project ever attempted. Alas, as is often the case with ambitious lexicographic enterprises, it is also one of the most slowly paced. In progress since the early 1970s, it has produced 31 volumes so far, covering only a portion of **a**, the first letter of the Sanskrit alphabet. (It should be noted that **a** is a common initial for Sanskrit words, so one hopes that the remaining letters will not take as long to complete).

The main strength of the Poona dictionary lies in the theoretical and methodological awareness of its editors, especially in matters related to the historical development of polysemy and other semantic relations.²² While they join in the general consensus that a full semantic history of the Sanskrit vocabulary is precluded by the paucity of information about the precise dates of the sources (Kelkar 1973, 60), the editors of this dictionary have made laudable efforts to provide at least some chronological information about word senses.

²² So, Gathage (1976, iii–iv) writes, "Far more important for the task of a historical dictionary are the clarification and implications of the concepts of homonymy, polysemy and hyponymy among words and process to handle them in a historical perspective."

The most tangible product of their efforts is a tabulated synopsis of the chronology of the sources referred to in the dictionary, divided by textual tradition (Gathage 1976, lxxiii–lxxxii). This chronological synopsis offers a double advantage. First, it provides a reference to approximative dates of the sources cited, and thus saves time to the user. Second, the division by textual tradition allows for parallel chronologies and thus for representing the different pace at which the vocabulary of different traditions evolved. The synopsis suffers from one major shortcoming. It cannot easily be updated. Some of the dates it proposes are now outdated, as with new findings and conjectures the consensus on the dates of several texts has changed.²³ The editors were surely aware of the problem, but the constraints of hard-copy publishing limited their ability to keep the chronological synopsis up to date (see Narahari 1973, 89).

Besides the chronological synopsis, the Poona dictionary also aims to embed diachronic information within its lexical entries. It sets out to order word-senses chronologically by the date of earliest found attestation in each textual tradition (Ghatage 1976, xi–xii). This constitutes an

²³ For instance, the dictionary synopsis dates an important proto-Yogācāra text, the *Bodhisattvabhūmi*, to 300–350 CE. These dates are now believed to correspond to the mature development of the Yogācāra school, and the *Bodhisattvabhūmi*'s likely dates have been moved to a century earlier (Deleanu 2006, 194 proposes 230–300 CE). To establish the actual date of the text is not possible, but it is important to preserve the likely relative position of texts within the development of traditions.

enormous improvement on earlier dictionaries.²⁴ Still its practical realisation is far from ideal. Begun before the boom of digital corpora and corpus-based dictionaries, the Poona dictionary is based on a selection of citations manually extracted from a limited corpus of Sanskrit literature. The earliest attestation found in these citations bears no necessary relation with the actual earliest use of the word, or word-sense. This is of course the case with all historical corpora. However, it seems particularly problematic in this case, as the extant corpus of Sanskrit literature is enormous and the dictionary compilers have only looked at an infinitesimal fraction of it.²⁵ Moreover, it is debatable to what extent the fraction they did consult is representative of the Sanskrit sources that are available to us at present. Ghatage (1976, x) appears to have held the belief that Vedic literature must be more represented than other types of literature: “All that can be done [...] is to effect a complete extraction of some selected books and such books would naturally be more for the Vedic period than for the later phases of the language.” It is unclear whether this belief derives from an idea of Vedic primacy, akin to that expressed by Boehtlingk and Roth (1868) and traditional grammarians, or by a quantitative appraisal of the Sanskrit material available at the time. Regardless, the corpus design that informed the Poona dictionary is now likely to be outdated, as the belief in the primacy of Vedic sources for linguistic study has

²⁴ It is also an improvement on other contemporary lexicographic projects. The *Sanskrit-Woerterbuch der buddhistischen Turfan-Funde*—which like the Poona dictionary, started in the early 1970s and is still ongoing—does not attempt to order word-senses chronologically at all.

