ERC Advanced Grant 2016
Research proposal [Part B1]

Open Philology: The Composition of Buddhist Scriptures

OpenPhilology

Cover Page:
- Principal Investigator (PI): Jonathan A. Silk
- PI's host institution for the project: Leiden University
- Proposal duration in months: 60

Proposal summary (identical to the abstract from the online proposal submission forms, section 1).

Buddhist *sūtra*s are fundamental sources for understanding the beliefs that once dominated, and largely continue to dominate, Asian societies. But classical philological approaches to these texts, assuming an Ur-text and linear development, misrepresent their nature and obscure their history. Like oral literature the *sūtra*s are authorless and textually fluid, their content formulaic and modular, and the situation complicated by their huge volume and the linguistic diversity of their extant versions. Their fluidity and the absence of an Ur-text makes traditional stemmatic editing inappropriate. Other options are also unacceptable: eclectic editions conflate distinct sources, creating texts which never existed historically, while exclusive focus on a single version ignores the tradition's true richness.

Utilizing the tools of Digital Humanities, OPEN PHILOLOGY offers a revolution in approaching the composition of these scriptures, and other literatures with common features, such as the Homeric corpus or Rabbinic literature. Given the vastness of the Buddhist canons, we will focus on a traditional subset of texts, and produce electronic editions which preserve their diachronic and synchronic fluidity, revealing the intertextuality inherent in their formulaic composition. Our editing environments will allow one to easily view the diversity of sources, in different versions of 'the same text' and in parallels in other texts, and permit one to view data at any level of granularity, according to one's interests.

The project will produce corpora of *sūtra*s mutually aligned in their Chinese, Tibetan and, when available, Sanskrit versions, a map of formulaic content, annotated text editions and translations, a general study of the corpus, and a series of scholarly publications on methodological and content-oriented issues. The tools and results we produce will bring our historical understanding of Buddhism, the most formative influence on Asian cultures, to a new level.

Section a: <u>*Extended Synopsis of the scientific proposal (max. 5 pages)*</u>

Buddhist scriptures are the most highly esteemed vessels of the tradition's ideas and ideals, and represent our best bases for understanding the beliefs and ideologies that once dominated, and largely continue to dominate, Asian societies, and that are now spreading vibrantly in the West as well. They are our **prime evidence for Buddhist worldviews and self-understandings**, and thus for approaching some of the most basic ideas and ideals of half the human race. But these texts, the *sūtra*s traditionally considered to record the Buddha's sermons, confront us with an array of challenges: like oral literature, they are authorless and their content often highly formulaic, their volume is huge, and they exist in a variety of languages. Our view of the tradition's very nature can be significantly biased not only by the texts we choose to study (in the past, often those most important to modern Japanese sectarian Buddhists), but also by the way in which the texts have been treated. The pioneering editors who brought the *sūtra*s under the gaze of scholarship approached them with a set of presuppositions inappropriate for their history, inherited from Classical and Biblical studies, ideas of a relationship among sources they thought would allow, through the tracing of paths of descent and the elimination of 'corruption,' the **reconstruction of an Ur-text**. Correspondingly, they viewed shared material through a framework of source and borrower. **Editors thus disregarded the actual lives of texts in communities widely scattered in time and space:** they overlooked both the inherent diversity of the literature and its role as witness to Buddhist life in favor of the creation of a hypothetical source which, in fact, never existed. **The technical limitations of two-dimensional print** and the theoretical model of a main text accompanied by variant readings further conspired to prevent scholars from accurately representing the manifold nature of the literature. The result has been an obscuration and misrepresentation of the history of Buddhism, its place in Asian civilizational history, and the rich medley of forms in which the tradition expressed itself over the ages. OPEN PHILOLOGY offers a **revolutionary new way of approaching these fluid texts**, taking advantage of **the flexibility of digital environments**. Critically examining a select representative portion of the vast Buddhist scriptural canon, we will develop **architectures for presenting multiple related texts in an range of languages, preserving, rather than simplifying, the complexity of the texts**, and facilitating explorations of their highly **formulaic modular composition.** The result will open a window onto Buddhist history that we had not even known existed before. This approach will be of value and utility, moreover, also to those who explore other literatures of a fluid nature, such as the Homeric, the Rabbinic, and Arabic Ḥadīth.

