ERC Advanced Grant 2016
Research proposal [Part B2]

Part B2: *The scientific proposal* (max. 15 pages)

Section a. State-of-the-art and objectives

The luxuriant cultural diversity of Asia, home to more than half of the world's population, has grown in a soil made fertile by an underlying and historically unifying Buddhist influence. Deep-seated aspects of the world-views of the great Asian civilizations—the Indian, the Chinese and so on—were foundationally and fundamentally shaped by Buddhist ideas and practices for more than 2500 years. The principal vehicle for the transmission of these ideas and practices is the **Buddhist sacred literature**, the ***sūtra*s**, traditionally believed to be records of the Buddha's sermons. Generations of Buddhists, seeking their timeless truths, have looked—and continue to look—to the *sūtra*s for guidance and wisdom. For scholars, on the other hand, these *sūtra*s offer a unparalleled window into the **diversity of Buddhist traditions**, particularly because, from their Indian origins through their subsequent spread to the edges of the continent, they were not subject to any centralized editorial effort or standardization.

As a direct result of this rich history, the ***sūtra*s also present us with profound challenges**, most significantly because of their **fluid**, **authorless nature**, the **many related versions** in which they have been passed down, their prolific **formulaic modularity and intertextuality** and, not least, the **variety of languages** in which they are preserved. Inheriting the legacy of Classical and Biblical studies, scholars long directed their efforts at recovering the earliest state of any *sūtra* under study, its 'original.' The **underlying methodological assumption** behind this quest, however, blinds us to the heterogeneity of the living, dynamic and vibrant communities which contributed to every text's history of production, growth and diffusion. Even editors with a broader awareness of these textual complexities foundered on the shoals of **the technical limitations of two-dimensional print** and, bound by the form of a main text accompanied by variant readings, found themselves unable to present editions or translations that respected the manifold nature of the literature. The history of Buddhism, its roles in Asian civilizational history, and the rich medley of forms in which the tradition expressed itself over the ages have thus remained obscured and misrepresented.

We now stand poised on the brink of a **revolution**. Taking advantage of **the flexibility of digital environments**, Open Philology will develop **architectures for presenting multiple related texts in an array of languages**, facilitating explorations of their highly **modular composition**—*composition* in the sense both of the way the texts are constructed, and of the processes through which that was achieved**.** For the first time, we will be able to honestly approach the Buddhist scriptural legacy in its spectacular variety, revealing it as a dynamic reflection of vibrant and ongoing processes of expression, processes that allow us to appreciate as never before the mosaic of Buddhist traditions over time and space.

We will achieve this **major development** through the creation of an **open environment** for editing, annotating, and translating Buddhist scriptures. This has two key elements: the preparation of a **multilingual and multi-recensional corpus**, and the creation of **detailed editions, translations and studies of individual texts**. Ultimately, careful treatment of the entirety of the extant Buddhist scriptural literature, which in Tibetan translation comes to some 70,000 pages, will be the work of decades. We will therefore focus our attentions on **a traditional sub-set of the canon**, the 49 *sūtra*s of the ***Mahāratnakūṭa* collection** (MRK), some 3500 Tibetan pages, a corpus of interest to the PI since his PhD (Silk 1994b), and one of central importance historically. The tools and the methods we will develop will help students of other literatures approach their own corpora, promising the project an impact far beyond the field of Buddhist Studies. In accord with Open Data, Open Access and Open Source ethics, all materials produced by the project, along with guidance to facilitate their use, will be made available to all without restriction or cost. Open Philology will provide us, at last, a means to engage, **openly and non-hegemonically**, the scriptural corpus of the Buddhist tradition, which has provided the intellectual and sacred foundation for Asian societies for more than two millennia.

Background

According to **Buddhist tradition**, in the 40 years of his teaching career the Buddha (± 5th c. BCE) wandered preaching from village to village. He must have given 'the same sermon' multiple times, each time, however, a bit differently, as a bard repeating a tale and, like a bard, he would have employed formulae to facilitate his discourses. In an age when memory ruled, these sermons were held in mind, in whole or in part, and the lessons available to be passed on orally, even if memorized verbatim by his followers, came to constitute a circulating **collection of variant—but nevertheless *equally authentic*—versions of the Buddha's teachings, the *sūtra*s**. These *sūtra*s, together with the monastic code (*vinaya*) through which the Buddha's community of followers was governed, were gathered by his disciples at a Council after his death, the two collectively constituting the core 'Buddhist canon' (on the complexities of the category, Silk 2015b). According to

tradition, alongside these 'early' *sūtra*s the Buddha also preached a more profound doctrine, the 'Great Vehicle' (sc. leading toward awakening, in Sanskrit *Mahāyāna*), for which the world, however, was not yet ready. These *sūtra*s were thus not revealed until long after the Buddha's death. The historicity of this scenario aside, it precisely reflects the actual situation: *sūtra*s are constituted of multiple closely related versions which, however, cannot stem from a unique archetype or 'Ur-text'. Moreover, as a result of their compositional process, discrete texts share substantial formulaic materials. Whether anything transmitted through the generations literally reflects what was spoken by the Buddha is unknown, but for historians, when the Mahāyāna movement arose some centuries after the Buddha's death, its proponents presented their new insights and understandings in the most hallowed format available, the *sūtra*. These **Mahāyāna *sūtra*s**, produced from the beginning of the Common Era until the 5th c., are, like the earlier *sūtra* literature upon which they are modeled, fluid and highly modular compositions. The MRK, the focus of the OPEN PHILOLOGY project, consists of a **representative selection of these Mahāyāna *sūtra*s**.

We generally presume that any text reflects the concerns of its author—but *sūtra*s do not have 'authors' as we are wont to think of them. Traditional, linear approaches to editing assume an **author and authorial intention**, the aim most commonly being to reconstruct a text as close as possible to that which left the author's pen. While valid in some cases, in many others this approach can fundamentally misrepresent the nature of the literature under study. Well-known instances are the oral bardic and Homeric literatures studied by Parry and Lord (e.g. Lord 1960). These literatures have only particular instantiations, discrete recitals and tellings which may have been recorded in writing, often in more than one time and place, but which by their nature partake of no unique original Ur-form liable to 'recovery'. Buddhist scriptures, as different as they are from oral poetry, are also **highly formulaic**. Although we know little of the actual mechanics of Buddhist text production, it was clearly facilitated by an open **pool of pericopes**, stock elements upon which authors—if one may call them that—drew. To understand these processes of composition in their historical complexity, **we must replace linear models which posit a base text** accompanied by 'variants' with approaches that model the literature's innate fluidity.

As Buddhism spread north and east out of its homeland, the *sūtra*s were not generally transmitted in Indic languages, such as Sanskrit and Pāli, but rather translated. Due to the demise of Buddhism in India by the 13th c., and the subsequent loss of the vast majority of the Indian scriptural legacy in its Indic language forms, **Mahāyāna *sūtra*s**, such as those collected in the MRK, are **chiefly available in translations in Chinese (from the 2nd c.) and Tibetan (from the 8th c.)**. The sources taken as bases for the respective translations into Chinese and Tibetan, their Vorlagen, being in each case merely one among a myriad of circulating Indic versions, were consequently also not identical, either to each other or to other versions to which we might have access today. As a result, the *sūtra*s exist in multiple versions, in multiple languages, in forms in principle not related to each other hierarchically or stemmatically. No single version, therefore, deserves more than provisional and situational priority. The challenge facing us is how to present such a text, devoid of any solid core, in a fashion that does not impose an arbitrary pivot onto the fluid textual situation. It is one thing to think this through theoretically (Silk, 2015c, 2016), another to actually do it. OPEN PHILOLOGY represents an effort to do precisely this.

The Choice of the Mahāratnakūṭa *Collection* (*MRK*)

The MRK provides an excellent basis upon which to build the tools necessary for this task, drawing our particular attention for both practical and principled reasons. The collection, which exists now in Chinese and Tibetan (for 10 out of the 49 *sūtra*s we also have some Sanskrit material), is preserved in 120 traditional Chinese volumes (roughly 1 million Chinese characters) corresponding to 282 traditional Tibetan volumes (structurally this Tibetan counterpart is based on the Chinese), a size which feasibly permits study by a single team. These 49 texts represent a cross-section of the types of Mahāyāna *sūtra*s found in the broader Mahāyāna canon; indeed, while the exact logic of the composition of the collection as a unit is not yet known, one has the impression of a sort of 'buffet', with one *sūtra* chosen from each thematic category in the canon. The MRK may, in fact, have been *intended* as a sort of '**mini-canon**,' a hypothesis which arises not only from this thematic variety but also from the circumstances of its composition. For the texts were edited, in some cases translated anew, and compiled together into their current form by the monk Bodhiruci in Tang dynasty China, officially presented to the court in 713. The context of Bodhiruci's work is crucial. The year 690 saw the enthronement of the Empress Wu Zetian (r. 690–705), the only woman to rule China in her own name. Although not published until 8 years after her death, I hypothesize that Bodhiruci's project (see esp. Forte 2002) was deeply connected with **Empress Wu's efforts to establish a Buddhist realm** in medieval China, or with the efforts of the Buddhist community, which had enjoyed her special patronage, to secure its ongoing privileged position after her fall from power. Many questions confront us in our efforts to understand **the historical situatedness of the MRK**: What is the true background of this collection? Was it intended as a 'best hits' collection from the Mahāyāna *sūtra* literature? Why were some texts translated anew, while in other cases existing translations were adopted? Why are the *sūtra*s ordered as they are? A study of this collection is consequently of great value not only as a **representative collection of Mahāyāna scriptural sources**—although it surely is this—but also in its historical context in relation to general issues including those of Church and State, gender and power, and the propaganda uses of religious literature. The individual texts in

the collection present us with **a variety of editorial situations**, **a wide range of doctrinal stances, rhetorical modes, literary styles and religious attitudes** and offer, moreover, an excellent opportunity to **map and begin to understand the modular and intertextual nature of the *sūtra* corpus**. One could hardly ask for a more ideal body of material upon which to work. The individual projects detailed below contribute, both jointly and severally, to the overall aims of the project by highlighting specific issues in the text-critical and editorial treatment of the literature, and by examining core themes of the individual *sūtra*s under study, and of the collection as a whole.