²⁵ On the problems of source selection and the limitations of manually searching for word instantiations, see Kelkar (1973, 79) and Ghatage (1976, x).

waned, and more Sanskrit manuscripts from later periods have emerged, especially Buddhist materials.

The methodology used to retrieve frequency information from the corpus is also outdated. In the 1970s, the dictionary's editor despaired that an accurate statistical representation of the frequency and dispersion of lemmata and their senses in extant Sanskrit literature could ever be achieved. He noted (Ghatage 1976, x) that to compile "complete indices verborum of 1500 books is an impossible task" and concluded that, while the dictionary entries contain remarks on the level of popularity of word-senses in different periods, these remarks are to be taken as merely "impressionistic" (Ghatage 1976, xxiii).

Almost half a century later, we are now in a better position to tackle word frequency and other lexicographic issues through the use of digital resources.

Chronology and Sanskrit lexicography in the digital age

The advent of the digital era has revolutionized lexicography. Digital corpora have improved the efficiency of lexicographic workflow, and electronic dictionaries have freed lexicographers from the constraints of print. Yet, so far, these technological advances have had but a marginal impact on Sanskrit lexicography, where it has been largely confined to the digitization of pre-existent dictionaries (see note 1). It is time for Sanskrit lexicography to fully embrace technology. The creation of a "natively digital" Sanskrit dictionary could improve, not only the representation of Sanskrit vocabulary, but also our understanding of its historical evolution.

Before embarking in a discussion of the benefits that an entirely digital workflow would bring to Sanskrit lexicography, a few words on the reasons why this has not yet happened are in

order.²⁶ Besides the difficulties of attracting funding for historical lexicography—which, much to the distress of this journal’s readers, are shared equally across all languages—Sanskrit poses an additional difficulty: it is not easy to tokenize.²⁷ Suffice here to note that, since the division of texts into discrete lexical units constitutes the very basis of corpus analysis, it comes as no surprise that Sanskrit lexicography is lagging behind in terms of digital corpus workflows. Still, this situation needs not continue. Unprocessed digitized corpora of Sanskrit are available (e.g., GRETIL (<http://gretil.sub.uni-goettingen.de/>); progress in Sanskrit Natural Language Processing is being made; and, even if accurate tokenization and other corpus processing tasks (e.g., part of speech tagging and lemmatization) may still loom in a distant future for Sanskrit,²⁸ basic word retrieval and word frequency statistics can be achieved through regex-powered searches and character-ngram tokenization.²⁹

²⁶ I am referring here to the development of a dictionary of the entire Sanskrit language.

Attempts to adopt a digital workflow for small scale specialized dictionaries are under way, e.g., The Buddhist Translators Workbench.

²⁷ An overview of the linguistic features and technical details that complicate Sanskrit tokenization exceeds the scope of this article, but see Hellwig (2015).

²⁸ There have been some over-optimistic depictions of the generalizability of existent segmenters, morphological analyzers and lemmatizers for Sanskrit (e.g., Goyal et. al 2012). None of the tools I have tried has performed well on my Sanskrit corpus of Buddhist texts.

²⁹ For an overview of different tokenization techniques for a language that share some of the same problems as Sanskrit, see Rehman et al. (2013, Section 2).

These advances in text processing have significantly improved on the situation Ghatage lamented in 1976, but they have not completely fixed it. Today is easy to obtain accurate representation of the frequency and dispersion of lemmata across all sources available in searchable digital format. All is needed to obtain information about the distribution of a lemma over time is a diachronic corpus enriched with chronological metadata.³⁰ Statistics about lemma frequency and distribution, however, are of limited usefulness in a dictionary. Information about the relative frequencies and distribution of each sense would be far more helpful to dictionary users. Translators and students would benefit from indications as to the likelihood that a word-sense is expressed in a text of certain period and tradition. Historical linguists would be able to take advantage of dictionary entries to study semantic developments. Brief, a chronological presentation of polysemy is a great desideratum in Sanskrit lexicography.