State of the Art
According to Buddhist tradition, in the 40 years of his teaching career the Buddha (± 5th c. BCE) wandered preaching from village to village. He gave 'the same sermon' multiple times, each time, however, a bit differently, as a bard repeating a tale, and like a bard, he employed formulae to facilitate his teaching. In an age when memory ruled, these sermons were held in mind, in whole or in part. Since each 'original' varied from every other 'original', the lessons passed on orally, even if memorized verbatim, came to constitute a circulating collection of **variant but *equally authentic* versions of the Buddha's teachings**, the *sūtra*s. Some sermons, so profound that they could not be revealed openly during the Buddha's lifetime, taught the 'Great Vehicle' (leading toward awakening, in Sanskrit, *Mahāyāna*). Setting aside the historicity of this scenario, it reflects the actual situation of *sūtra*s **themselves, constituted of multiple closely related versions, sharing substantial formulaic material** which, however, **cannot stem from a unique archetype or 'Ur-text'**. For the historian, when the Mahāyāna movement arose some centuries after the death of the Buddha (±0?), its proponents presented their new insights and understandings in the most hallowed format available, the *sūtra*. These **Mahāyāna *sūtra*s**, produced in India from the beginning of the Common Era to the 5th c., are, like the earlier works upon which they are modelled, **fluid, without an Ur-text and highly formulaic.**

As Buddhism spread from its Indian homeland throughout Asia, missionary monks carried the *sūtra*s with them, in manuscripts or in memory. Since Buddhism vanished from India by the 13th c., the scriptures are for the greater part **preserved in renderings in Chinese (from the 2nd c.) and Tibetan (from the 8th c.).** The volume of this material is huge, in Tibetan translation over 70,000 pages, of which only a very small portion is yet critically edited or translated in any modern language. For both practical and principled reasons, OPEN PHILOLOGY focuses on one representative set of Mahāyāna *sūtra*s, the ***Mahāratnakūṭa*** collection (MRK). From a practical perspective, while still sizable—some 3500 Tibetan pages—the MRK permits study by a single team. From the perspective of its content, its 49 texts represent a cross-section of the varieties of Mahāyāna *sūtra*s found in the broader canon. The MRK may even have been compiled with the intention to create a sort of '**mini-canon**,' a hypothesis which arises not only from its thematic variety but also from the circumstances of its composition.

Although its separate *sūtra*s were all composed in India, whether the MRK was considered a unit in India itself is unclear. It is certain that the form in which we have it is a result of the editorial work of the monk Bodhiruci in Tang China, presented to the court in 713 but long in preparation. In the immediately preceding period, the year 690 had seen the enthronement of the Empress Wu Zetian (r. 690–705), the only woman to rule China in her own name. I hypothesize that Bodhiruci's project was intimately connected with **Empress Wu's efforts to establish a Buddhist realm** in medieval China, or with the compensatory efforts of the Buddhist community to secure itself after her fall from power. Many questions surround **the historical situatedness of the MRK:**

What is its background? Was it intended as a 'best hits' collection from the Mahāyāna *sūtra* literature? If so, what was its aim, and why are the *sūtra*s ordered as they are—is this ordering related to their broader political purpose? This collection is consequently of great interest not only philologically as a **representative collection of sources**, but both for the light its contents will shed on the development of Mahāyāna Buddhism in India, a hotly debated topic, and—specifically in terms of its Chinese compilation—in relation to issues such as the tension between Church and State, gender and power, and the propaganda uses of religious literature. The study of individual texts in the collection presents us with a variety of editorial situations, a wide range of doctrinal stances, rhetorical modes, literary styles and religious attitudes, and with an excellent opportunity to map and begin to understand the modular and formulaic nature of the *sūtra* corpus.