Key Concepts

Since 1844 when E. Burnouf published his epoch-making study of the history of Indian Buddhism, scholars of Buddhist scriptures were principally Indologists, who focused their attention on works which survive in Indic languages, leading them to pay less attention to the bulk of Mahāyāna *sūtra*s. The vast majority are even today yet to be edited, translated or carefully studied. When these texts *are* studied, the usual manner of their presentation only accentuates the limits of traditional editing practices. The PI himself, though now having come to recognize the problem, had earlier publishing editions which collected variants and suppressed readings other than those selected for the main text, placing such readings in a 'single readings apparatus' and even referring to them as 'noise', as opposed to the 'signal' of the established text (Silk 1994a, 1994b). Even the best editions of Mahāyāna *sūtra*s are similarly based on an assumption of an Ur-text, and present rejected readings in a fashion that renders them effectively inaccessible. Even concerted efforts to reconstruct from published critical editions the sources upon which they are based are not likely to succeed. The most important problem here is not merely that the efforts of scholars are thus wasted, or that one might disagree with an editor's choices in a given case. The problem is rather that different users come to texts with different questions, and an edition which does not allow a user to ask her own questions is not useful to her. (An example of a difficult-to-foresee use might be that of the historical linguist who is principally interested in 'obviously mistaken' spellings for the evidence they may provide of the dialect or mother-tongue of a scribe.)

Although our problems differ, in part in their much greater scale and scope, due to the breadth and diversity of the literature and the multiple languages in which it is preserved, scholars of Buddhist Studies are not alone in urgently requiring a **conceptual revolution** to deal with literature whose composition is fluid and amorphous. Those in other fields have thought about similar challenges, and our work often dovetails with theirs. Scholars of Rabbinic literature, for example, offer a typology descriptive of a compositional situation similar to the one we too confront. Milikowsky (2006, 82) helpfully differentiates logical levels: a '*Work'* is what is produced by an author or editor, though it "may theoretically never have existed in any concrete mode of expression such as a manuscript or book," to which we would add 'oral recitation.' Less abstract is a '*Document*,' a concrete mode of expressing a work, while a '*Text'* is an actual word-after-word presentation. The largest unit, the Work, at its broadest may be no more than something like a hypothetical generic class. What really exist are multiple documents (including recitations), whose words—the Text—(re)present the Work. We refer to these actual documents as 'witnesses,' the conveyors of the text.[1] A crucial theoretical question asks in what sense, then, if we read a single version we can consider ourselves to be studying the Work, which we generally presume to be the ultimate object of our interest. We require a model that allows users to change the resolution or the granularity at which they look at a particular object, our gaze falling on the Work, on the Document, on the Text, on a phrase or even word, as circumstances dictate.[2] How can we present these? There are traditionally said to be two forms of edition: the diplomatic edition, a strict transcript of a source, and the eclectic text, anything other than an exact transcription, a text "with at least one deviation from the text of the document serving as [its] base" (Milikowsky 2006, 86). As long as we imagine the limits of printed output, these are indeed the only general possibilities. But **the tools of Digital Humanities now allow us to think outside the narrow bindings of books**.

Further comparisons can clarify our thinking about the nature of Buddhist texts, and what it will take to treat them fairly. The Hebrew Bible text constituting the *textus receptus* is found in manuscripts agreeing extremely closely among themselves, and is tracked very closely by the Septuagint and other translations in Aramaic and Syriac. These sources present a tradition with few variations. While we are largely ignorant of the pre-redactional history of the Bible, we do know, most importantly from Qumran, that the true historical situation,

---

1. This bibliographic hierarchy approximates the "FRBR" (Functional Requirements for Bibliographic Records) scheme, a 1998 recommendation of the International Federation of Library Associations and Institutions (IFLA Study Group 1998). Humanist scholarship extends beyond the bibliographic level into the citation structure of a work and its contents. One of the models for part of what we will accomplish, the Homer Multitext Project, has extended and adapted FRBR to allow identification and retrieval of texts at any level of granularity, from the notional "text group" (e.g. "Homeric epic") and (notional) "work" (e.g. "Iliad") down to specific tokens in specific versions of a text, e.g. "the second instance of the letter 'alpha' in Book 1, line of the Iliad as it appears on Manuscript A." The Canonical Text Services protocol captures these semantics in a concise, technologically agnostic, machine-actionable scheme of citation, the CTS-URN: Blackwell and Smith 2012, 2014a, 2014b. This work may be of great utility to us as we work to develop our own architectures, and C. Blackwell, Information Architect for the Homer Multitext Project, has eagerly offered his guidance and assistance.

2. This reminds us of computer scientist Katy Börner's notion of the "macroscope": "Macroscopes provide a 'vision of the whole,' helping us 'synthesize' the related elements and detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, macroscopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend" (Börner 2011, 60).

rather than being uniform, is closer to what we see with Rabbinic literature, in which "[E]very writing has two histories: namely, a pre-redactional and a post-redactional history. In the middle of these two histories stands firmly and unshakably the zero-point. [...] The redactional identity of a work happens at this zero-point. All that precedes it is not yet 'work' but 'sources used by the redactor.' All that follows belongs to the 'history of transmission' of the work defined through the zero-point of the single redaction" (Schäfer 1989, 9). This zero-point—which can be visualized as the narrow neck of an hourglass—represents a form of a theoretically recoverable Ur-text, even if it does not stand at the very fount of the tradition. In the case of the Hebrew Bible, and with Classical literature, the corpora on the basis of which modern text criticism as a field was developed, this is almost all that remains. (For a lively account of the history and practice of Classical text-critique, see Trovato 2014.) *Sūtra* literature is different, in that we have no such generalized zero point, but similar in that when we *seem* to encounter uniformity, it is overwhelmingly due to a paucity of evidence, other more textually fluid witnesses—witnesses carrying text which varies from that otherwise preserved—having been lost. We see typologically similar processes elsewhere: "For Homeric epic, the relative uniformity of the medieval manuscripts is the accident of transmission, and multiformity is the natural result of the process by which they were created".[3] Buddhist scriptures also resemble Rabbinic literature, and in some respects Homer as well (cf. Nagy 2001, 116), in their large-scale use of pericopes, formulaic and stock elements, free of definable stages of pre- and post-redaction: hence, no zero-point and no Ur-text. The same issues arise in the study of texts culturally closer to the Buddhist *sūtra*s, including the Indian Epics, the *Mahābhārata* and *Rāmāyaṇa*, and the *Purāṇa*s, but possible phenomenological and historical ties between these Indian literatures remain barely explored. A reasonable hypothesis links efforts to control the diversity of texts, and thus to define canonicity and orthodoxy, to exercises of power, whether that power be located in the Temple in Jerusalem, in the Library in Alexandria, or in the government bureaus of the Imperial Chinese state. These questions too, for Buddhism, remain largely untouched, and their investigation might teach us more general lessons that we cannot yet foresee.

The typological distinction between Work, Document and Text, and the notion of the intertextuality of pericopes, relate to another helpful set of concepts invoked in discussions of Rabbinic literature, **the microform and the macroform**. The microform is an independently transmitted redactional unit, while the macroform is both a "fictional or imaginary single text" and the "manifestations of this text in the various manuscripts." (Schäfer 1992, 6n14). Although the genres differ, much of what the folklorist L. Honko (2000, 18–19) says applies to Buddhist literature as well: "[T]he elements are free to vary and combine, and it is in the variation and combination of multiforms, themes and formulas that the individual novelty can be found. [...] a pool of generic rules, storylines, mental images of epic events, linguistically preprocessed descriptions of repeatable scenes, sets of established terms and attributes, phrases and formulas, which every performer may utilize in an imaginative way, vary and reorganize according to the needs and potentials present at a new performance. [...] Whatever is shared by more than one singer belongs to the pool of tradition. The pool holds a multiplicity of traditions, a coexistence of expressive forms and genres, mostly in a latent state, only parts of it becoming activated by the individual user." While this scenario describes central aspects of Buddhist scriptural composition, it leaves us where Schäfer (1986, 150) was left—with a series of questions: "How do different recensions of a 'text' relate to one another in respect to the redactional identity of the text? How should the individual tradition, the smallest literary unity, be assessed in relation to the macroform of the 'work' in which it appears? What is the meaning of the presence of parts of one 'work' in another more or less delimitable 'work'? What is redaction or final redaction? Are there several 'redactions' of a 'work'—in chronological order—but only one final redaction? What distinguishes redaction from final redaction? What lends authority to the redaction? Or is the final redaction merely the more or less incidental discontinuation of the manuscript tradition?" These questions cannot be answered meaningfully in the abstract, but only by means of careful examinations of actual text traditions. The MRK is a perfect body of material with which to approach such questions.

Specific Challenges

Single exemplars are valuable to us above all *as witnesses to* a Work, and we must be careful not to confuse them with the Work itself, in particular since we aspire to keep in view the Buddhist tradition as a whole, its historical gestalt, rather than aiming at an atomized history of a single version. Privileging a single version (which may, of course, be transmitted in multiple witnesses) is an emic, and **inherently teleological**, stance, since it assumes the (logical or chronological) priority of that version when in fact, viewed etically, it is merely contingently valued by a certain group. Our approach, however, in no way conflicts with traditional Buddhist attitudes toward their own sacred literature, attitudes which *do* privilege a given version: the approaches belong, rather, to two entirely distinct domains. Emically speaking, Buddhists—even editors of texts—see only a timeless text. Transmissional errors can be recognized, and editors (such as the Korean monk Sugi, who edited the Chinese canon in the early 15th c.) did often seek a 'best text.' The tradition, nevertheless, always saw the scriptures synchronically, or better, ahistorically. Consequently, the goals of our scholarly approach to

3. Dué and Ebbott 2009, 25. West 2001 engages with the position of Nagy 2000, and implicitly of his students Dué and Ebbott, and these debates help us to clarify the larger issues at play.

diachronic textual history, and traditional Buddhist attitudes toward the same literature, are typologically and conceptually distinct.

Our concern with history and with the tradition as a whole precludes us from limiting our gaze to any particular local community and the version of a text it might have come to inherit. We need instead to think about how to approach **simultaneous treatment of witnesses** which contain what is, after all, the 'same' Work, although each may articulate it in different words. We need to think of practical solutions to the theoretical challenges outlined above. But some avenues are not open to us.

