

## Editing without an Ur-text:

### Buddhist Sūtras, Rabbinic Text Criticism, and the Open Philology Digital Humanities Project

Jonathan A. Silk

*The works attributed to Homer, folktales, Rabbinic literature and Buddhist scriptures share the feature that they all lack an Ur-text, an originary core. How are we to understand the nature of such literature, and how are we to edit it in a fashion that does not impose a unitary perspective onto what is inherently diverse and fluid? The ERC-funded Open Philology project is building a digital environment within which to edit Buddhist scriptures. This paper introduces the problem of fluid texts, sketches how reflections on Rabbinic and Biblical literature shed light on it, and outlines the approach being taken by our team to our multi-lingual corpora of Buddhist materials.*

Indian Buddhist literature has been the object of scientific study for something like a century and a half. The the founding fathers of our field were in many cases schooled in the Classics, although not necessarily, of course, specifically in textual criticism. Such scholars consequently brought with them implicit views about the nature of texts. It is not the goal here to explore the history of the study of Indian Buddhist literature, or the impact that the personal backgrounds of our predecessors may have had on the tradition of inquiry they bequeathed to us. Rather, we will look instead at the present day situation of Indian Buddhist textual studies from a different point of view, looking forward rather than backward, so to speak. In the course of this discussion, I outline some theoretical concerns, some specific problems that face those who study Buddhist scriptures, and the solutions we are trying to develop in our Leiden-based ERC-funded project called “Open Philology.”

Let us begin with a big question: just what is our goal when we study a text? In its broadest frame, the answer can be nothing other than this: we aim to *understand* the text, which means, to understand what the authors of the text intended to say. Or alternatively, but also entirely legitimately—and I would venture to say, often even more interestingly—we may aim to understand what some *audience* understood the authors to have intended to say.

If our goal is different from one of these two options, we have no need of text criticism. If we are content only with our own readings, content to be inspired by our own understandings of what we read without consideration of whether others ever came to the same conclusions, it then does not matter to us what the authors meant, or even really what they wrote. If *we* are the source of meaning, introspection is our only necessary tool. Another way of putting this is to say: we only need philology if we care what persons in the past said and meant.

If we *do* care what others said and meant—and this means, in complicated ways, what they thought—as a historian must, then we have no choice but to strive to grasp the intent of the authors,

in one fashion or another, directly or through the ways this intent has been imagined by historical communities of reception. The way to do this is through the words of the texts, and this requires in the first place determining just what those words were, or were believed to be, which for us comes down to the same thing.

It is simple to articulate this goal, but there the simplicity ends. For the moment we begin to think about this, we must ask ourselves: what *is* the text, exactly? As we all know, a tremendous amount of ink has been spilt by those who have tried to answer this question, which they have done, or attempted to do, often in quite abstract and theoretical ways. Our initial concern is also theoretical, to be sure, but in a rather pragmatic mode. We will turn not in the direction of philosophical reflections on authorial intent and the hermeneutic circle, but in the direction of philology. Therefore, one of the first things we want to know is *how to locate* this text we are talking about.

This does not at first seem to be a difficult question. We have manuscripts, perhaps a number of them. And in the case of Indian Mahāyāna sūtra literature, we probably have translations too, in the first place into Chinese and Tibetan, each of which is witnessed by its own manuscripts or blockprinted exemplars. And we can collect these materials and put them together. But does putting them together mean that we then have “the” text?

Just what *is* the relationship between individual manuscripts and a text? We need to be a bit Confucian here, and involve ourselves for a moment in the ‘rectification of names’ (正名). And here scholarship on rabbinics may be of some help.

Chaim Milikowsky<sup>1</sup> uses the term “Work” to mean “the author’s or editor’s product,” which he then immediately states “may theoretically never have existed in any concrete mode of expression such as a manuscript or book.” This is the broadest framing.

Next comes “Document,” which Milikowsky explains as “a concrete mode of expressing a work.” He then goes on to say: “The text of a *work* is the actual word-after-word presentation of the original product, and the text of the *document* is the word-after-word presentation found in the manuscript. Generally, texts of documents are used to try to reconstruct texts of works, although it is of course legitimate to take as one’s goal the presentation of the text of a specific document.”

For Milikowsky, therefore, the largest unit is the *Work*. But a Work may well be, at its broadest, no more than something like a hypothetical generic class. What really exists is a Document. Or rather—and this is where things get interesting—multiple documents, the words in which—called the Text—represent, or perhaps better simply present, the Work. How can we move from text—words in a document—to a Work? This is a basic question for any editor.