Impact

Buddhist scriptures are not the only literature free of an Ur-text. In fact, it is likely the unique author who is the rarity in history, and anonymous or bardic authorship the norm. The Homeric corpus and Rabbinic texts provide helpful points of comparison (see e.g. Davila 1994, Dué and Ebbot 2009, Lord 1960, Milikowsky 2006, Nagy 2011, Schäfer 1986). The fluidity and lack of a unique, original form, and the ample use of **common pools of material** in these literatures make them natural comparators for the *sūtra*s. Texts evolve in a fluid way partly through appropriation of shared elements. Pericopes and stock phrases were merged with narratives and doctrinal elements to create new presentations which, remaining in the same key, are accepted as belonging to the same authentic dispensation. The Buddhist case differs from both the Homeric and Rabbinic, however, particularly in the enormous complexity of its corpora, and the many languages in which its texts are preserved. If study of Homer had to wait for the Homer Multitext Project (http://www.homermultitext.org/), how much more does the Buddhist scriptural corpus demand a digital treatment.

A measure of the confusion risked by a failure to appreciate the fluid textual nature of Buddhist scriptures is evident in translations even of such excellent scholars as É. Lamotte (1962) and J. Nattier (2003). In both cited cases—and many more could be adduced—a translation of one source, a Tibetan canonical text, is conflated with historically unrelated Chinese materials. The result is a translation of a source which never existed, anywhere, anytime. An obvious but painful conclusion is that the *historical* value of such a translation is virtually nil. Only a faith in a timeless spiritual message transcending the historical reality of the sources could justify such an approach, for which reason, of course, there are no grounds to criticize the worth of such works for faith communities. The alternative, namely focusing solely on one version, produces in turn a *Rashomon*-like problem: as long as one hears only one version, it is easy to accept this as 'true,' but confrontation with multiple, related but distinct, versions evaporates any fantasy of a real and unique truth waiting to be discovered. As historians, we must acknowledge that only a careful accounting of all sources which respects their autonomy can bring us close to valid statements about the past. If we want to know how Mahāyāna Buddhism—one of the major religious movements of human history—developed, and how its ideas came to influence half the world, we need the best possible access to its textual record. This is the only legitimate starting point. The editorial approach of Open Philology will make available in a clear manner the sources of Mahāyāna literature, and thus provide the necessary starting point for historically sensitive investigations of the tradition. The lessons we learn, and the tools we develop, will help others in typologically similar situations treat their own sources more authentically as well.

Objectives

Scholars of Rabbinics provide us with a typology useful for our Buddhist sources, a way to think past the conflation which has plagued many earlier efforts. Milikowsky (2006, 82) differentiates between logical levels: a *'Work'* is what is produced by an author or editor, without ever having existed in any concrete form, be it a manuscript, book, or oral recitation. Less abstract is a *'Document,'* a concrete form of 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 a hypothetical generic class. What really exist are multiple Documents (including recitations), whose words—the Text—(re)present the Work. A crucial theoretical question asks in what sense, then, if we read a single version we can consider ourselves to study the Work—generally presumed to be the ultimate object of our interest. On a more practical level, we require a model that allows users, in accord with individual interests, to change the resolution, or 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.

Given the structural history of Buddhist texts and the nature of their composition, traditional linear editions, citing parallels and variants, are not an adequate, and rather more importantly not an *honest*, representation of their actual situation. Editors have as their first and foremost task a reliable representation of *all* of the available evidence, not only that which supports some predetermined vision. This entails a *moral responsibility* to ensure that texts are visible as artifacts not only of their authors, but also of the communities that received and venerated them. In the case of Buddhist scriptural literature, by nature authorless, this is all the more true: only a full account of the evolving and nonlinear textual history of a given work has any claim to represent it, and thus the tradition, truthfully. Practically speaking, then, what can we as editors do?

Methods

Given the fluidity of the textual tradition, a first crucial task is to **prepare the available materials**, in the case of the MRK in principle in **Chinese, Tibetan and Sanskrit**, in a manner that renders them more accessible, and makes possible the discovery and exploration of their complex lineages and intertexuality. This is best achieved through the creation of aligned multilingual corpora.

As sources for the preparation of the aligned corpora, we have secured access to high-quality (multiply proof-read) machine-readable text data: the **Chinese Buddhist canon**, in the standard Taishō edition (1924-1935), from the **SAT** Chinese Tripiṭaka project of the International Institute for Digital Humanities (Tokyo), and the **Tibetan Buddhist canon**, Derge edition (1733), from the Buddhist Digital Resource Center, **BDRC** (formerly the Tibetan Buddhist Resource Center, TBRC). Both the SAT and BDRC will partner with us in this project, offering technical advice and support.