We cannot, in particular, follow the path of the eclectic presentation. Such an edition accepts certain readings, and rejects or suppresses others, because it assumes an archetype. This path leads immediately to what J. Grigely memorably called "**Textual Eugenics**" (1995, esp. chpt. 2), a crucible in which is created the "engineered superior version" of a work (Phillips-Rodriguez 2007, 167). We tend not to think of our editorial work in such Nietzschean terms, but what else could it mean to read a Work when we have contact only with its static instantiations? Our access to the Work is always and inherently partial; it exists only in the imagination, not in words. Cannot we, then, as is most usual, represent the **multiformity of our sources** by establishing a text along with its 'variants'? Albert Lord, speaking of bardic oral songs (1960, 101), emphatically denies this: "we cannot correctly speak of a 'variant,' since there is no 'original' to be varied!" Morally speaking, the problem is perhaps even worse: **by making something a 'variant,' we hierarchize**, a process inherently complicit in the eugenic program of ourselves assuming and asserting authority over the Work-cum-text. This path could also bring us into conflict with the tradition, for it has the potential to assert that a given reading is correct, and others—perhaps traditionally sanctioned by faith communities—are wrong, something we are simply never in a position to do. In addition, the eclectic edition is based on the **metaphor of corruptions**, which are to be identified and eliminated, "the remorseless corrupting influence that eats away at a text during the course of its transmission" (Bowers 1959, 8). **We categorically reject this picture** of textual composition and history (while acknowledging that transmissional errors, "corruptions," do of course occur).

A Way Forward

We are not alone in thinking about such problems, and Open Philology situates itself in the ongoing discussion, but what uniquely characterizes *our* task is the complexities of our materials, and their historical depth. E. Vanhoutte offers an alternative path forward to the editing crisis in his electronic edition of the 1948 Flemish novel *De trein der traagheid* (http://edities.kantl.be/daisne/index.htm), which "guarantees the completely equal treatment of each version of the text in the generating processes invoked by the user." Such an approach "deliberately puts some central concepts and issues of conventional textual scholarship in crisis. Amongst them the base text, the edited text, the textual apparatus, and the variant. All of these concepts are dependent on the static perception of the scholarly edition" (2007, 165–166). To avoid Lord's problem of variants in the absence of an invariant core, Vanhoutte appeals to van Hulle (2004, 514) for the idea of *relative calibration*. "[T]he conventional absolute classification of variants has to be replaced by a relative classification which depends on the specific moment of calibration. This means that the class to which a variant belongs is no property of the variant proper, but of the orientation of the set of witnesses in the collation" (Vanhoutte 2007, 166-167). No editor can reject the idea that some readings have chronological priority over others; van Hulle's approach, however, **allows a user to set parameters against which 'variants' will be arranged**, rather than insisting on the priority of the oldest recoverable state of a text.

Given the challenge of presenting a Work, which by definition does not have a literal core, for practical reasons it will be best to use a system in which, at a basic level of text constitution, a **frame text is supported by a comprehensive interactive linkage**, on the lexeme level, to annotated diplomatic versions of each textual witness. A hypothetical Ur-text can serve as a baseline for textual presentation, including provision of a standard numbering scheme for the text, keeping in mind that it is a heuristic fiction rather than a historical (re)construction. Although full flexibility of access to all aspects of the textual record is essential, simplified access is equally important.

All models by definition simplify; the challenge is to simplify complexity with as little distortion as possible. While it is essential to prepare and present *all* data, most users will not need—or want—this. 'Good editions' are considered good because they answer most of the questions most users might have. When we agree with their criteria and their aims, we are generally fully justified in trusting the choices made by editors. What is more, it often makes pragmatic sense for an editor to choose for the 'main text' a reading that can be demonstrated to be oldest, and which produced, through error or emendation, other readings. From this point of view, the editor performs a vital task in making choices which most readers will welcome, whether their interest in the text be historical, doctrinal or, in the case of religious texts, faith-based. In the case of Chinese or Tibetan translations produced at a specific historical moment, we are, moreover, usually justified in assuming a genetic relationship among witnesses to the text, and thus it is reasonable to think that the above-mentioned 'zero-point' actually exists. An editor will then often be able to determine with confidence which readings of a specific translation—though not necessarily of the Work as a whole—stood earlier in the tradition, and which represent transmissional errors or emendations. We may, in some cases, also be able to correlate the hypothesized sources of these translations, to imagine, to some extent, an Indic source which

Chinese and Tibetan translations both, respectively, reflect.[4] The provision of a tentative and hypothetical frame text simplifies some of the complexity our model engages.

A Test Case

A **concrete example of the challenges we face** with individual Buddhist Works comes with the *sūtra* which forms the historical core of the MRK, the *Kāśyapaparivarta* (MRK 43), comprehensive research into which will be carried out together by all team members as a **group exercise**. (In fact, the historically correct title of this *sūtra* is not *Kāśyapaparivarta* but *Ratnakūṭa*, from which the MRK collection took its name.) This text presents all at once most of the issues we encounter in the corpus at large, and this very complexity advances it as a superb candidate for communal work and discussion. As sources we possess, principally, versions in Sanskrit (7th c.), Tibetan (8th c.), Chinese (in five translations, 174 CE to the 10th c.), a commentary which quotes the text in extenso (extant in Tibetan and Chinese), and numerous quotations in later philosophical works (Silk 2009, 2010, 2013); many passages find parallels elsewhere, highlighting the high degree of intertextuality the text presents. With the exception of the Sanskrit, for which we have a single manuscript, some fragments and quotations, each version is represented by multiple witnesses. Some of the theoretical challenges posed by this material can be accentuated by a few focused questions: In order to establish an edition, should we take the Sanskrit manuscript—which is centuries newer than the oldest Chinese translation—and omit some of its passages from our frame text because they are absent in (one or more versions in) Chinese, on the grounds that these passages are lacking in a witness redacted at *another* moment and place, a moment *unrelated* to the redactional process that generated the extant Sanskrit manuscript itself? What do we do with the paragraphs of the text that are indeed closely parallel in different versions, but appear in different orderings? If the macroform "the" *Kāśyapaparivarta* did not grow in a linear fashion, such that no historical tree diagram of its expansion is possible, how should we treat the relations among witnesses, versions and the Work? On the other hand, if we were to treat each version, or even each witness, separately, how would we justify using one family of evidence to shed light on another, or even compare or contrast them at all? That is, if we deny their genetic relation, would we ever be justified in 'correcting' one version based on another? What status should be accorded parallel passages found also in other texts: in what sense do they 'belong to' the *Kāśyapaparivarta*, if they also 'belong to' other texts? Similar questions are posed by the study of each *sūtra* of the MRK collection, and therefore our group work on the *Kāśyapaparivarta* and our work on our individual projects will promote a synergy, each study offering insight to every other inquiry.

Section b. Methodology

In the face of the challenges sketched above, it sounds as if, in practical terms, the diversity of our sources leaves us no choice other than an exhaustive listing; we must renounce the idea of editing all together. This, happily, is certainly not the case. Following Davila (1994, 220), we hold that "the ideal way to study is to create a massive critical edition that reconstruct[s] every level of development of the document in all MSS [= manuscripts], from the earliest redactional levels to the forms in the latest and most expanded MSS." This is the only way to assure non-hegemonic treatment of all phases of the tradition, and thus open access to the historical instantiations of the text. An ideal edition aims at an accounting of all of the evidence, in a historically aware framework. This is a far cry from the mechanical process implied by exhaustive listing, requiring at each stage careful consideration of the mutual relations between witnesses, and a clear determination of just what it is that they are witness *to*. It is also the only way to assure egalitarian representation of the richness of the text tradition, and the breadth of possible uses to which scholars and believers alike might want to put these results. Moreover, by allowing us to highlight intertextual commonalities, it positions us to begin to address questions of interest far beyond Buddhist Studies.

OPEN PHILOLOGY seeks to understand both the vertical history of a text, its historical dimensions and diversity, and its horizontal history, its relations with 'other texts' with which it shares content, themes or motifs. There are thus **two main aspects to the project: 1) the general preparation of the MRK corpus, and 2) the preparation of editions of select texts**. The former is a team task, the latter a series of individual tasks, but conceived of in a synergistic way: the individual editions are not disjointed efforts but rather woven together with the whole, constituting its warp and woof. Two major software elements will furnish the environment for realization of our editing vision. One will provide for mark-up and alignment, the correlation of parallel MRK corpora in Chinese and Tibetan (and, when available, Sanskrit); the other will provide a digital working environment for the open editorial treatment of individual texts. In the sketch that follows, we outline methods to achieve these goals.

These environments must be created and tested. For this we will execute an initial **pilot project**. Several years ago the PI identified two texts in a Dunhuang manuscript kept in Paris, Pelliot tibétain 89, the *Gaṅgottaraparipṛcchā* (MRK 31) and the *Maitreyaparipṛcchā* (MRK 42), as Tibetan translations from Chinese originals (Silk

4. A correlation between the two translations may, moreover, help us make decisions about how to edit each version. If, for example, one reading in a Chinese text would conform to what is found in the corresponding Tibetan translation, while another possible reading would not, this would support identification of the latter reading as secondary (Silk 2011). A study of such overall patterns would be of great value.

2014a). Corresponding Tibetan translations of these *sūtra*s from Sanskrit are also extant, respectively Derge 75 and 86. The second of these texts has two Chinese translations, and a small amount of material preserved in Sanskrit. The PI prepared aligned editions and translations of all sources of the *Gaṅgottaraparipṛcchā* (based on 9 Tibetan witnesses and three Chinese), his student C. Li prepared the same for the *Maitreyaparipṛcchā*. This already prepared material thus constitutes a small-sized sample of MRK data and is perfect as the basis for a pilot project for the alignment and editing environments (see below for the *Phases of the Project*).

Sources and their Treatment: Alignment

The term 'alignment' refers to the ability to locate parallel passages in different versions of a text. When one looks at, for instance, a bilingual edition of a poem of Goethe, the German on the left-hand page has been aligned with the English on the right. The absence of aligned corpora of Buddhist texts is not merely a matter of inconvenience to the scholar, who might wish to quickly consult other versions of a given passage. The more fundamental problem is that, given the fluidity of the textual tradition, **lack of aligned corpora conceals from the reader the innate diversity of the literature**. This difficulty in readily seeing the variety of *parallel but different* versions of the same text, and parallel passages in other texts, imposes the **highly misleading impression of a uniformity to the textual tradition** which it, in reality, does not possess. Consequently, the absence of aligned corpora lure us into seeing the historically protean Buddhist tradition as changeless.