Unfortunately, automatic detection of word-sense attestations is not yet possible for Sanskrit.

³¹ Some indications about the frequency of word-senses can be gleaned from co-textual patterns (regularities and discontinuities in the lexical surroundings of a lemma) and morphological features (i.e., number, case, compounding behaviour, or gender). Accurate word-sense identification, however, can only be achieved through manual semantic annotation. This practice is obviously labor-intensive and time consuming. Hence it can only be applied to small portions of the corpus, or to selected citations.

³⁰ There is an urgent need in the field for metadata-enriched diachronic Sanskrit corpora.

Currently, the only available example is the Digital Sanskrit Corpus, a small 4 million word manually tagged corpus available at <http://kjc-sv013.kjc.uni-heidelberg.de/dcs/index.php>.

³¹ Hellwig (2017) reports some progress in this regard.

All of this recalls the same problems of representativeness highlighted above, regarding the Poona dictionary. Digital corpora and statistical techniques, however, can help mitigate the risks that derive from using small samples (for an overview, see Glynn 2014). On the assumption that semantics correlates with grammatical and co-textual features, lexicographic sampling can be designed to target a certain degree of morpho-syntactic and contextual variation to capture a balanced representation of the words behaviour.³² Of course, in case no correspondence between semantics and co-textual or morphological patterns emerge, generalization about the frequency of word-senses over the entire corpus should remain “impressionistic,” as in the Poona dictionary. Even in these cases, however, a “natively digital” dictionary would greatly improve on the current state of Sanskrit lexicography.

From stasis to dynamism: iterative lexicography to advance chronological knowledge

The Poona dictionary relegates explicit chronological information to a table printed in the first volume of the dictionary. The word-senses, examples, and references in the dictionary entries are ordered chronologically according to that table. This system is not as user-friendly as it could be and it is prone to nearly immediate obsolescence. Both issues can be solved instantly by switching from a static printed table to the dynamic display of an electronic dictionary.

Hyperlinks facilitate users’ access to the chronological metadata associated with the dictionary

³² For a concise overview on meaning as correlated to morphosyntactic features, see e.g., Janda (2013); for word-sense disambiguation through co-textual patterns, especially collocation, the *locus classicus* is Kilgariff (1997); on sampling a corpus for specific word-senses using contextual cues, see Zenner et al. (2013, 266).

examples and references. Free from the space constraints of print, these metadata may be extended to include a comprehensive summary of the available chronology information on the text. A typology of chronological information, distinguishing, for example, between widely accepted dates, disputed dates, “hard” *terminus ante quem* dates, and date of the earliest manuscript witness would provide a valuable compass in navigating the ever-fluctuating relative chronology of Sanskrit sources.³³

The main advantage of digital metadata vis à vis printed information, however, lies in their easily updatable nature. Sanskrit lexicography has a small market and new editions are likely to prove too costly to be released frequently, if at all. By contrast, revising chronological metadata in a well-designed lexicographic database is financially sustainable and efficient.³⁴ If the consensus on the likely date of a text changes, its metadata can be updated and all the examples and word-senses associated to that text across the entire dictionary could be instantly re-arranged to conform to the new periodization. Thus, the historical representation of words’ polysemy could cease to be a static lexicographic artefact, forever tied to the limitations of the knowledge

³³ On the usefulness of providing multiple forms of dating in cases of chronologically problematic textual transmission, see Mishor (2002, 133).

³⁴ “Well-designed” means fit for its specific lexicographic purpose. Given the lack of stability of in the accepted chronology of Sanskrit texts, for a Sanskrit dictionary ease of updating chronological metadata should be a priority in database design; cf. Bergenholtz, and Nielsen (2013).

available at the time of the dictionary's compilation.³⁵ It could evolve dynamically with the progress of the field. More strikingly, it could help advance the very progress of field by shedding new light on the relationship between Sanskrit sources.