One approach is easy: we can decide that we are interested in one historical moment, that represented by a single version—let’s allow ourselves to use that ambiguous word without further clarification for the moment. We can read our single version—let us say a single manuscript—as is. This might be the legitimate stance of one who is interested in an understanding at a particular historical

---

<sup>1</sup>. Milikowsky 1999: 138n4 ≈ 2006: 82.

moment, in some reception of a work, perhaps for instance a manuscript owned and used by a particular person. But if we choose this approach, in what sense are we studying *the Work*? Moreover, how would we be able to read—*read* in the sense of *understand*—a single manuscript, whether this means reading it as presenting a Work or even simply as coherent in itself? We must ask this question because, if we mean what we say when we speak of restricting ourselves to a single source, there is no way for us to correct even apparently *obvious errors* in that single source. And this is a crucial point. It is possible to transcribe a source, but not necessarily to read it as presenting a Work, or even to read it at all. But there is inevitably a certain amount of editing that goes on as we *read* anything.

#### A Shaggy Dog Story

Once upon a time there was a man. That man had a dog. It was a small, shaggy, white somewhat cute dog, but it wasn't a happy dog. And then one day the dog dog bit the man, and the man did not like that at all. So he bit the dog back.

If you read the sentence “The dog dog bit the man,” and you notice the repetition, you will delete one of the dogs. What are your grounds for doing so? You know that there is no kind of dog called a “dog dog,” so the first “dog” cannot be modifying the second. But as soon as you do this, you are no longer reading your source as it is.

Here we enter more directly the question of what an edition is and should be. At one pole is what is often called the diplomatic edition, a transcript of a source. For Milikowsky, the term:<sup>2</sup> “diplomatic edition’ ... can be used only when the editor means to present the text of a document, not that of a work.” He goes so far as to say that “if one starts from the methodological presupposition that each manuscript represents a unique redactorial moment, we cannot correct the reading of one manuscript according to the reading of another manuscript.”<sup>3</sup> The key point here is that you either transcribe or you don't. We will come back to this phrase “unique redactorial moment” in a moment, but let us just emphasize the intimate connection between editing a text, including correcting it, and reading and understanding it. The second cannot happen without the first, so when Milikowsky claims that one cannot correct one manuscript according to another, he is saying that the reader who treats each source as sovereign cannot appeal to sources outside his manuscript to read and understand it.

The alternative to this diplomatic edition, in the usual vocabulary, is the *eclectic text*. For Milikowsky:<sup>4</sup> “any text presented by an editor who does not purport to give his readership an exact transcription of one document, and presents the text of the work with at least one deviation from the text of the document serving as his base—such a text must be termed an eclectic text.”

---

2. Milikowsky 2006: 86.

3. Milikowsky 2002: 553.

4. Milikowsky 2006: 86.

But now we have a problem: what is the relation between multiple sources and a single work? Milikowsky is a specialist in rabbinic literature, a genre which in very broad terms typologically has something in common with Mahāyāna sūtra literature, and a comparison between the two genres may teach those of us studying Buddhist literature a great deal.

But first let us leave rabbinics for a moment, and think about the Bible. For here things are a bit simpler, at least in this respect. And the reason they are simpler is important. The Bible—speaking here only of the Hebrew Bible—is a well-established Work, the Hebrew text constituting the *textus receptus* found in manuscripts not only agreeing extremely well amongst themselves but also being tracked very closely by the Septuagint and other translations in Aramaic and Syriac. But this is so only thanks to the earlier suppression of almost all alternatives to the Masoretic Text, and an early unification of the textual tradition.<sup>5</sup> This is a point to emphasize, and one that is very important for those in Buddhist Studies to understand. The *historical circumstances* of the transmission of biblical literature have led to a situation in which it is plausible, in the case of the Hebrew Bible, to speak in a practical sense of an Ur-text, something that cannot be done for Buddhist scriptures. And this is so because the Hebrew Bible—or at least most of it—went through a definable “unique redactional moment.” In other words, not only did the textual tradition of the work undergo a process of editing and conscious redaction; in addition, alternatives were subsequently suppressed, for what we might term simply political reasons, leaving us, basically, with the Masoretic text.