Alignment of Chinese, Tibetan and Sanskrit Buddhist texts has so far been very limited. Even the few existing alignments of different versions within a given language—e.g. the alignment of multiple Chinese versions of the same *sūtra*—are sparse and rudimentary. We possess very few editions allowing a scholar interested in a given Chinese *sūtra* passage to easily compare its correspondent in Tibetan, or even in another Chinese translation. There do exist general identifications of correspondences of whole text units in Chinese and Tibetan. These texts in turn can have internal divisions, such as chapters, and some catalogues list these alignments, usually very roughly (the best example is Sakurabe 1930-1932). A more precise and useful alignment, however, remains to be done. First steps have been made monolingually in Tibetan by the main database clearinghouse for the Tibetan canon, the University of Vienna "Resources for Kanjur & Tanjur Studies" project, led by H. Tauscher (https://www.istb.univie.ac.at/kanjur/xml4/xml/index.php), which has agreed to cooperate with us in this regard, but no similar projects exist for the Chinese canon.

We have secured access to high-quality (multiply proof-read) machine-readable text corpora: the **Chinese Buddhist canon** (*Dazangjing*), in the standard Taishō edition (1924-1935), from **SAT** Chinese Tripiṭaka project of the International Institute for Digital Humanities (Tokyo), and the **Tibetan Buddhist canon** (*Kanjur*), Derge edition (1733), from the Buddhist Digital Resource Center, **BDRC** (formerly the Tibetan Buddhist Resource Center, TBRC).[5] These will serve as the first bases upon which we will build our alignments. Because critical alignment relies for its fine detail on a careful collation of all witnesses, something which is only possible through the painstaking study of each individual text, the **corpus-wide alignment** of the Chinese and Tibetan MRK collections will be subject to further adjustment as we, and other scholars in the future, carefully study each individual text. A preliminary result, however, which is nearly perfect 'for all intents and purposes,' can be achieved on the basis of the heretofore standard text corpora. We will produce this alignment for the MRK in the first years of our project, a process that will be facilitated by the fact that both the Tibetan and Chinese canons in digital form are punctuated (the former reflecting tradition, the latter produced by modern scholars); punctuation will provide one of the keys for future automated alignment of the broader corpora.

In light of the limited size of our corpus, we will prepare a **manual alignment** of the MRK; this in its turn will be used as **training data for alignments of the canon as a whole** to be produced through trained algorithmic processing (on grounds for optimism, see Xu and Chen 2011). For automated alignment, we will use a version of the Needleman-Wunsch algorithm to produce character-level alignment among Chinese texts. For sentence level alignment between Chinese and Tibetan, the best choice at present appears to be Open Source GIZA++, which is explicitly designed for bilingual corpora (in the environment called Moses, http://www.statmt.org/moses/?n=Moses.Overview). Alternatively, Open Source Doc2Vec (Le and Mikolov 2014) is a machine learning-based distributed representation of a passage. It modifies the word2vec algorithm (Bengio 2003) to perform unsupervised learning on larger blocks of text, such as sentences, paragraphs or entire documents. This would provide great help to our alignment task. Open Source CollateX (http://collatex.net/; Dekker & Middell 2011) facilitates tokenization of the text, that is, breaking it into component 'words,' and this software can be embedded in a larger environment, or be run as a standalone. (It must be noted that when studying, editing and translating individual texts we will, of course, take full account of all available Sanskrit evidence. However, the problems of dealing with Sanskrit computationally are so formidable [Nath Jha 2010] that, in terms of corpus preparation, we will eschew the Sanskrit materials, aside from manual alignment and numbering.)

The Natural Language Processing (NLP) problems associated with automated bilingual alignment can be

5. A version of the Tibetan canon in some ways parallel to the Taishō edition, which cites minimal variant readings, is the recent *Bka' 'gyur* (*dpe bsdur ma*) [Comparative Edition of the Kanjur], Tibetan Tripitaka Collation Bureau of the China Tibetology Research Center. 108 volumes. Beijing, 2006-2009, based on the Derge edition with references to 6 other editions. Its collations, however, are not reliable, and its format of presentation makes correlations between the printed text and its sources impractical.

substantial. Although these processes for Chinese present challenges (see most recently the encouraging results of Deng et al. 2016), these are significantly less serious than those presented by Tibetan (lemmatization of which will profit from cooperation with N. Hill through his project "Tibetan in Digital Communication: Corpus Linguistics and Lexicography"; http://gtr.rcuk.ac.uk/project/47AAC0E8-8058-447B-962C-31F290FE6503, and particularly his team's cutting-edge work on Tibetan Part-of-Speech tagging, Garrett et al. 2014). In contrast to the general case with Classical Chinese as a whole, the tokenization, lexicalization and classification of individual items of the Chinese Buddhist canon is greatly facilitated by the existence of a ready-made listing of technical terms, names and so on with a detailed ontology (or classifying organization) in the published index to the Taishō canon (Daizōkyō Gakujutsu Yōgo Kenkyūkai 1926-1985). As a **pilot program**, at our instigation the Center for Evolving Humanities at the University of Tokyo **digitized the index of the MRK, and applied its ontology to the base Chinese text**. The result is being proof-read at this writing. We will use this as the basis of mark-up (see below), to which we will subsequently add data drawn from specialized dictionaries, such as those of plant names (Waku 1979) and metaphors in Buddhist texts (Mori 1988); we have further obtained permission to integrate material from the Digital Dictionary of Buddhism (http://www.buddhism-dict.net/ddb).

A crucial step toward individual text preparation will be the **provision of standards** to the Chinese and Tibetan texts (perhaps with TEI compliant XML, but see Schmidt 2014 for caveats). We will cooperate with the SARIT project (B. Kellner, Vienna, director), which has developed a set of encoding guidelines (http://sarit.indology.info/exist/apps/sarit/docs/encoding-guidelines.html) that may provide a point of departure. However, we will not duplicate the work of projects such as SARIT which, like the better known Perseus project (http://www.perseus.tufts.edu/hopper) for Greek and Latin, so far makes printed editions available digitally, albeit in searchable and linkable format. We aim to do something different, taking advantage of the possibilities inherent in a native digital environment.

Mark-up

The term '**mark-up**' has a broad semantic range, but is used here to indicate a process by which structured information is attached to a term. For the Chinese character 馬, for instance, we might want to indicate its pronunciation(s), meaning (horse), possible Sanskrit and/or Tibetan equivalents, that it is a mammal, etc. All of this data can then be used to interrogate 馬 as it appears in the text corpus. Some of this information can be associated with the Unicode encoding of 馬, but other elements must be stored elsewhere. A good candidate for the model of this data is the Resource Description Framework (RDF). A flexible and extensible ontology can be developed to identify the base text and its variants, specify terms, annotations can be attached, and the resulting Triplestore can be queried using a semantic query engine such as SPARQL.

Text Reuse and Parallelism

When, previously, scholars have noticed the modularity of Buddhist scriptures, their sharing of stock expressions and pericopes (themselves often fluid in their expression, rather than literal repetitions), they have usually sought to explain this phenomenon through the scenario of borrowing: when Mahāyāna *sūtra*s share material with works assumed to be chronologically earlier, the former were considered to have reused material from the latter; when two Mahāyāna *sūtra*s share material amongst themselves, one was assumed to have been influenced by the other. A more fruitful, and more historically valid, hypothesis imagines a large volume of floating material, or a pool from which compilers were able to draw (explored in Silk 2014b). Following our rejection of the source-borrower model, studies of **text reuse**—identification of which is also a form of alignment—will contribute importantly to our project.

For text reuse specialists, "Quoting a text passage always implies purposeful re-use" (Büchler et al. 2013, 65), and hence influence. In contrast, we recognize shared material without implication of (chronological or logical) priority.[6] Results of reuse-like studies of Buddhist scriptures, therefore, indicate **the pervasion of a certain set of expressions of ideas**, rather than the influence of a specific work. Our focus is on ideas and their articulations, not on some specific (putative) original source for those ideas. Since we imagine a shared cultural base, the mappings (the visualizations of reuse patterns) that will result from our reuse studies could produce a virtual picture of the interrelations of the (so far unidentifiable) communities which produced the *sūtra*s, and within which they evolved.[7] There are a number of software packages well suited for such visualizations, such as the Open Source Cytoscape (http://www.cytoscape.org/index.html).[8]

---

6. This does not rule out real reuse, which certainly takes place, for instance when treatises cite proof texts from scripture; we thus distinguish 'reuse' from 'citation'.

7. An earlier study by the PI (Silk 2002, 374-375) suggested the use of Cluster Analysis in such a scenario, but at that time practically speaking it was not possible to carry the idea further.

8. An area of common interest to a number of scholarly communities is identification of nonliteral parallelism. This quest has some important features in common not only with efforts to track text reuse, but also with modern plagiarism detection. A fundamental question, connected with abstract considerations on the nature of a Work, is: how can we identify two blocks of text which 'say the same thing' when they do not share the same wording? This formulation differs in some important respects from text reuse studies which rely on shared lexical items. (Bamman and Crane [2009] suggest identification of sentences using dependency grammar, an excellent choice interlinguistically. A question they raise, however, is the size of the window of reuse, and the corresponding granularity of the searches. They also point

Traditional commentators may cite proof-texts for a certain doctrinal position, and in so doing refer to similar articulations in a variety of scriptural sources: we might term this an identification of parallelism *avant la lettre*. This type of citation-cum-parallel identification is important, and should be systematically exploited by collecting examples. A potentially even more interesting approach, which is certain to be more comprehensive and therefore more revealing, is what Tangherlini (a member of our Advisory Board) and Leonard (2013) call "sub-corpus topic modeling" (STM), based on Latent Dirichlet Allocation (LDA). As they describe it, (2013, 728), "As opposed to keyword search, which requires that the researcher know what to look for *a priori*, the topic modeling approach asks the algorithm to reveal latent semantic patterns in the data, and it couples these latent patterns with expert-applied labels." The approach seeks, in the metaphor of the authors, to fish through digitized corpora for topics or themes that may have otherwise remained undetected. A similar sort of detection can be, and has been, done with Buddhist literature by hand (e.g. Silk 2008), but this requires the scholar's prior awareness of a theme; the great advantage of STM is that it allows the text corpus *itself* to generate patterns and themes. Based on our mark-up of the MRK, we expect this to reveal **patterns and motifs which have hitherto gone unnoticed**, and therefore to inform our picture of the overall nature of the literature in unexpected and revealing ways.