So far, I have considered the diachronic representation of polysemy in dictionaries as a function of pre-existent knowledge about the chronology of the sources. This makes perfect sense. When we know the dates of the sources, we can diligently order the senses to mirror their successive chronological appearance in increasingly more recent sources. With Sanskrit sources we rarely enjoy the luxury of such historical knowledge. One could, at first, believe that relying on chronological approximations and conjectures is pretty much the best we can do. This might have been the harsh reality that Burnouf and Ghatage had to face. Today we are allowed a less pessimistic outlook. We could take advantage of the very process of dictionary making to advance our historical knowledge, and refine, confirm, or refute those approximations and conjectures about the chronology of the sources.

Lexicographic work is typically associated with the division of the semantic spectrum of words into discrete units of meaning and with the drafting of dictionary entries. While dictionary entries are indeed the most ostensible product of lexicographers' efforts, they are not the only ones. An enormous amount of work lies behind the shortest dictionary entry.³⁶ Much of this

³⁵ For classic articles on the advantages of dynamically adjusting entry display in electronic dictionaries see Lew (2009, 256 ff.) and Schryver (2003 189-190).

³⁶ Sometimes, this work is elaborated and published in the interest of lexical research independent of or complementary to the dictionary product, obviously, as in the case of the

work consists in the semantic categorization of attestations of lemmata in context. All too frequently, the steps that lead to this categorization find no use apart from feeding into an entry. Still, the semantic analysis they involve is valuable in and of itself, especially so if it follows a systematic workflow and its output is stored in a lexicographic database in a clearly structured machine-readable format (e.g., TEI-compliant xml).

Such a lexicographic database of semantically categorized examples can function as a high-quality dataset for quantitative linguistic analysis. If the examples are derived from a carefully designed corpus, they could provide a wealth of information as to the distribution of word uses across Sanskrit literature. Importantly, they would provide the kind of information that is difficult to extract from the corpus automatically, as it involves sense disambiguation, which is currently beyond the reach of Sanskrit NLP (see note 31). Besides its immediate use in a dictionary, these data can serve as the basis for research into the historical relationship between the sources. Using similarities measures and other statistical techniques, texts could be clustered according to the lexico-semantic patterns they display (see, e.g., Moisl 2015). While much of the similarities in vocabulary usage will probably turn out to be traceable to traditional affiliation—Buddhist texts will cluster together, and, within Buddhism, texts belonging to the same school will in turn cluster together—a well-balanced corpus may allow us to detect patterns across different textual traditions. These patterns may point to geographical and chronological relations

Dictionary of Old English (Healy 2002), or obliquely, as in the case of the never-completed *Early Modern English Dictionary* (Adams 2010).

between the communities that produced those texts, thus affording us precious historical insight.³⁷

Needless to say, quantitative analysis of lexico-semantic features is no panacea. A century of unsatisfactory attempts at gaining chronological knowledge through statistical analysis cannot but have a sobering effect on even the most enthusiastic quantitative linguists. Yet, there are grounds for optimism. So far, efforts to historicize the sources through statistical techniques have mostly relied on morphological features. As outlined at the beginning of this article, morphological forms are subject to much variation in textual transmission. My intuition, as a Sanskrit philologist, is that lexico-semantic features remain more stable during text transmission than morphology and spelling. Hence, I am inclined to believe that the vocabulary will provide a more reliable ground for statistical analysis. This intuition needs to be confirmed or disconfirmed through a study of patterns of variation across different recensions of texts. An initial survey of 200 cases of variants across four recensions (for a total 38 witnesses) of one Buddhist Sanskrit text suggests that lemmata are twice as stable as either morphological forms or spelling.³⁸ A

³⁷ Research along these lines is under way as part of the project *Lexis and Tradition: Variation in the vocabulary of Sanskrit Mahāyāna literature* at King's College London. The corpus developed and the data collected for this research will be released in 2019.