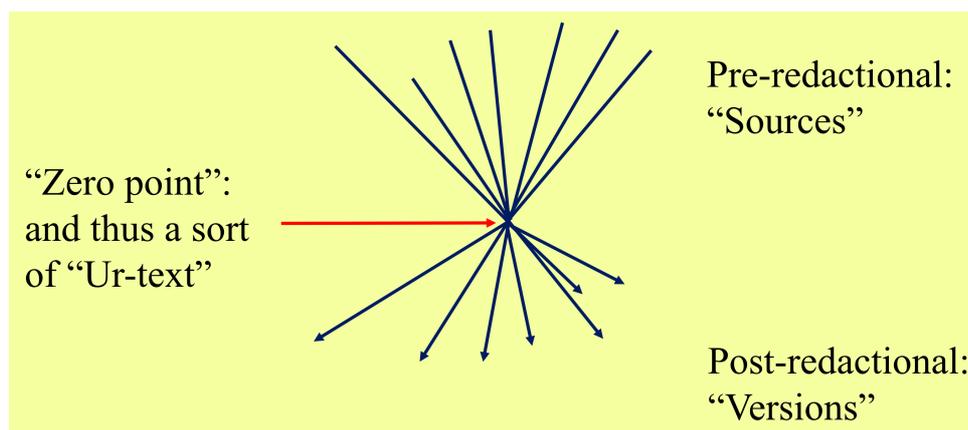
A point to stress here is that, like Buddhist texts, the Bible text also underwent development. We are, however, ignorant of most of that process—its pre-redactional history—due to the political happenstance that a single version was made so dominant as to efface other versions. But this is not so elsewhere in Jewish literature.

Peter Schäfer, in the context of a debate with Milikowsky over rabbinic literature, lays out what is for us a useful scenario. “[E]very writing of rabbinic literature 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.”<sup>6</sup> For the Bible, this is indeed a quite proper formulation. The zero-point represents a theoretically recoverable Ur-text.

---

<sup>5</sup>. James Davila 1994: 219 puts it succinctly: “The Massoretic Text is so uniform only because all other forms of the texts were suppressed.”

<sup>6</sup>. Schäfer 1989: 90.



Indian Mahāyāna sūtra literature, however, for what may be no more than random historical reasons, is not like this. Or at least, it does not appear to us that it is like this. Or to put it yet another way: whatever redactional work may have been undertaken, whatever redactional winnowing may have been attempted, has not left evident traces. There may have been efforts at the canonization of Indian Mahāyāna scripture—in the sense of intentional editing—but if anything like this ever took place in India, these efforts are entirely unknown to us as yet, and they did not succeed in anything like the monolithic fashion one finds with the Hebrew Bible.

However, even those facts that might be learned about the historical development of Indian Mahāyāna scriptures remain unknown to us because we have usually come at this literature with an interpretive framework that has prevented us from detecting this history.

What does one do when one encounters, as happens all too frequently, passages which simply don't make good sense? And the only good answer is that one tolerate nonsense or emend the text. As the doyen of biblical text-critical scholars, Emanuel Tov, reminds us: “the need for emending the text derives from exegesis, but the emendations themselves also need to be acceptable from a textual point of view.”<sup>7</sup> That is to say, we emend when we cannot *make good sense* of the text, but the basis of our emendations must be the form and history of the text we seek to understand. If we cannot make sense of the readings presented in our evidence, and we have no other sources, or refuse or decline to consider other sources because of our views on the nature of the history of the texts, we are at a loss. We must guess. In fancy terminology, this is called “conjectural emendation,” or emendation proper, but it's actually nothing more than educated guessing. It is another question what one seeks to achieve through conjectural emendation, but we must leave this interesting topic aside for now.

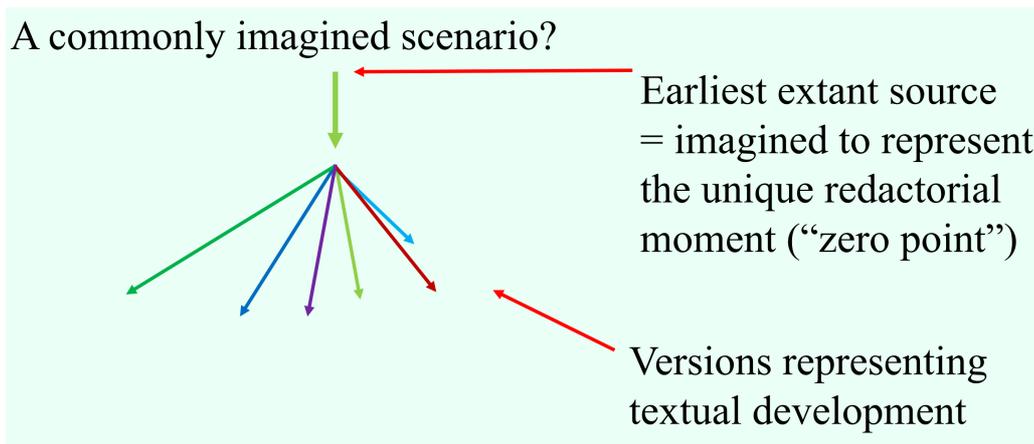
Let us go back to Mahāyāna sūtra literature, and see what we can do to understand its nature. Here we limit our attentions to Indian materials, although some similar things can be said about scriptures composed, for instance, in China, but the case is different enough that we can leave it aside for now.