Another important task of NLP is Named Entity Recognition (NER), which highlights personal, place and other names. We can expect very accurate results from the data we have tagged, which in turn will serve as a reliable source for future semi-supervised automated application to a wider corpus of Chinese Buddhist scripture translations. The relative uniformity of the vocabulary in Buddhist *sūtra*s, and the very limited number of personal and place names across this corpus—in comparison to the vast variety in Classical Chinese as a whole—will make what is otherwise a daunting task manageable and feasible. (In addition to the Taishō index for technical terms, for personal names we have a virtually complete listing in Akanuma 1931). Our problem is therefore a vastly simplified one compared to that presented by Classical Chinese as a whole, or even by the Buddhist Gazetteers studied by Bingenheimer (2015), due to the relative homogeneity of the *sūtra* corpus in terms of vocabulary and grammar, and the fact that our data is already punctuated, since this greatly increases the accuracy not only of alignment but also of entity recognition.

Editions

The corpora we will align on a sentence level in Chinese and Tibetan will make use of the above-mentioned existing data sets (Taishō, Derge). Given the quality of the data, we expect very high accuracy, but the results will be provisional pending the preparation of full editions of all component scriptures. The critical editions we will produce will employ all available primary sources: for Chinese, blockprinted (and stone rubbing) editions, and manuscripts, particularly those preserved in Japanese libraries, and when available, at Dunhuang; for Tibetan, approximately 20 separate editions, both blockprinted and manuscript, are available. The results of these editions will be as complete an accounting of the textual record of each text as is possible. Under the direction of the PI, the team will cooperate in the preparation of the overall corpus of the MRK through alignment and mark-up, the whole team will study the *Kāśyapaparivarta*, and individual team members will study selected texts and topics, and prepare critical editions of individual *sūtra*s, always with an eye on integration with the whole. The PI will author a comprehensive study of the MRK collection, considering its history (including its political situation and connection with the Empress Wu), organizational principles, influence and reception, and so on, to result in a monograph.

Of the 49 texts of the MRK, 17 have been previously treated in published or unpublished work. Studies of 7 are ongoing, and we have contacted, or will contact, the scholars involved in these efforts, inviting them to offer their results to our environment, under curated conditions. Our project to construct the software environments will begin with the pilot project based around the *Gaṅgottaraparipṛcchā* (MRK 31) and the *Maitreyaparipṛcchā* (MRK 42). In toto our team will produce editions and studies of 11 MRK texts. As a result, through the efforts of our team and others, substantially more than half of the MRK will be made available in some form of modern edition.

Providing a range of both thematic and text-critical situations, each individual project plays an integral role in the overall study of the MRK, and of Mahāyāna Buddhist scripture more broadly, and allows an appropriate level of challenge for the relevant team member, with more difficult tasks and longer texts reserved for the Post-docs, and less complex and shorter ones for the PhD students.

- the **PI** will make an **in-depth study** of a medium-length ***sūtra* on dreams**, the ***Svapnanirdeśa*** (MRK 4), extant in Tibetan and Chinese. It presents 108 dreams (the number is traditional in Buddhism), each accompanied by an elaborate allegorical explanation. The presence of similar imagery elsewhere will be

---

out that most reuse studies apply only within a single language. Studies of Buddhist literature face all of these are problems as well [cf. Metzler et al. 2005]. Syntactic *n-gram* analysis may be a better—less strict—approach than literal matching, but this requires further investigation.) Existing efforts to explore Buddhist literature do not address the problem of nonliteral parallelism, since when they search for similarities they assume an extensive use of common vocabulary. Examples include the TACL program for Chinese (https://github.com/ajenhl/tacl/; Radich 2014, 208), and the work of Klein et al. 2014 on "Inexact Quotations" in Tibetan corpora. Radich has promised his close cooperation with our project. Such approaches can, of course, produce impressive results, such as those attained by Nattier (2008), who mapped early Chinese Buddhist translation teams through patterns of word usage. However, the types of parallelism we seek to identify are not generally amenable to this approach.

explored to triangulate the intertextuality of the *sūtra*. The translation will be accompanied by a study of the rhetoric of dreams in Buddhist literature, in which the image, central to Buddhist thinking about reality, is used as a device for prophecy, an image of false mentality, of transience and ephemerality.

- **Post-doc 1** (years 1-4) will focus on an early Mahāyāna *sūtra* dealing with the ethics of the ideal Buddhist aspirant, the *bodhisattva*, and containing important references to the **previous-life stories** of the Buddha (*jātaka*). The ***Raṣṭrapālaparipṛcchā*** (MRK 18) is of medium length, extant in Sanskrit, Tibetan and 3 Chinese versions. Although one of the first Mahāyāna *sūtra*s to be published in Sanskrit (Finot 1901), and studied several times (Boucher 2008), no edition taking account of all extant sources has ever been prepared. The textual situation presented by this text is precious, and its contents a valuable window into the early period of the Mahāyāna movement in India, and the effort to redefine the place and meaning of ascetic practice in Buddhism.
- **Post-doc 2** (years 2-5) will consider the interaction between multiple sources of a text and a commentary which cites the text. The previously unstudied ***Ratnacūḍaparipṛcchā*** (MRK 47), also of medium length, is extant in Tibetan and 2 Chinese versions, with its **commentary** attributed to the great Indian scholastic Vasubandhu. Comparatively few Mahāyāna *sūtra*s have commentaries; in the group of those we will study in this project, only the *Kāśyapaparivarta* and *Ratnacūḍaparipṛcchā* are so equipped. The relationship between a 'root text' and its commentary will be carefully explored, as will the question why so few other Mahāyāna *sūtra*s are accompanied by comparable commentaries.
- **PhD 1** (years 1-4) will concentrate on issues of **intertextuality, modular composition and text reuse** and, in terms of doctrine, the **spiritual status of women** and its narrative depiction, through examination of 3 short but related texts, the ***Sumatidārikāparipṛcchā*** (MRK 30; Tibetan and 4 Chinese), ***Aśokadattavyākaraṇa*** (MRK 32; Tibetan and 2 Chinese) and ***Vimaladattāparipṛcchā*** (MRK 33; Tibetan and 3 Chinese; on the three cf. Silk 2014b). These texts have a great deal in common, including some of their literal phraseology, and thus provide an opportunity to investigate possible commonalities in their composition.
- **PhD 2** (years 2-5) will examine claims to the **universality of the Buddha's teaching**, something asserted by the ***Adhyāśayasaṁcodanasūtra*** (MRK 25) in its famous and doctrinally influential phrase that "Everything well said is the speech of the Buddha." The text exists in Tibetan and 2 Chinese translations, and significant Sanskrit quotations. The latter provide an excellent opportunity to study another form of **text reuse**, namely that of **citation**.
- **PhD 3** (years 2-5) will address a topic of profound philosophical importance and centrality in the Mahāyāna tradition, the abstruse universal principle joining all things, the ***dharmadhātu*** (cf. Silk 2015a). The vehicle for this study is the ***Dharmadhātuprakṛtyasambhedanirdeśa*** (MRK 8), for which we have 1 Tibetan and 1 Chinese version. Its study gives ample range to engage both with philology and the more abstract reaches of Buddhist philosophy.

Individual *sūtra*s are Works, but they exist in witnesses. These individual witnesses must be prepared, transcribed, arranged, and ultimately annotated and translated. The **editing environment** to be implemented by the project will allow each element of evidence to be combined with other elements such that a user can decide, for example, whether she is more interested in the purely Indic shape of a text and its vicissitudes, in its evolution from India to China or Tibet, in its relations to other texts—or in a host of questions which will occur to a user but perhaps not to those who build the system. Its structure must therefore be flexible enough to accommodate queries we have not yet imagined. The environment must have two linked elements: an editing environment and the capacity to produce printed output.

Our envisioned environment will allow a reader great latitude, including the ability to:[9]

- navigate easily among versions, to annotation, to translation, to parallel passage
- read any witness in its entirety, from beginning to end
- read an editor's reconstruction of a particular version of the text
- examine the genesis and relations of the work, as hypothesized by the editor, by means of a dynamic illustration of textual layers
- shift from any point in one witness or version to the corresponding location in any other version
- display any witness, version, or editorial reconstruction chosen as the frame text, and have an apparatus in traditional form, which scrolls along in sequence with the frame text (or as pop-up or by-hover)
- have two or more witnesses on screen, scrolling in parallel, with apparatus
- make the apparatus selective by witness (showing only selected manuscripts) or by type of variation (e.g. suppressing, or highlighting, all purely orthographic variants)
- show an entire tradition in line-by-line synoptic form, or matching selected witnesses
- display annotation (or markers indicating the availability of notes) selectively by type
- search intelligently (morphologically, fuzzily)
- offer for mediation comments, additions, and questions on any of the aspects of the presentation[10]

9. This listing is inspired in part by Sperberg-McQueen 2009, 32-33.

In order to provide for **interaccessibility among versions in different languages**, **different witnesses, translations and notes**, the texts will, in the course of being edited, be numerated, in a manner similar to that used for chapter and verse in Bibles (for an example of the imposition of such a system on a Buddhist *sūtra*, see Silk 2015a). In this manner, **we will establish a unique notation system independent of source** for each text. This system can only be established after careful study of each text, and thus will not form part of the initially aligned corpus. The system is dynamic, and without any implication of hierarchy or priority. Moreover, 'pluses' and 'minuses' (as in Biblical studies [Tov 1997, 123-132]) are easily accommodated, such that inclusion and exclusion of materials in an already standardly numbered corpus are indicated in a non-normative fashion.[11]

From an interface perspective, an excellent model is provided by a commercial and proprietary software package prepared for the study of the Bible in Hebrew, Greek, and other versions, Accordance (http://www.accordancebible.com).[12] However, it is to be noted that Accordance software works so very well in part because generations of scholars performed manual analysis of the texts of the Hebrew, Greek and Latin Bible. Therefore, there was no need to employ automatic part-of-speech taggers on the corpus, no need for lemmatization, and so on. No comparable material has ever been prepared for any Buddhist scripture. It might be possible to build our own editing environment on the Open Source software developed by the project "Buddhistische Handschriften aus Gandhāra," funded by the Bayerischen Akademie der Wissenschaften, namely READ (Research Environment for Ancient Documents), but although we have had extensive fruitful discussion with the project leader S. Baums (Munich), as this is not yet released no tests have been possible. This environment allows input/output with Open Source EpiDoc (https://sourceforge.net/p/epidoc/wiki/Home/), which is another possibility for a base environment. Both of these options, however, were designed in the first place for use with monolingual inscriptions, and thus some modification would be required. Other open and non-commercial environments which, although different in some respects, are attempting similar presentations include the Beckett Digital Manuscript Project, which proclaims, in line with what we wish to accomplish, that it "functions as a research environment that is non-hierarchical in the sense that each text can be compared to any other text and that no text is singled out as being more important or 'definitive' than the other versions. In the underlying markup, this system of relative calibration is based on a numbering system keyed to a so-called 'base text' [...] a 'text chosen by an editor to compare with other texts of the same work in order to record textual variation among them'."