³⁸ My preliminary analysis focuses on 44 verses from four different sections of taken from the 17th chapter of the *Samādhirājasūtra* (Skilton 1997). The critical edition collates variant readings from 38 witnesses grouped in four main recensions of the text (excluding Gilgit). The edition lists variant readings for virtually each and every word in the text. I have categorized each variant as due to spelling (including both likely scribal mistakes and likely genuine

more formal and larger scale analysis over different texts and textual traditions is needed to establish to what extent the vocabulary can justifiably be thought to have remained stable throughout textual transmission.

For the time being, I can only suggest that structured semantic categorisation of words in context should not remain an inert by-product of lexicographers' work. It should serve as the basis for further lexicological study aimed at the historicization of the sources. Chronological insights gleaned from this study could then feed back into the lexicographic database in the form of updated chronological metadata. Subsequent changes in chronological metadata, due to new findings or theories in the field, would lead to the progressive a modification the first lexico-semantic dataset, which, once sufficiently different from its original state, could inform a new cycle of lexicological research. Benefitting from constantly updated information, such iterative model of lexicography seems better suited than its linear counterpart to tackle the fluidity of Sanskrit chronology and rescue its vocabulary from drifting in timeless polysemy.

Conclusions

Sanskrit lexicography has a long history. Throughout this history the diachronic development of word-senses has never been satisfactorily addressed. This is due, on one hand, to scarce interest in chronological matters on the part of traditional lexicographers, and, on the other, to the paucity

phonological differences), morphology (e.g., differing number, gender or case), or attestation of an altogether different lemma (e.g., *sarva* for *dharmā*). I have identified 46 cases where one or more of the recensions displays different lemmata at the same location, 99 cases where they display different morphological forms, and 90 cases where the variant reading is due to spelling.

of historical data available in modern times. The vocabulary's semantic development over the ages is further obfuscated by the vast extent of words' polysemy in Sanskrit. This polysemy partly derives from the sheer size and diversity of Sanskrit literature and partly stems from an artificial inflation of word meanings on the part of traditional lexica. As a result, modern and contemporary Sanskrit dictionaries all too often offer a confusing depiction of word senses, where no reliable information is given to help the user understand which senses are more likely to be expressed in a certain period or textual tradition.

I argue here that creation of a "natively digital" Sanskrit dictionary would improve the situation. An electronic dictionary that draws on a database of semantically categorized examples associated with chronological metadata would serve a double purpose. First, it would facilitate keeping diachronic sense-ordering within entries abreast of the latest discoveries on the relative chronology of the sources. Second, the dataset of semantically categorized examples of word in context could serve as the basis for further quantitative analysis, whose findings could, in turn, feed in the lexicographic database in an iterative loop. This iterative workflow holds the potential to generate new knowledge as to the chronology of the sources. Rather than passively relying on the dating proposed by philologists and historians, lexicography could thus play an active role in dating its own sources.

Acknowledgments

I would like to thank the British Academy for sponsoring aspects of this research through a Newton International Fellowship and the Mangalam Research Center for Buddhist Languages for supporting my work on Buddhist Sanskrit lexicography.

REFERENCES

Dictionaries

Bechert, Heinz, Klaus Röhborn, K., and Jens-Uwe Hartmann, 1973– . *Sanskrit-Wörterbuch der buddhistischen Turfan-Funde*. Göttingen: Vandenhoeck und Reuprecht.

Böhtlink, Otto, and Rudolph Roth. 1855–1875. *Sanskrit Wörterbuch*. 7 vols. St. Petersburg: Eggers.

Burnouf, Émile, and L. Leupol. 1866. *Dictionnaire classique sanscrit-français: où sont coordonnés, révisés et complétés les travaux de Wilson, Bopp, Westergaard, Johnson, etc. et contenant le dévanâgari, sa transcription européenne, l'interprétation, les racines et de nombreux rapprochements philologiques*. Paris, Maissonneuve.