<sup>7</sup>. Tov 1997: 5.

We can, of course, deduce the nature of Indian Mahāyāna sūtra literature only on the basis of the sources known to us. We may think of the question which these materials present in two aspects: what *sources* do we have, and how were they *composed*. The two questions are mutually informative.

The sources we may have include Sanskrit manuscripts, sometimes of widely variant dates and places of discovery, often multiple Chinese translations, and Tibetan translations, in addition to some other materials on occasion, in Khotanese for instance. Now, in this context, manuscripts differ from translations in that a manuscript is a unique object. At least as we have them now, our translations in Chinese and Tibetan represent the results of self-conscious editorial activity, and are transmitted in multiple versions, some of which of course are in manuscript form. Each Chinese and Tibetan translation, therefore, confronts us with its own problems of textual history, with the lineage of works preserved in Tibetan canonical collections, the Kanjurs, being sometimes especially difficult to untangle. It is possible moreover that some translations represent the results of what is technically called stemmatic contamination, making tracing a text's lineage quite difficult. This is very important, but again, in order to maintain focus, we must leave this problem aside for the moment as well, so that we can return to the core issue.

The various texts that we find in Sanskrit, Tibetan and Chinese in many instances present what appear to be versions of different *layers* of the formation of the text. As the scenario is usually described, typically the text grows over time, that is, the layers represent moments of historical evolution. While this may historically *also* be the case, for Mahāyāna sūtras their apparent fluidity and multiplicity is the nature of their textual production *from the very beginning*. It appears to us to be a feature of the post-compositional development of the literature primarily because of the poverty of our evidence. Without reflection, we do probably generally imagine a “unique redactional moment,” which may be none other than the moment of creation of our least developed source, whether this be an early Chinese translation or something else. This is then postulated as representing the source of other versions. This way of imagining things, however, is an error.



But it is not an error in that we simply cannot access the “original” text, which is perhaps now lost but nevertheless existed. It is an error because the very nature of the genre does not permit the existence of such a kernel or core composition. Thinking again for a moment about the nature of

rabbinic literature may provide a somewhat different model.

Schäfer, referring to the work of Arnold Goldberg, speaks of “textual units the original contexts of which have been lost and which ‘only’ exist in newly coined (and changing) redactional connections.”<sup>8</sup> Schäfer himself introduces a vocabulary motivated by this view of the nature of rabbinic literature, and speaks of Goldberg’s textual units as follows:<sup>9</sup>

I employ the term *macroform* for a superimposed literary unit, instead of the terms *writing* or *work* .... The term *macroform* concretely denotes both the fictional or imaginary single text ... as well as the often different manifestations of this text in the various manuscripts. The border between *micro-* and *macroforms* is thereby fluent; certain definable textual units can be both part of a superimposed entirety (thus a *macroform*) as well as an independently transmitted redactional unit (thus a *microform*).

Elsewhere Schäfer asks the following questions:<sup>10</sup>

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?

And here, for rabbinic literature Schäfer rejects Milikowsky’s view of a “unique redactional moment.” For Schäfer, on the contrary:<sup>11</sup> “the redaction which constitutes the work does not occur at the zero-point, which assigns a firm place to all the individual elements, but is a process which does not allow any neat division between pre- and post-redactional history.” James Davila intends much the same thing when he says:<sup>12</sup> “various redactors worked on the material”—he refers to “the ‘multi-forms’ we have today”—“as it was passed along [so] the text was never redacted into a universal final, or better, canonical form.”