While the full flexibility of the electronic architecture will only be possible in a virtual environment, we will **enable and encourage the production of printed output**. We are in full agreement with Dahlström (n.d.), who opined, "A S[cholarly, critical] E[dition] is intended to fulfill two perhaps contradictory user demands: a) the clear, economical, selective guiding through the textual mass in such a way that the user can benefit from the editor's insights and competent judgement, and b) the broadest possible presentation of the textual material, enabling the user to choose different paths and variants than has the editor. [...] There is no reason why a digital archive could not result in a frozen print edition as an out product. The important difference is that a print edition from such a digital archive is *one potential* bi-product, not *the final* end product. The complexity of such envisioned archives allows one to rather imagine many possible edition types, be it reading, student, diplomatic, variorum, modernised, genealogical, multiple, or critical ones." Since most readers (as opposed to users, in the sense of those who wish to delve into details) will wish to *read* a published text, we will produce such editions, both digitally through Open Access and in inexpensive printed form; an example is Silk (2015a), published as a book while simultaneously available for legal, free PDF download.

Support Structure and Embedding in the Field

As a member of the Advisory Board of the **Buddhist Digital Resource Center**'s (BDRC) Buddhist Text Preservation and Digital Archive, the PI is intimately familiar with the current situation of other state-of-the-art efforts at digitalization of Buddhist scriptural sources. Our project has secured both moral and practical technical support from the major players in this field, a number of whom will serve as members of our Advisory Board. Digitization of some earlier Buddhist scriptures in Chinese translations, in Pāli and in Sanskrit fragments, was carried out (see http://suttacentral.net), also in parallel with Chinese sources (http://buddhist-informatics.ddbc.edu.tw/BZA), in an archive built principally by M. Bingenheimer (Temple Univ.), J. Hung (Dharma Drum Institute of Liberal Arts) and Chao-Lin Liu (National Chengchi University), all of whom have agreed to advise and collaborate with our project. Dharma Drum is also the creator of the **CBETA** (Chinese Buddhist Electronic Text Association) archive of the Chinese Tripiṭaka, one of the major digital repositories of

10. While not a priority, a possible future functionality is linking witnesses to their photographic reproductions, as implemented for instance in the Beckett Digital Manuscript Project (http://www.beckettarchive.org/editorial.jsp) and the READ project of the Buddhistische Handschriften aus Gandhāra project in Munich. At the moment, while some *sūtra* materials are available (e.g. through the BDRC), for others reproduction is not realistic for copyright or legal reasons.

11. Until each text is provided an enumeration, reference can continue to be made to the standard Taishō and Derge editions. In some cases in which editions have been published, their numbering may be adopted, if adequate, but it would be preferable to establish a standardized system for the entire corpus.

12. A simple but important feature of our interface will be that it is multilingual. The interface will be built not only in English but initially also in Japanese and Chinese. It is core to the project's stance that its results belong also to Buddhist communities in Asia and the West, and not merely to a narrow cadre of scholars.

the Chinese Buddhist canon. Likewise we have secured close cooperation for the construction of our tools and overall conceptualization from the Center for Evolving Humanities at the University of Tokyo (M. Shimoda, director) and the **SAT** Chinese Tripiṭaka, a project of the International Institute for Digital Humanities, (K. Nagasaki, Information Architect) which, as mentioned above, has carried out our pilot project to digitize the index of the Taishō MRK. Another very close partner is the likewise above-mentioned Buddhist Digital Resource Center, **BDRC** (formerly the Tibetan Buddhist Resource Center, TBRC) and its major project, the **Buddhist Universal Digital Archive, BUDA** (J. Wallman, Executive Director). Further cooperation has been agreed with another important project of multilingual Buddhist text input and presentation, the **Thesaurus Literaturae Buddhicae** of The Norwegian Institute of Palaeography and Historical Philology (J. Braarvig, Director). Finally, the Homer Multitext Project (http://www.homermultitext.org/index.html) and its Information Architects, C. Blackwell and N. Smith, whose work, in an entirely different cultural domain, presents important conceptual similarities to ours, have offered their generous support, and Blackwell will sit on our Advisory Board.

Open Data allows sharing and adaptation, but openness alone leads to chaos. All areas of our archive and environment will be available to users, and we will encourage adoption of the environment and use of the data, but contributions must be mediated by qualified scholars. During the 5-year lifetime of the project, team members will be responsible for **mediation and peer-review**. Thereafter, in conjunction with the BUDA project we will establish an editorial board to oversee such contributions. Therefore, the openness of the project is not a wiki-like free-for-all but a carefully controlled system which encourages qualified input. In this respect, we aspire to aims similar to those articulated by the project at http://papyri.info, whose Papyrological Editor "enables multi-author, version controlled, peer reviewed scholarly curation of papyrological texts, translations, commentary, scholarly metadata, institutional catalog records, bibliography, and images."

The Team and the Leiden Context

All team members, in addition to relevant philological skills and background, will be recruited with an eye on skills in Digital Humanities. The Leiden University Faculty of Humanities has newly appointed two Assistant Professors in Digital Humanities, P. Vierthaler (http://pvierth.herokuapp.com/) and J. Cha (https://javiercha.com/), both of whom work in Chinese and who will work closely with the team to improve our skills. This **significant expertise in the emerging field of Chinese Digital Humanities in the faculty in which OPEN PHILOLOGY will be housed** assures us direct and high quality advice and guidance close at hand. The PI will work 50% in the project for its full 5 years. His Institute (Leiden University Institute for Area Studies, LIAS) has, moreover, formally agreed to limit his teaching to one course per term for the duration. The Post-docs will be appointed at 70% for 4 years, with an additional 30% paid by the LIAS, to allow them to teach, since in the Humanities we consider teaching essential for professional development, and as preparation for any future academic position. The PhDs will be appointed at 100% for four years. We will recruit in open competition world-wide; the PI is aware of several fine candidates. A student assistant will work one day a week (0.2fte) to assist with clerical matters, scheduling, meeting and conference planning, and so forth.

Programmers

We will sub-contract the services of two **specialists in programming and software design**, one in **database creation**, the other in **interface development**. Their positions will involve concentrated time in Phase I of the project, less in Phase II, and relatively little, primarily devoted to debugging and maintenance, in Phase III. For this we estimate total time at 3500 hours. To ensure compliance with the best value for money principle, we will prepare an estimate of the costs for this type of work and a full specification of the work to be sub-contracted. We will launch a call for expressions of interest to all firms and individuals specialized in database creation and interface development located within reasonable distance of Leiden (in principle, the EU and UK), in view of the need for regular face-to-face consultations. We will select the best candidates according to: 1) price; 2) time requested; 3) years of experience; 4) familiarity with similar academic work. We do not anticipate any commercial exploitation of the results. Provisionally, based on the market at present, we budget on a scale of €125/hr (+VAT).

Advisory Board

In addition to an initial meeting in Leiden, one visit during the project and another during our International Conference in year 4, members of the international, interdisciplinary Advisory Board will meet 2~3 times per year in a virtual, online setting, to assess and evaluate the project, and advise on all relevant aspects, in particular those concerning Digital Humanities. They will provide feedback on effectiveness of tools and technologies and the user interface of the project, via email, phone calls, online and personal meetings. The Advisory Board, whose members have been mentioned in the body of this proposal, consists of:

M. Bingenheimer: Temple University
J. Braarvig: University of Oslo
J. Hung: Dharma Drum Institute of Liberal Arts
K. Nagasaki: International Institute for Digital Humanities
T. Tangherlini: UCLA
C. Blackwell: Homer Multitext Project
N. Hill: SOAS
C.-L. Liu: National Chengchi University
M. Shimoda: Center for Evolving Humanities
H. Tauscher: University of Vienna

J. Wallman: BDRC

Materials
Aside from the digital collections which will serve as the bases for corpus alignment, for our editions primary sources are the Chinese and Tibetan canonical texts. In the case of the former, we have direct access to 4 woodblock printed canons, to old manuscripts from the cave-temple of Dunhuang on the Silk Road, and from Japanese monasteries (Nanatsudera and others). Some of these are published, but some require purchase. For Tibetan sources, we have 9 woodblock printed canons, and in recent years increasing access to manuscript canons. As many as 20 witnesses are available. Especially through the above-mentioned Vienna Kanjur project, we have access to much of this Tibetan material at minimal cost. Additional sources, (e.g. the recently published Urga Kanjur and a newly published, much better, digital color print of the Peking Kanjur) can be commercially purchased. In addition, a number of scholarly publications, especially from China and Japan, are necessary. Leiden University Libraries does not have the financial resources to purchase all of this material, so we must acquire it within the project.