Ghatage, A. M. 1976. *An Encyclopaedic Dictionary of Sanskrit on Historical Principles*. Poona: Deccan College.

Monier-Williams, Monier. 1899. *A Sanskrit-English Dictionary: Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages*. Oxford: Clarendon Press

Digital Resources

Cologne Digital Sanskrit Dictionaries (<http://www.sanskrit-lexicon.uni-koeln.de/>).

GRETIL: Göttingen Register of Electronic Texts in Indian Languages (<http://gretil.sub.uni-goettingen.de/>).

Digital Corpus of Sanskrit (<http://kjc-sv013.kjc.uni-heidelberg.de/dcs/index.php>)

The Buddhist Translators Workbench (<https://btw.mangalamresearch.org/en-us/>)

Secondary Sources

- Adams, Michael. 2010. Legacies of the *Early Modern English Dictionary*. In *Adventuring in Dictionaries: New Studies in the History of Lexicography*, edited by John Considine, 290–308. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Adams, Michael. 2013. *DARE*, dialect, and techniques of historical lexicography. In *Selected Proceedings of the 2012 Symposium on New Approaches in English Historical Lexis (HEL-LEX 3)*, edited by R. W. McConchie, Teo Juvonen, Mark Kaunisto, Minna Nevala, and Jukka Tyrkkö, 1–13. Somerville, MA: Cascadilla Proceedings Project.
- Atkins, Sue., and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Aklujkar, Ashok N., The early history of Sanskrit as supreme language. In *Ideology and the Status of Sanskrit*, edited by Jan M. E. Houben, 59–86. Leiden: Brill.
- Apple, James B. 2009. “Wordplay”: Emergent ideology through semantic elucidation: Rhetorical technique in Mahāyāna Buddhism. *Bulletin for the Institute of Oriental Philosophy* 25: 161–173.
- Aussant, Émilie. 2014. Sanskrit theories on homonymy and polysemy. *Bulletin d'Études Indiennes* 32: 13–36.
- Bergenholtz, Henning, and Jesper Skovgard Nielsen. 2013. What is a lexicographical database? *Lexikos* 23:77–87.
- Boisson, Claude, Pablo Kirtchuk, and Henri Béjoint. 1991. Aux origines de la lexicographie: Les premiers dictionnaires monolingues et bilingues. *International Journal of Lexicography* 4: 261–315.
- Bronner, Yigal. 2010. *Extreme Poetry: The South Asian Movement of Simultaneous Narration*. New York: Columbia University Press.

- Cardona, George. 1990. On attitudes towards language in Ancient India. *Sino-Platonic Papers* 15: 1–19.
- Colas, Gérard. 1999. The criticism and transmission of texts in Classical India. *Diogenes* 47.2: 30–43.
- Condotti, Maria Piera. 2011. Loke, vede, śāstre: Grammarians' partition of tradition and related linguistic domains. In *Boundaries, Dynamics and Construction of Traditions in South Asia* edited by Federico Squarcini, 389–410. London: Anthem Press.
- Cox, Collett. 2004. From category to ontology: The changing role of Dharma in Sarvāstivāda Abhidharma. *Journal of Indian Philosophy* 32: 543–597.
- Deleanu, Florian. 2006. *The Chapter on the Mundane Path (Laukikamārga) in the Śrāvaka bhūmi: A Trilingual Edition (Sanskrit, Tibetan, Chinese), Annotated Translation and Introductory Study*. 2 vols. Tōkyō: The International Institute for Buddhist Studies.
- Deshpande, Madhav M. 1985. Historical change and the theology of eternal Sanskrit. *Zeitschrift fuer Vergleichende Sprachforschung* 98: 122–49.
- Deshpande, Madhav M. 1996. Contextualising the eternal language: Features of priestly Sanskrit. In *Ideology and the Status of Sanskrit*, edited by Jan E. M. Houben, 401–38. Leiden: Brill.
- Dundas, P. 1996. Jain attitudes towards the Sanskrit language. In *Ideology and the Status of Sanskrit*, edited by Jan E. M. Houben, 137–56. Leiden: Brill.
- Fosse, Lars Martin. 1997. *The Crux of Chronology in Sanskrit Literature: Statistics and Indology. A study of method*. Oslo: Faculty of Arts, University of Oslo.