Microforms, macroforms, multiforms. This terminology, while not necessarily in conflict with the world of the Text constituting the Document constituting the Work, speaks of a world of *fluidity*, of ongoing textual creation, a world in which the text is dynamic and alive, not frozen at some particular moment to be locked in a book, only to slip away again thanks to unreliable scribes and sloppy copyists. We do not have here an hourglass with a narrow point through which all passes, only to

---

8. Schäfer 1986: 144.

9. Boustán 2007: 139, quoting Schäfer 1992: 6n14.

10. Schäfer 1986: 150.

11. Schäfer 1989: 9.

12. Davila 1994: 214.

spray out again in a wide, chaotic shower, but a kind of meandering river whose course, and the location of whose banks, is oftentimes not very clear.

As Martin Jaffee explains:<sup>13</sup> “A given microform might consist of a narrative or other unit of tradition that is cycled and recycled in diverse textual versions and is placed in interlocking relationships with other microforms in a variety of documentary contexts. These documentary contexts are not ‘works,’ but ‘macroforms.’” Works, in this view, are built out of shared parts. A given text, a macroform, may acquire new units, and may rearrange them internally, all the while retaining its identity as a unity. It is important to note in this context that the relation between one macroform and another is also fluid, since two more-or-less unified macroforms may also share material.

We do not dare to speak of the social environment within which the ‘editing’ or ‘composing’ activity of Mahāyāna scriptures took place in India, or to speculate on the process through which texts were able to grow, to accrete new materials. Almost nothing is known socio-historically of these processes. Nevertheless, these are some of the most key questions regarding the nature of Mahāyāna scriptures, namely those concerning the nature of their processes of ongoing composition. And here the central insight to be gained from the study of rabbinical literature is that for Buddhist texts we need not imagine, indeed we *should* not imagine, a clean break between pre- and post-redactional stages of composition. The sources we have from the Indian world itself—Sanskrit manuscripts, translations, quotations—all represent essentially randomly preserved evidence of an ongoing and highly fluid process.

And thus a key point: the question of how to understand the *growth* of Mahāyāna scripture is, in this view, identical with the question of how to understand the *nature* of their initial composition. And this becomes clear when we think about the structure of such texts, and most especially with how much they share with each other, and with non-Mahāyāna scripture.

Mahāyāna scriptures are filled with stock phrases, common formulations, shared narratives, and of course, in ways we hardly think about, such scriptures are through and through permeated by the phrases, formulae and structures found in the earlier—that is to say, what is generally considered pre-Mahāyāna—Buddhist literature. If we consider smaller units of text, pericopes, as not merely pre-redactional units, but as part of the stock out of which the ongoing process of composition takes place, this will allow us to appreciate that the fluidity of the macroforms, or the multiforms of macroforms, is a basic feature not of the *growth* of Mahāyāna scriptures as they develop over time post-redactionally, but of their very identity *from the beginning*. It also is of fundamental importance for the way we view relations between different so-called “texts,” or if we prefer, different “Works.” And this, of course, has profound implications for our editorial stance toward these materials.

It can often be useful to visualize abstractions, and one way of visualizing this scenario is to think of building blocks; even if the blocks remain much the same, they can be put together with other blocks in a variety of ways, and sometimes close to the same set of blocks can be put together in

---

<sup>13</sup>. Jaffee 1999: 21.

somewhat different ways.

How can the scholar deal with such a literature? For Davila, there are three possible approaches, which we might rephrase as follows:<sup>14</sup>

1. We can reconstruct an Ur-text.
2. We can reconstruct a hyparchetype, leaving aside the question of the relation between this early redacted form and its predecessors.
3. We can do step two for all versions or lineages known to us, and then attempt to study the relations between them.

There is a hierarchy inherent in these suggestions, of course, as well as a not very well hidden set of assumptions. We can only *reconstruct* an Ur-text if such a thing existed. If we “reconstruct” an Ur-text for a work which did not develop from a single nucleus, then we have created something which we can argue to have been the progenitor of all known descendants, but which in reality never played that role. This does not, however, mean it is pointless.

In some sense, this would be a process akin to *linguistic reconstruction*. It is possible to postulate a proto-form of a language family, such as Indo-European, in such a way that significant features of its latter-day survivals can be explained in an evolutionary manner. But we are long past the days when scholars thought of Proto-Indo-European as something which at one time existed as an actual language. The utility of postulating Proto-Indo-European is that it enables the *modeling* of linguistic *development*, and a greater appreciation of the nature of language. For example, it allows one to understand words otherwise lost in a language, as for instance words in the Veda which no longer exist in Indic but which can be understood through Old Church Slavonic, and so on. It is conceivable, in a parallel manner, that we will find it somehow useful to create a proto-scripture. But if so, we must realize that the result of our construction never existed as such in the past, any more than Proto-Indo-European did.