Scholars wishing to study a Work face the practical problem of moving among its various extant versions. One reading, for instance, a Chinese *sūtra* translation cannot, without considerable investment of time and effort, locate the corresponding spot in a Tibetan translation, or even in parallel Chinese texts. What results is a failure to take cognizance of all available sources, almost forcing a distorted picture of the historical situation. Heretofore, **alignment**—coordination of the corresponding portions of text—among Chinese, Tibetan and Sanskrit Buddhist texts has been done only on an extremely limited basis. While a number of approaches to the problem of automated correlation between two languages have been suggested (e.g. Och & Ney 2003, Bamman et al. 2010), the Natural Language Processing challenges of this approach remain significant. The feasible size of the MRK, however, will allow us to build the necessary correlated text corpora manually. This in turn, will serve as training material for future automated alignment of the canon as a whole.

Automated alignment requires prior **identification of lexical elements**. We might expect the classification of individual lexical items of the Chinese canon to present a sizable challenge, as it does in Classical Chinese in general. However, Buddhist materials offer several substantial advantages. The corpus of available data is already punctuated, a tremendous help in alignment. Moreover, a ready-made (manually created) listing of technical terms, names and so on with a **detailed ontology** (or classifying organization) already exists in the published index to the Taishō edition (Daizōkyō Gakujutsu Yōgo Kenkyūkai 1926-1985). To assess the utility of this tool, we have begun a **pilot project** with the Center for Evolving Humanities at the University of Tokyo to digitize the index of the MRK, and apply its ontology to the base Chinese text. The result is being proof-read at this writing. This material will allow the creation of a propagated shallow ontology across the corpus, facilitating lemmatization of the text. This ontology will allow one to search for topics of interest on a level more abstract than that possible through mere key-word search, and will contribute both to identification of shared modular elements and to future efforts at automated alignment. The ultimate goal is a full alignment of the Buddhist canons in Chinese, Tibetan and Sanskrit (as available), toward which the production of a manual alignment of the MRK is a vital first step.

The **editions of individual texts** are an integral element in our larger project, and the way we treat them is indicative of our overall philosophical approach to this literature. Single exemplars of a given Work must be treated in their context as *witnesses to* that Work, rather than as *identical with* it. We cannot follow the common pattern of generations of editors, namely an **eclectic** presentation, in pursuit of which certain readings are accepted and others rejected or suppressed, because this process assumes an archetype or Ur-text, which does not exist. The eclectic edition, in fact, entails 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 task in such Nietzschean terms, but what else could it mean to read a Work when we have contact only with its static instantiations? Nor can we represent **the multiformity of our sources** by establishing a text along with its 'variants', because (Lord 1960, 101): "we cannot correctly speak of a 'variant,' since there is no 'original' to be varied!" The problem indeed has a moral dimension: by making something a 'variant,' we **hierarchize**, a process inherently complicit in the eugenic and hegemonic program of assuming and asserting authority over the Work-cum-text.

An alternative path is offered by E. Vanhoutte, who suggests "the completely equal treatment of each version of the text in the generating processes invoked by the user" (2007, 165). 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). Van Hulle's approach 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.

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*, comprehensive research into which will be carried out by all team members as a **group project**. We have, principally, witnesses in Sanskrit (7th c.), Tibetan (8th c.), Chinese (in five independent translations, from 174 CE to the 10th c.), a commentary that quotes the text in extenso (extant in Tibetan and Chinese), and numerous quotations in later philosophical treatises. If the

Work "the" *Kāśyapaparivarta* did not grow in a linear fashion, how should we treat the relations among witnesses, versions and the Work? From what standpoint should we arrange our data? And what would such a presentation allow us to learn about the history of the text, the communities that produced it, and their interrelations? 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? Similar questions are posed by the study of each *sūtra* of the MRK collection.