- Ghatage A. M. 1973. Traditional lexicons and Sanskrit historical dictionary. In *Studies in Historical Sanskrit Lexicography*, edited by A. M. Ghatage, R. N. Dandekar, and M. A. Mehendale, 25–39. Poona: Deccan College.
- Glynn, Dylan A. 2014. Techniques and tools: Corpus methods and statistics for semantics. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, edited by Dylan A. Glynn and Justyna Robinson, 307–41, Amsterdam: John Benjamins.
- Goyal, Pawan, Gérard P. Huet, Amba Kulkarni et al. 2012. A collaborative platform for Sanskrit processing, In *Proceedings of the 24th Computational Linguistics Conference*, edited by Martin Kay and Christian Boitet, 1011–28. Mumbai: Indian Institute of Technology.
- Healy, Antonette diPaolo. 2002. The *Dictionary of Old English*: From manuscripts to megabytes. *Dictionaries* 23: 156–79.
- Hellwig, Oliver. 2015. Using recurrent neural networks for joint compound splitting and Sandhi resolution in Sanskrit. In *Human Language Technologies as a Challenge for Computer Science and Linguistics*, edited by Zygmunt Vetulani, 289–293. Poznan: Adam Mickiewicz University.
- Hellwig, Oliver. 2017. Coarse semantic classification of rare nouns using cross-lingual data and recurrent neural networks. In *IWCS 2017: 12th International Conference on Computational Semantics - Long papers, Montpellier, France, September 19 - 22, 2017*, edited by Claire Gardent and Christian Retoré. The Association for Computer Linguistics.
- Houben, Jan E. M. 1996. Socio-linguistic attitudes reflected in the work of Bhartrhari and later Grammarians. In *Ideology and the status of Sanskrit*, edited by Jan E. M. Houben, 157–96. Leiden: Brill.

- Janda, Laura A., and Valery D. Solovyev. 2013. What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS. In *Cognitive Linguistics: The Quantitative Turn*, edited by Laura A. Janda, 295–324. Berlin: Walter De Gruyter.
- Jamison, Stephanie W. 1999. Review of Laura M. Fosse, *The Crux of Chronology in Sanskrit Literature: Statistics and Indology. A Study of Method*. *Indo-Iranian Journal* 42: 249–252.
- Kahrs, Eivind. 1998. *Indic Semantic Analysis: The Nirvacana Tradition*. Cambridge: Cambridge University Press.
- Kamboj, Jiya Lal. 1986. *Semantic Change in Sanskrit*. Delhi: Nirman Prakashan.
- Katre, S. M., and P. K. Gode. 1954. *Introduction to Indian Textual Criticism*. Poona: Deccan College.
- Kilgariff, Adam. 1997. I do not believe in word senses. *Computers and the Humanities* 31: 91–113.
- Kelkar, A. R. 1973. The scope of a historical dictionary. In *Studies in Historical Sanskrit Lexicography*, edited by A. M. Ghatage, R. N. Dandekar, and M. A. Mehendale, 57–69. Poona: Deccan College.
- Lew, Robert. 2009. Towards variable function-dependent sense ordering in future dictionaries. In *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*, edited by Henning Bergenholtz, Sandro Nielsen, and Sven Tarp, 237–64. Bern: Peter Lang.
- Lewis, Robert E. 2002. Aspects of polysemy in the *Middle English Dictionary*. In *Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology, at the University of Leicester, 2002*, edited by Julie Coleman and Anne McDermott, 149–56. Tübingen: Max Niemeyer.