Whether we choose to follow this course, or instead prefer to maintain the integrity of the sources to a greater degree, we must still negotiate the relations between extant materials we wish to consider part of the same family, in a Wittgensteinian sense, something we might still be content to refer to as a “Work.”

For an editor, however, recognizing the sources that must be accommodated, and accepting that something beyond a diplomatic edition is desirable, is only the first step. What next?

We can take Davila’s second option, and aim to reconstruct a *hyparchetype*. This requires editorial choices, and as Tov reminds us, claiming a certain reading is *preferable* to another implies an acceptance of an “original” since it is a claim that this reading better reflects the original composition, whether this be from the point of view of language, vocabulary, ideas or meaning.<sup>15</sup>

---

<sup>14</sup>. Davila 1994: 215.

Let us consider a concrete example. For some years I have been studying a text we know as the *Kāśyapaparivarta*. But the more carefully I study the Sanskrit, two Tibetan, five complete Chinese translations, commentary and numerous quotations, the less sure I become of just where to find the Work. And what should be my approach to establishing any text in my quest to establish a critical edition? Am I really ready to take the almost complete Sanskrit manuscript and omit passages from it, because for one reason or another I think these did not form part of the Work as redacted—can one even redact a Work?—at *another* moment, a moment therefore either unrelated, or only indirectly related, to the redaction process which generated the extant Sanskrit manuscript itself? Do I want to assume a linear development over time, and if so, do I believe that this chronology is reflected in the chronology of the extant sources?

**A Case Study: \**Kāśyapaparivarta* = *Ratnakūṭa***  
Sources:

**Sanskrit:** nearly complete ms; fragments; quotations.

**Tibetan:** complete in Kanjurs; partial in Dunhuang ms.; quotations.

**Chinese:** 5 complete; 1 extract; quotations.

**Commentary:** In Tibetan and Chinese, quoting sūtra, and its model in another text, without citations.

**Others:** Khotanese fragments, etc.

### Chinese translations

H: *Yiyue monibao-jing* 遺日摩尼寶經. Attributed to Lokakṣema of the Later Han. T. 350.

J: *Moheyan baoyan jing* 摩訶衍寶嚴經. Translated in the Jin 晉 dynasty by an unknown translator. T. 351.

Q: *Da Baoji-jing Puming pusa-hui* 大寶積經普明菩薩會. Translated in the Qin 秦 dynasty by an unknown translator. T. 310(43).

L: *Jiashe jinjie-jing* 迦葉禁戒經, translated by Cuqu jingsheng 沮渠京聲 of the Liu-Song (early- to mid-5th century). T. 1469 (§111-138 in H).

M: *Dasheng baoyun-jing* 大乘寶雲經, *juan 7*, the *Baoji-pin* 寶積品, attributed to \*Maṇḍalasena and \*Saṅghabhara. T. 659.

S: *Dajiashe wen Da baoji zhengfa-jing* 大迦葉問大寶積正法經. Translated by Shihu 施護 (Dānapāla ?) during the Song 宋. T. 352.

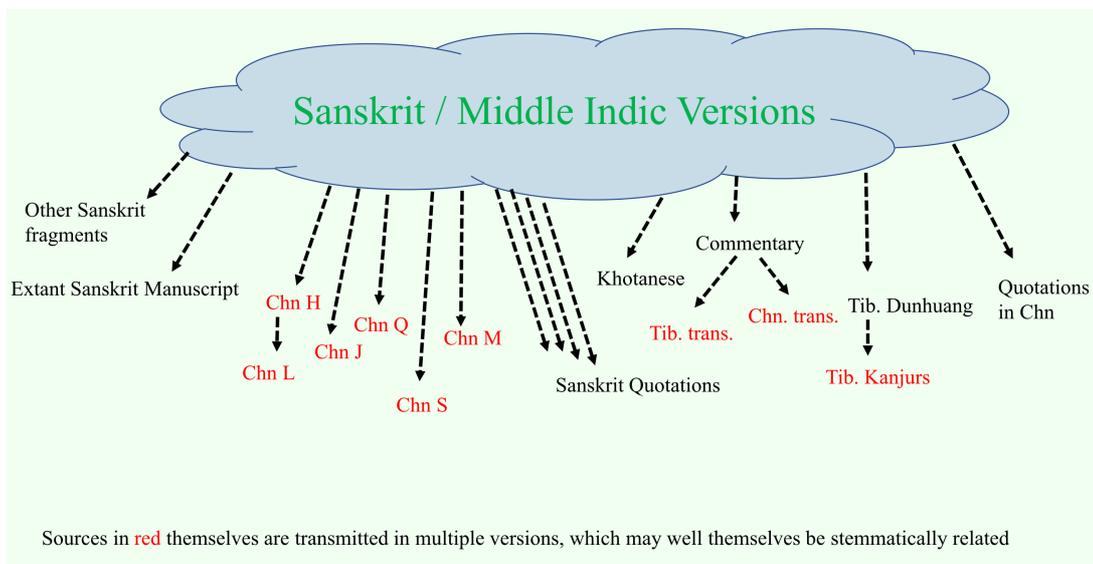
<sup>15</sup> Tov 1992: 168. Note that Tov points out that one can still argue that one reading is better than another, without reference to an Ur-text, if one argues that one reading is a corrupt version of the other.