Following Davila (1994, 220) we agree 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." Such an edition will be 'critical' chiefly in that it will provide all of the relevant data, and the editor's interpretation, while not imposing any particular viewpoint on the user. This is the only way to assure an **overall non-hegemonic treatment**, and thus access to the historical instantiations of the text and consequently to the communities to which they belong. As long as we imagined an edition as something compiled on paper, as ideal as this solution seemed, it remained a logistical impossibility, and the absence of this honest presentation of the fluidity of the textual tradition was one factor leading either to a focus on a single version, ignoring the tradition's inherent diversity, or to the type of conflated, and thus historically meaningless, translation referred to above. **The tools now being made available by Digital Humanities bring the project of creating editions which reflect the textual tradition's inherent complexity within reach for the first time.**

We will adapt **an editing environment** that, building on an alignment among the various available versions, will **do away with the assumption of an Ur-text**. It will allow when appropriate, but not require, the establishment of stemmatic relations among witnesses (as will be fitting, for instance, with multiple witnesses to one Tibetan translation). Presenting a Work, which by definition does not have a literal core, is a challenge, even digitally, which we will meet using a system in which 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 will serve as a baseline for textual presentation, keeping in mind that it is a heuristic fiction rather than a historical (re)construction. 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 standardly numerated, in a manner similar to that used for chapter and verse in Bibles. Since this numeration requires careful study of each witness, it can only be carried out on a text-by-text basis as we edit each Work. One candidate for a basis upon which to build our software environment is the Open Source READ (Research Environment for Ancient Documents) system developed by the project "Buddhistische Handschriften aus Gandhāra," concerning which we have had extensive discussion with the project head S. Baums (Munich).

Individual *sūtra*s are Works, but they exist in witnesses. These witnesses must be prepared, transcribed, arranged, annotated and translated. Our editing environment will allow each element of evidence to be combined with other elements, such that a user can decide 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 that will occur to a user but perhaps not to those who build the system. Its structure must therefore also be as flexible as possible, and allow the user to see as much or as little complexity as she wishes at any moment. Our envisioned editing environment will thus allow a reader to do many things, including most importantly:

- navigate easily among versions, to annotation, to translation, to parallel passage
- read any witness in its entirety, from beginning to end, or two or more witnesses in parallel
- read the editor's reconstruction of a particular version
- examine the genesis and relations of the Work, as hypothesized by the editor, by means of a dynamic illustration of textual layers
- 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)
- search intelligently (morphologically, fuzzily)
- offer for mediation comments, additions, and questions on any of the aspects of the presentation.

The openness of the project will not be a wiki-like free-for-all but **a controlled system which will encourage qualified input.** 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." Moreover, we consider the role of the editor absolutely central. 'Good editions' are considered good for a reason: 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, who are, after all, the most highly qualified persons to make such decisions. The editor performs a vital task in making choices which most readers will welcome. In the specific case of translations of Buddhist *sūtra*s in Chinese and Tibetan, for instance, since translations were published at one particular moment it will usually be possible to determine with confidence which readings of a specific translation stood earlier in the tradition, and which represent transmissional errors or emendations. Although the Chinese and Tibetan versions are not likely to be directly related to one another, we may still be able to imagine an Indic source behind an individual term or passage

found in both sources, and this can anchor our alignments. If we could not do this to some degree, it would be impossible to posit that two versions, in two or more entirely different languages, in fact represent the same Work. Most readers will welcome this simplified manner of presenting a text. But for others it may not be primarily the common form of witnesses to a Work that is of interest. Traditional editions make it difficult, and often impossible, to extract information their editors did not consider important, such as 'mere scribal errors'. Such errors, however, while perhaps irrelevant for the establishment of a text as such, might provide invaluable data on, for example, the substrate language of scribes (as 'errors' may provide clues to a scribe's pronunciation). Any edition which does not make the evidence of its sources equally available contributes to a homogenizing process, and hence distorts history. We must therefore carefully balance the editor's expert contribution with the autonomy of users of the edition.

The Project Phases and Results
The project is divided into three phases. In **Phase I** (year 1~2), the PI, Post-doc 1 PhD 1 and our programmers begin work, and we plan for and begin adaptation of our software environments. As a **pilot** we will input in our editing environment already prepared multilingual critical editions of the *Gaṅgottaraparipṛcchā* (MRK 31) and *Maitreyaparipṛcchā* (MRK 42), created by the PI and a PhD student, allowing us to test our systems, evaluating and adjusting as needed. We will hold the first Leiden meeting with our International Advisory Board, consisting of experts in Digital Humanities and Buddhist Studies. We will begin **weekly team study** of the *Kāśyapaparivarta* (MRK 43), alternating with discussions on philological theory. The **PI** will produce a comprehensive edition of a text on dreams, the *Svapnanirdeśa* (MRK 4), and author a book on the MRK as a whole. Post-doc 1 will focus on a *sūtra* dealing with the ethics of the ideal Buddhist aspirant, the *bodhisattva*, the *Raṣṭrapālaparipṛcchā* (MRK 18), while PhD 1 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, related texts, the *Sumatidārikāparipṛcchā*, *Aśokadattavyākaraṇa* and *Vimaladattā-paripṛcchā* (MRK 30, 32, 33).