Phases of the Project, and Timeline:
In **Phase I** (year 1, continuing in year 2), we will plan for and begin construction of the online platform and interface, and develop or adapt corpus architecture, data curation, and tools (tokenizer, part-of-speech tagger, lemmatizer), and begin to prepare documentation. The first test will be input of the already prepared editions of the *Gaṅgottaraparipṛcchā* and the *Maitreyaparipṛcchā*. These steps will be followed by evaluation and adjustment in light of experiences and lessons learnt. We will hold a joint working session with our Advisory Board in Leiden. At the 6 month mark we will know in detail how our software environments are functioning, and will have begun to make the modifications necessary for our particular needs. Team members will have been trained in Digital Humanities methodologies and relevant programming languages. Simultaneously, in Phase I we will begin weekly team study of the *Kāśyapaparivarta*, alternating with discussions on philological theory and related issues. Team members will begin the preparation of data for alignment using standard corpora (Taishō/Derge), to be followed by preparation of critical editions, and will begin study of topics related to their individual texts. After one year, having completed the pilot project, we will schedule a second meeting with our Advisory Board to profit from their evaluation and further advice. For those team members who join in year 2, in the first few months we will introduce them to the results of year 1 discussions and experiments, and begin to train them in methodologies and relevant programming languages.

While some aspects of Phase I continue in year 2, including continuing evaluation of technologies and methodologies by testing and applying tools developed in Phase I, in **Phase II** we will develop additional tools, as needed, and refine data curation and metadata standards, and the alignment and editing environments. We will implement the marked-up Chinese text of the MRK section of the Taishō prepared by our partners in Tokyo, and roll out the alignment environment to the MRK as a whole. We will continue to refine both, and begin inputting transcriptions of witnesses for the texts we will study in detail (in Unicode, not further marked-up; only pages and line-breaks indicated). Team members will study individual texts, while paying special attention to seeking parallelisms and pericopes shared across the corpus. This effort will continue through years 3 and 4. In year 3 we will begin planning for the International Conference we will hold in year 4, which we will coordinate with a third joint working group with the Advisory Board. The conference and its subsequent proceedings will build on our contact with others around the world engaged in related work. We will invite not only specialists in Buddhist literature but also those working in conceptually related projects in other areas, such as Homer and Rabbinic literature.

Phase III belongs to the conclusion of our project. It involves the presentation of our results and the complete release of a new version of the corpus (see below under Data Management Plan). We will write and disseminate all necessary documentation, and in addition prepare for the continuation and expansion of the methods and tools we developed and refined in Phases I and II. By this time, our continued frequent communication with our partners, the publicity attending our International Conference and the subsequent publication of its papers, and the availability of our data and tools will have prepared the way for others to continue our approach, expanding its application to the remainder of the Buddhist scriptural corpus, and beyond. We will hold our final working group meeting with our main partners to discuss continuation issues. During this phase, we will make special efforts to travel abroad to publicize our accomplishments, especially in Asia.

The project will have several kinds of **Output**. Conventionally, we will publish a study of the MRK as a whole, and studies, editions and annotated translations of our individual texts. These publications will include in particular a consideration of the modular intertextuality of the MRK as a unit. The electronic corpus of the MRK, aligned between all extant versions, will be completed. Further, papers on both the content-wise topics raised by the study of the texts, and on method and practical matters raised by the digitization and alignment efforts, will be both presented at conferences (including our own, and with a special effort to present in Asia), and published in both Buddhist Studies and Digital Humanities venues. A conference volume which will synthesize our efforts, and document our discussions with other scholars from around the world working on related efforts, will result from our Leiden meeting in year 4.

As an essential part of their scholarly training, and in view of the formal academic regulations of Leiden University, the project's PhD students will produce **dissertations in a traditional format.** The theses, as well as the studies of the PI and Post-docs, who will also produce **books**, will contain not only editions but substantial studies of the texts under examination, along with considerations of the methods explored in the project. They will be **accompanied by electronic editions in the new architecture.**[13] Other academic contributions, including **articles in peer-reviewed journals**, both in Buddhist Studies and Digital Humanities, will deal with the doctrines and other content of the texts, text-critical issues and with methodology. The senior team members will more actively contribute to the building of tools and participate in conferences and other activities to share our experiences.

Person Year	Team	PI	P-D 1	P-D 2	PhD 1	PhD 2	PhD 3
1 (Phase I)	collate MRK: Chinese/Tibetan; build editing environment; *Gaṅgottara* & *Maitreya* pilot project; meet with Advisory Board	begin general study of MRK; begin *sūtra* project	begin collation & study		begin collation & study; improve DH skills		
2 (Phase I/II)	continue collations & refine editorial environment; meet with Advisory Board	continue, and attend conference	continue, and attend conference	begin collation & study	continue	begin collation & study; improve DH skills	begin collation & study; improve DH skills
3	publish MRK collation, and elicit additions from scholars; publish editing environment for feedback	continue, and attend conference on DH; organize our conference	continue; attend conference on DH; organize our conference; publish book;	continue, and attend conference on DH; organize our conference	continue, and attend conference	continue	continue
4	project conference and meeting with Advisory Board; make public editing environment	lead our conference; edit conference volume; continue research	complete research; participate in our conference	continue; participate in our conference	complete dissertation; participate in our conference	continue, and participate in our conference	continue, and participate in our conference
5 (Phase III)	integrate editing environment into BDRC site; meet with partners for further exploitation of results	complete research; publish conference volume; publish MRK and *sūtra* books	(attend final meeting as invitee)	complete research; attend conference, and publish book	(attend final meeting as invitee)	complete dissertation	complete dissertation

Data Management Plan

Server space will be provided by Leiden University, and other work will be carried out on existing desk-top or laptop machines, or off-the-rack machines purchased for team members. No special hardware is required.

Data produced by the project will be jointly managed by the PI and the BDRC (Jeff Wallman). (For general issues of data curation, see https://guide.dhcuration.org/.) After completion of the project, in collaboration with the BDRC the PI will head a panel of scholars to curate the data produced by the project, integrating this with any data subsequently produced with our tools. All data, including that obtained from the SAT project and BDRC, will remain Open Source. The data produced by project members will consist of collated primary sources, marked-up text (in principle in Unicode and XML), and software code produced in the project itself. We will work with the BDRC to produce a sustainable accessible environment of aligned corpora, and the open research environment, including application programing interfaces (API), Open Data feeds, and backed-up storage of all digital objects created in the project. All data will be stored on servers of Leiden University, the BDRC, and the International Institute for Digital Humanities (Tokyo). In addition to daily local backup, we will update and back-up the files with these partners on a monthly basis. Data will be open as soon as

13. The team will also produce tutorials and other aids to assist users in making the best use of the tools and data we will produce. These will take the form of web tutorials, videos or other guidance as appropriate.

produced. No confidential data will be produced, and no embargoes imposed. No commercial application of any data produced is foreseen. Following domain-local standards and widespread usage, we will deploy recognized standards, including library cataloging standards, for our metadata.

Digitized Text: The digitized text will be raw data in the form of text and XML files. These will be stored on GitHub, making use of this platform both as a version control system and depository. Under intellectual property law in effect, the text from the blockprints and manuscripts we will use is in the public domain. Our digital textual data is based on transcriptions of ancient texts, which are no longer under copyright restrictions. We will not reproduce images of the objects themselves without permission from their repositories. We will use published materials (e.g., existing editions) only if the copyright has expired, if permission has been granted, or if we have also consulted with the original manuscripts to produce our own original editorial work. We expect no legal or ethical restrictions on our data.

Tools and Technologies: The digital tools to annotate and format the text files will be written in Java, Python, or other scripting languages. We will pursue the adaptation of existing Open Source tools and the development of our own tools. We will distribute the tools via links on our web platform and through the BDRC as free public downloads under Open Source licenses. The software will be distributed on GitHub. The files will be created using digital tools and manual annotations, which can produce text, .csv, Excel, and XML files. The final output will be produced by the Open Source converter framework SaltNPepper, which enables easy standoff markup of tokenized text. (For this and other relevant software see http://corpus-tools.org/home/.)

Literature:

Akanuma Chizen. *Indo Bukkyō koyū meishi jiten* [Dictionary of Indian Buddhist proper nouns]. Kyoto, 1931.
Bamman, D. et al. 2010. “Transferring structural markup across translations using multilingual alignment and projection.” In *Proceedings of the 10th annual joint conference on Digital libraries* (JCDL ’10). New York, 11-20. http://doi.acm.org/10.1145/1816123.1816126
Bamman, D., and G. Crane. 2009. “Discovering Multilingual Text Reuse in Literary Texts.” In *White Paper, Perseus Digital Library* (http://www.perseus.tufts.edu/publications/2009-Bamman.pdf).
Bingenheimer, M. 2015. “The Digital Archive of Buddhist Temple Gazetteers and Named Entity Recognition (NER) in classical Chinese.” *Lingua Sinica* 1/8: 1–19.
Blackwell, C., and N. Smith. 2014a. The CTS Protocol Specification (version 5.0). Center for Hellenic Studies Technical Documentation. http://www.homermultitext.org/hmt-docs/specifications/cts/.
Blackwell, C., and N. Smith. 2014ab. The CTS URN Specification (version 5.0). Center for Hellenic Studies Technical Documentation, 2014. http://www.homermultitext.org/hmt-docs/specifications/ctsurn/.
Blackwell, C., and N. Smith. 2012. “Four URLs, Limitless Apps: Separation of Concerns in the Homer Multitext Architecture.” In L. Muellner, ed. *Donum Natalicium Digitaliter Confectum Gregorio Nagy Septuagenario a Discipulis Collegis Familiaribus Oblatum*. Washington, DC. http://folio.furman.edu/projects/cite/four_urls.html.
Börner, K. 2011. “Plug-and-Play Macroscopes.” *Communications of the ACM* 54/3: 60-69.
Boucher, D. 2008. *Bodhisattvas of the Forest and the Formation of the Mahāyāna*. Honolulu.
Bowers, F. 1959. *Textual and Literary Criticism*. Cambridge.
Büchler, M., A. Geßner, M. Berti, T. Eckart, 2013. “Measuring The Influence Of A Work By Text Re-Use.” *The Digital Classicist 2013* (*Bulletin of the Institute of Classical Studies* Supplement 122), 63-79.
Burnouf, E. 1844. *Introduction à l’Histoire du Buddhisme Indien.* Paris.
Dahlström, M. n.d. “Drowning by Versions.” http://etjanst.hb.se/bhs/ith/4-00/md.htm.
Daizōkyō Gakujutsu Yōgo Kenkyūkai, ed. 1926–1985. *Taishō Shinshū Daizōkyō Sakuin* [Indexes to the Taishō canon]. Tokyo.
Davila, J. R. 1994. "Prolegomena to a Critical Edition of the Hekhalot Rabbati." *Journal of Jewish Studies* 45: 208-226.
Dekker, R. H. and G. Middell. 2011. “Computer-Supported Collation with CollateX: Managing Textual Variance in an Environment with Varying Requirements”, Supporting Digital Humanities, University of Copenhagen. <http://crdo.up.univ-aix.fr/SLDRdata/doc/show/copenhagen/SDH-2011/submissions/sdh2011_submission_54.pdf>
Deng,K., P.K. Bol, K. J. Li, and J. S. Liu. 2016. “On the unsupervised analysis of domain-specific Chinese texts.” *Proceedings of the National Academy of Sciences* 113/22: 6154–6159
Dué, C. and M. Ebbott. 2009. “Criticism: Editorial Standards for the Homer Multitext” *Digital Humanities Quarterly* 3/1: (http://www.digitalhumanities.org/dhq/vol/3/1/000029/000029.html).
Finot, L. 1901. *Rāṣṭrapālaparipṛcchā*. (St. Pétersbourg: Imperial Academy).
Forte, A. 2002. “The South Indian Monk Bodhiruci (d. 727). Biographical Evidence.” In A. Forte and F. Masini, eds., *A Life Journey to the East. Sinological Studies in Memory of Giuliano Bertuccioli* (*1923–2001*). Kyoto, 77–116.
Garrett, E., N.W. Hill, and A. Zadoks. 2014. ‘A Rule-based Part-of-speech Tagger for Classical Tibetan.’ *Himalayan Linguistics*, 13/1, 9-57.
Grigely, J. 1995. *Textualterity*: *Art, Theory , and Textual Criticism*. Ann Arbor.
Honko, L. 2000. ‘Text as process and practice: the textualization of oral epics.’ In L. Honko, ed. *Textualization of Oral Epics*. Berlin, 3–54.
van Hulle, D. 2004. “Compositional variants in modern manuscripts.” In A. Bozzi, L. Cignoni and J.-L. Lebrave, eds., *Digital Technology and Philological Disciplines.* Pisa, 513-527.
IFLA Study Group on the Functional Requirements of Bibliographic Records. 1998. "Functional Requirements of Bibliographic Records: final report." München. http://www.ifla.org/VII/s13/frbr/frbr.pdf. (Downloaded 29 August 2002.)
Klein, B. et al. 2014, “Finding Inexact Quotations Within a Tibetan Buddhist Corpus (Poster)”, Digital Humanities