## Chronology (provisional!)

H Chn. Later Han translation:	174-184
J Chn. (Western) Jin translation:	265-316
Q Chn. (Western) Qin translation:	384-431
L Chn. <i>Jinjie jing</i> (extract of H?):	(400-470)
Cy Chn. Later Wei, <i>Baoji jinglun</i> :	508-535
M Chn. Chen, <i>Dasheng baoyun jing</i> :	557-589
-----	
Sanskrit manuscript (main):	7 <sup>th</sup> -8 <sup>th</sup> c.
-----	
Dunhuang Tibetan translation:	8 <sup>th</sup> -9 <sup>th</sup> c.?
Kanjur Tibetan translation:	788-824
-----	
S Chn. Song translation:	Late 10 <sup>th</sup> c.

If we do want to drop a passage or a sentence or a word attested in one source, what will be our criteria for that removal? Sometimes such changes might be less controversial. There appear, for instance, to be duplications, sometimes words but occasionally larger strings of text which appear to have been repeated by a scribe. Could one object to removing them from an edition which is anything more than a transcript? Shall one correct apparent misspellings? What shall we do when Chinese or Tibetan evidence suggests a rearrangement of some sections in comparison with their ordering in the Sanskrit manuscript? Or shall we, at the other extreme, be content with a collection of various diplomatic editions of my sources? If so, what permits us on the one hand to consider the macroform to exist at all, or on the other hand to 'correct' any one version in any fashion? If each witness is sovereign, how can we intrude even in the slightest way?

Let us look at a visualization of the situation of our sources.



The best explanation of the available evidence is that the Work existed in a very fluid situation in India. What we have is accidentally preserved Sanskrit evidence, Chinese translations from Sanskrit (or Middle Indic, but we may ignore this for the moment), a Tibetan translation, which was revised in Tibet, independent Khotanese evidence, a commentary extant only in Chinese and Tibetan translation, and quotations in Indian works, found in Sanskrit, and in Tibetan or Chinese translation.

To illustrate the situation we have sketched, we may look at what the sources present us, taking some sentences almost at random. In the first example (for the examples, see the texts below), we have a Sanskrit text from our almost complete manuscript, several Sanskrit citations, a Tibetan translation and a number of Chinese translations; the passage is not found in the commentary. In the next example, we have even more Sanskrit evidence, including a similar but not identical expression in a completely different “Work.” This represents a clear case of a very similar modular element which found a home in more than one Work. Sometimes the wording is not precisely the same between versions, or some versions do not have the material at all. Sometimes even fewer versions attest a passage, which is however found in otherwise unexpected sources, in this case a metal strip fragment from Sri Lanka (EZ). Sometimes the Sanskrit and Tibetan agree, but are poorly attested in Chinese. Sometimes not at all. And sometimes we have only one source.