While aspects of Phase I will continue in year 2, in **Phase II** Post-doc 2 and two further PhD students will join the team. We will develop additional tools, as needed, and refine data curation and metadata standards; emphasis will shift to **implementation of the marked-up Chinese text** of the MRK section of the Taishō prepared by our partners in Tokyo, and **roll out of the alignment environment** to the MRK as a whole. We will hold a second meeting with our Advisory Board. Post-doc 2 will consider the interaction among multiple sources of a text and its commentary in the previously unstudied *Ratnacūḍaparipṛcchā* (MRK 47), PhD 2 will examine claims to the universality of the Buddha's teaching in the *Adhyāśayasaṁcodanasūtra* (MRK 25), and PhD 3 will address a topic of profound philosophical importance, the abstruse universal principle joining all things, the *dharmadhātu*, in the *Dharmadhātuprakṛtyasambhedanirdeśa* (MRK 8). We will pay special attention to seeking parallelisms and pericopes shared across the corpus.

In year 3 we will begin planning for the **International Conference** to be held in year 4, which we will coordinate with a meeting with our Advisory Board. The conference will host not only specialists in Buddhist literature but those working in conceptually related areas, such as Homer, Rabbinic literature and Arabic Ḥadīth. We know of no meeting so far in which specialists from across such diverse areas have shared their common interests.

Phase III involves the presentation of our results, and of a new version of the MRK corpus, aligned among all extant versions. 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 abroad), and published, in Buddhist Studies and in Digital Humanities venues. A conference volume which will synthesize our efforts, and document our discussions with scholars from around the world, will result from our meeting. We will continue our efforts to encourage adaptation and adoption by others of the systems we have developed, and will organize the extended application of our tools to the complete Buddhist canons in Chinese and Tibetan.

Feasibility
Limiting ourselves to the MRK allows us to feasibly align the Chinese, Tibetan and, as available, Sanskrit sources of the MRK *sūtra*s manually, making use of standard, already digitized corpora. We have both the expertise and the proper tools to carry out this task. The resultant corpora will be invaluable not only for the team members but for other scholars as well, and we aim to widely release this material as quickly as possible. Preparation of the edition of the *Kāśyapaparivarta* we will create as a team exercise, and of individual editions of other *sūtra*s, will begin with the collation and digitization of all witnesses, during which time we will modify an existing system to obtain the necessary editing infrastructure. We will begin in year 1 with a **pilot project**, evaluating and adapting our systems before entering the main corpus. We will provide fully Open Data, and all software we develop will be entirely Open Source and Open Access. We will cooperate principally with the BDRC for storage and dissemination of the data we will produce, and the incorporation of our editing environment within their state-of the-art website (http://www.tbrc.org/#!home). We will also prepare tutorials to enable interested users to learn how to deploy the editing environment we create.

References:

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

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.

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).

Grigely, J. 1995. *Textualterity: Art, Theory , and Textual Criticism*. Ann Arbor.

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.

Lamotte, É. 1962. *L'Enseignement de Vimalakīrti* (*Vimalakīrtinirdeśa*). Louvain.

Lord, A. B. 1960. *The Singer of Tales*. Cambridge.

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.

Nagy, G. 2011. "Homeric Poetry and Problems of Multiformity: the 'Panathenaic Bottleneck.'" *Classical Philology* 96, 109-119.

Nattier, J. 2003. *A Few Good Men: The Bodhisattva Path According to the Inquiry of Ugra.* Honolulu

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.

Schäfer, P. 1986. "Research into Rabbinic Literature: An Attempt to Define the Status Quaestionis." *Journal of Jewish Studies*, 37, 139-152.

Silk, J. A. 2008. *Riven by Lust: Incest and Schism in Indian Buddhist Legend and Historiography.* Honolulu.

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.

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.