(DH) 2014, Lausanne, Switzerland, pp. 486-488. <http://www.cs.tau.ac.il/%7Enachumd/papers/textalignment.pdf>
Le, Q. V., and T. Mikolov. 2014. "Distributed Representations of Sentences and Documents." *Proceedings of the* 31st *International Conference on Machine Learning*, Beijing, China, 2014. JMLR: W&CP volume 32. http://arxiv.org/abs/1405.4053.
Lord, A. B. 1960. *The Singer of Tales*. Cambridge.
Metzler, D., et al. 2005. "Similarity measures for tracking information flow." In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*. New York, 517–524.
Milikowsky, C. 2002. "On the Formation and Transmission of Bereshit Rabba and the Yerushalmi: Questions of Redaction, Text-Criticism and Literary Relationships." *Jewish Quarterly Review* 92/3-4, 521-567.
Milikowsky, C. 2006. "Reflections on the Practice of Textual Criticism in the Study of Midrash Aggada. The legitimacy, the indispensability and the feasibility of recovering and presenting the (most) original text." C. Bakhos, ed., *Current Trends in the Study of Midrash*. Leiden, 79-110.
Mori Shōji. 1988. *Bukkyō Hiyu Reiwa Jiten* [Metaphors in Buddhist Texts]. Tokyo.
Nagy, G. 2000. Review of M. L. West, *Homeri Ilias. Bryn Mawr Classical Review* 2000.09.12: http://bmcr.brynmawr.edu/2000/2000-09-12.html.
Nagy, G. 2011. "Homeric Poetry and Problems of Multiformity: the 'Panathenaic Bottleneck.'" *Classical Philology* 96, 109-119.
Nath Jha, G., ed. 2010. *Sanskrit Computational Linguistics: 4th International Symposium, New Delhi, India, December 10-12, 2010. Proceedings.* Berlin.
Nattier, J. 2008. *A Guide to the Earliest Chinese Buddhist Translations*. Tokyo.
Phillips-Rodriguez, W. J. 2007. "A Discussion about Textual Eugenics: Still Searching for the Perfect *Mahābhārata*?" *Variants – The Journal of the European Society for Textual Scholarship* 6, 163–175.
Radich, M. 2014. "On the Sources, Style and Authorship of Chapters of the Synoptic *Suvarṇaprabhāsottama-sūtra* T664 Ascribed to Paramārtha (Part 1)." *Annual Report of The International Research Institute for Advanced Buddhology* 17, 207-244.
Sakurabe Bunkyō. 1930-1932. *A Comparative Analytical Catalogue of the Kanjur Division of the Tibetan Tripitaka.* Kyoto.
Schäfer, P. 1986. "Research into Rabbinic Literature: An Attempt to Define the Status Quaestionis." *Journal of Jewish Studies*, 37,139-152.
Schäfer, P. 1989. "Once Again the Status Quaestionis of Research in Rabbinic Literature: An Answer to Chaim Milikowsky." *Journal of Jewish Studies* 40/1, 89-94.
Schäfer, P. 1992 . *The Hidden and Manifest God: Some Major Themes in Early Jewish Mysticism.* Albany.
Schmidt, D. 2014. "Towards an Interoperable Digital Scholarly Edition." *Journal of the Text Encoding Initiative* 7: 1-20.
Silk, J.A. 2016. "Peering Through a Funhouse Mirror: Trying to Read Indic Texts Through Tibetan and Chinese Translations." In D. Wangchuk, ed., *Cross-Cultural Transmission of Buddhist Texts: Theories and Practices of Translation*. Hamburg: 287–311.
Silk, J. A. 2015c. "Establishing / Interpreting / Translating: Is It Just That Easy?" *The Journal of the International Association of Buddhist Studies* 36/37 (2013-2014): 205-225.
Silk, J. A. 2015b. "Canonicity." In *Brill's Encyclopedia of Buddhism*. Vol. I.
Silk, J.A. 2015a. *Buddhist Cosmic Unity: An Edition, Translation and Study of the* Anūnatvāpūrṇatvanirdeśaparivarta. Hamburg. Open Access: http://hup.sub.uni-hamburg.de/volltexte/2015/154/pdf/HamburgUP_HBS4_Silk_Unity.pdf.
Silk, J.A. 2014b. "Taking the *Vimalakīrtinirdeśa* Seriously." *Annual Report of The International Research Institute for Advanced Buddhology at Soka University* 17, 157-188.
Silk, J.A. 2014a. "Out of the East: Tibetan Scripture Translations from Chinese." *Bod rig pa'i dus deb / Zangxue xuekan | Journal of Tibetology* 9: 29-36.
Silk, J. A. 2013. "The Nature of the Verses of the *Kāśyapaparivarta*." *Bulletin of the Asia Institute* 23 (Evo ṣuyadi: Essays in Honor of Richard Salomon's 65th Birthday), 181–190.
Silk, J. A. 2011. "The Utility of Tibetan Translations for the Editing of Chinese Scripture Translation and the Question of Lineages of the *Dazangjing*: Some First Notes from the *Ratnarāśi-sūtra*." *The First International Conference of Tripitaka Studies*, Gaoxiong, Taiwan, 295-306.
Silk, J. A. 2010. "Test Sailing the Ship of the Teachings: Hesitant Notes on *Kāśyapaparivarta* §153–154." In Eli Franco and Monika Zin, eds., *From Turfan to Ajanta*. Lumbini, II.897–924.
Silk, J. A. 2009. "Remarks on the *Kāśyapaparivarta* Commentary." In M. Straube, R. Steiner, J. Soni, M. Hahn und M. Demoto, eds., *Pāsādikadānaṃ.* Marburg, 381–397.
Silk, J. A. 2008. *Riven by Lust: Incest and Schism in Indian Buddhist Legend and Historiography.* Honolulu.
Silk, J. A. 2002. "What, If Anything, is Mahāyāna Buddhism? Problems of Definitions and Classifications." *Numen* 49/4, 355–405.
Silk, J. A. 1994a. *The Heart Sutra in Tibetan: A Critical Edition of the Two Recensions Contained in the Kanjur*. Vienna.
Silk, J. A. 1994b. "The Origins and Early History of the *Mahāratnakūṭa* Tradition of Mahāyāna Buddhism, With a Study of the *Ratnarāśisūtra* and Related Materials." Ann Arbor.
Silk, J. A., and C. Li. Forthcoming. Two Chinese *Sūtra*s in Tibetan Translation.
Sperberg-McQueen, C. M. 2009. "How to teach your edition how to swim." *Literary and Linguistic Computing* 24/1, 27-39.
Tangherlini, T. and P. Leonard. 2013. "Trawling in the Sea of the Great Unread: Sub-corpus topic modeling and Humanities research." *Poetics* 41/6, 725-749.
Tov, E. 1997. *The Text-Critical Use of the Septuagint in Biblical Research*. 2nd ed. Jerusalem.
Trovato, P. 2014. *Everything You Always Wanted to Know about Lachmann's Method: A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Cladistics, and Copy-Text*. Padova.

Vanhoutte, E. 2007. "Traditional Editorial Standards and the Digital Edition." *Learned Love. Proceedings of the Emblem Project, Utrecht Conference on Dutch Love Emblems and the Internet* (November 2006), ed. E. Stronks and P. Boot. The Hague, 157-174. http://emblems.let.uu.nl/static/images/project/learned_love_157-174.pdf.
Waku Hakuryū. 1979. *Bukkyō Shokubutsu Jiten* [Dictionary of Plants in Buddhism]. Tokyo.
West, M. 2001. "West on Nagy and Nardelli on West." *Bryn Mawr Classical Review* 2001.09.06. http://bmcr.brynmawr.edu/2001/2001-09-06.html.
Xu, J, and J. Chen. 2011. "How Much Can We Gain from Supervised Word Alignment?" *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers*: 165-169.