- Moisl, Hermann. 2015. *Cluster Analysis for Corpus Linguistics*. Berlin: De Gruyter Mouton.
- Nattier, Jan. 2008. *A Guide to the Earliest Chinese Buddhist Translations: Texts from the Eastern Han 'Dong Han' and Three Kingdoms 'San Guo' Periods*. Tōkyō: International Research Institute for Advanced Buddhology.
- Narahari, H. G. 1973. Some problems in the chronology of Nyāya-Vaiśeṣika works. In *Studies in Historical Sanskrit Lexicography*, edited by A. M. Ghatage, R. N. Dandekar, and M. A. Medenhale, 89–94. Poona: Deccan College.
- Norman, K. R. 1984. On translating from Pali. *One Vehicle: Journal of the National University of Singapore Buddhist Society* 1: 77–87.
- Norman, K. R. 1997. *A Philological Approach to Buddhism. Bukkyō Dendō Kyōkai Lectures*. London: SOAS.
- Mishor, M. 2002. The philological treatment of ancient sources. *International Journal of Lexicography* 5.2: 132–138.
- Olivelle, Patrick. 2004. The semantic history of *Dharma*: The Middle and Late Vedic periods. *Journal of Indian Philosophy* 32: 491–511.
- Pruden, Leo M. (trans.) 1990. *Abhidharmakośabhāṣya of Vasubandhu. Translated into French by Louis de La Vallee Poussin. English Version by Leo M. Pruden*. Berkeley: Asian Humanities Press.
- Rehman, Zobria, Waqas Anwar, Usama Ijaz Bajwa et al. 2013. Morpheme matching based text tokenization for a scarce resourced language. *PLoS ONE* 8.8, e68178.
<http://doi.org/10.1371/journal.pone.0068178>
- Silk, Jonathan A. 2009. The nature of the verses of the *Kāśyapaparivarta*. *Bulletin of the Asia Institute* 23: 181–190.

- Schryver, Gilles-Maurice de. 2003. Lexicographer's dreams in the electronic dictionary age. *International Journal of Lexicography* 16: 143–199.
- Sheidlower, J. 2011. How quotation paragraphs in historical dictionaries work: *The Oxford English Dictionary*. In *Contours of English and English Language Studies*, edited by Michael Adams and Anne Curzan, 191–212. Ann Arbor, MI: University of Michigan Press.
- Skilton Andrew. 1997. *The Samādhirāja Sūtra: A Study Incorporating a Critical Edition and Translation of Chapter 17*. D.Phil dissertation, University of Oxford.
- Squarcini, Federico (ed.) 2015. *Patañjali: Yogasūtra*. Torino: Einaudi.
- Thapar, Romila. 2013. *The Past before Us*. Cambridge, MA: Harvard University Press.
- Vogel, Claus. 1979. *Indian Lexicography*. Wiesbaden: Otto Harrassowitz.
- Zenner, Eline, Dirk Speelman, and Dirk Geeraerts. 2013. Cognitive sociolinguistics meets loanword search: measuring variation in the success of anglicisms in Dutch. In *Cognitive Linguistics: The Quantitative Turn*, edited by Laura A. Janda, 251–94. Berlin: De Gruyter Mouton.

Note on Contributor

Ligeia Lugli is Newton International Fellow at King's College London, where she research on Buddhist Sanskrit lexicology. She graduated in Asian Languages from Ca' Foscari University (Venice) and received her Ph.D. in Study of Religions from SOAS (University of London) in 2011. She also studied at the Central Institute for Higher Tibetan Studies in Sarnath and at Ōtani Daigaku in Kyōto, where she read Buddhist texts in Sanskrit, Japanese and Tibetan. She was Research Associate in London in 2012-2013 and Visiting Scholar at the University of California

at Berkeley until 2016. Ligeia has been working on lexicographic projects involving a number of ancient and minority languages including Sanskrit, Tibetan and Limburgish.