§64

Skt	na khalu punaḥ kāśyapa pudgalabhāvināśāya śūnyatā pudgalāś caiva śūnyatā śūnyatā caiva śūnyatā • atyantāśūnyatā • purvānta-śūnyatā • aparāntāśūnyatā pratyutpannaśūnyatā
Mv	na khalu pudgalavināśāya śūnyatā   api tu śūnyataiva śūnyā   pūrvvāntāśūnyatā   aparāntāśūnyatā ity evamādivistarab
Mt	na khalu punaḥ kāśyapa pudgalavināśāya śūnyatā caiva śūnyatā api tu śūnyataiva śūnyā ... pūrvvāntāśūnyā aparāntāśūnyā pratyutpannaśūnyā ...
Tib	yang 'od srung gang zag rnam par gzhig pa'i phyir stong pa nyid ma yin gyi stong pa nyid stong pa'o    'od srung sngon gyi mtha' stong pa   phyi ma'i mtha' stong pa   da ltar byung ba stong pa ste stong pa nyid la rton par bya'i gang zag la ma yin no
H	不自分別解身爲空也。空棄空中之空本自空甫當來空現在空。
J	不以無人觀諸法空。諸法自空故。如是本空末空。現在世空。當令依空莫依於人。
Q	復次，迦葉，非無人故名曰爲空。但空自空。前際空，後際空，中際亦空。
M	復次，善男子，非無人故名之曰空。但空自空。前際、後際、中際亦空。但法自空非色滅空。
S	復次，迦葉，補特伽羅非破壞空。即體是空，本非有故。非前際空，非後際空，現在即空。

Skt	varam khalu puna kāśyapa sumerumātrā pudgaladṛṣṭir āśrītā na tv evādhimānikasya śūnyatādṛṣṭi-m-ālinā
Pras	varam khalu kāśyapa sumerumātrā pudgaladṛṣṭir āśrītā na tv evādhimānikasya śūnyatādṛṣṭiḥ
RGV	varam khalu puna kāśyapa sumerumātrā pudgaladṛṣṭir na tv evādhimānikasya śūnyatādṛṣṭir
Tib	'od srung gang zag tu lta ba ri rab tsam la gnas pa bla'i   mngon pa'i nga rgyal can stong pa nyid du lta ba ni de lta ba yin no
H	佛語迦葉：人寧著礙大如須彌山。呼爲有其過不足言耳。人有著空言有空。其過甚大。
J	如是迦葉。寧荷我見積若須彌。不以憍慢。
Q	如是，迦葉，寧起我見積若須彌。非以空見起增上慢。
M	善男子，寧起我見積若須彌。莫以空見起增上慢。
S	佛言：迦葉，汝言非也。迦葉，寧可見彼補特伽羅如須彌山量。勿得離我而見彼空。
Lank	idaṁ ca mahāmate saṁdhāyoktaṁ mayā varam khalu sumerumātrā pudgaladṛṣṭir na tv eva nāsty astitvādhimānikasya śūnyatādṛṣṭiḥ

Skt	yā evaṃ pratyavekṣā iyam ucyate kāśyapa madhyamā pratipad dharmāṇāṃ bhūtapratyavekṣā
Pras	yaivaṃ pratyavekṣā iyam ucyate kāśyapa madhyamā pratipad dharmāṇāṃ bhūtapratyavekṣā
Mt	yaivaṃ pratyavekṣā iyam ucyate kāśyapa dharmāṇāṃ bhūtapratyavekṣā
Tib	gang dag de ltar so sor rtog pa de ni 'od srung dbu ma'i lam chos rnams la yang dag par so sor rtog pa zhes bya'o
H	
J	是謂中道眞實觀法。不以無人觀諸法空，諸法自空故。如是本空未空、現在世空、當今依空莫依於人。若有依空，倚此空者，我說是人遠離此法。
Q	如是觀者，是名實觀。
M	若如是觀，是名實觀。
S	如是正觀此說影像中法。

Skt	evaṃ yo n[ā]svabhāvena dharmān asvabhāvikaroti dharmā caivāsvabhāva
Mt	yan nāsvabhāvatayā dharmān asvabhāvan karoti dharmā evāsvabhāva iti
EZ	dharmān asvabhāvan karoti dha///
Tib	gang ngo bo nyid med pas chos rnams ngo bo nyid med par mi byed de
H	
J	不以如觀諸法如，諸法自如故。
Q	
M	
S	法無自性。無性即空。

Skt	evaṃ yan na svabhāvena dharmāsvabhāvatā dharmāṇāṃ yat svabhāvaṃ nopalabhate
Tib	chos rnams nyid ngo bo nyid med par
H	
J	
Q	
M	
S	

Skt	eva-rīṃ yan na agrāhyā dharmān agrāhyā karoti dharmā caivāgrāhyā
Tib	
H	
J	
Q	
M	
S	

How can we possibly begin to build a picture of a Work with this sort of variety of sources?

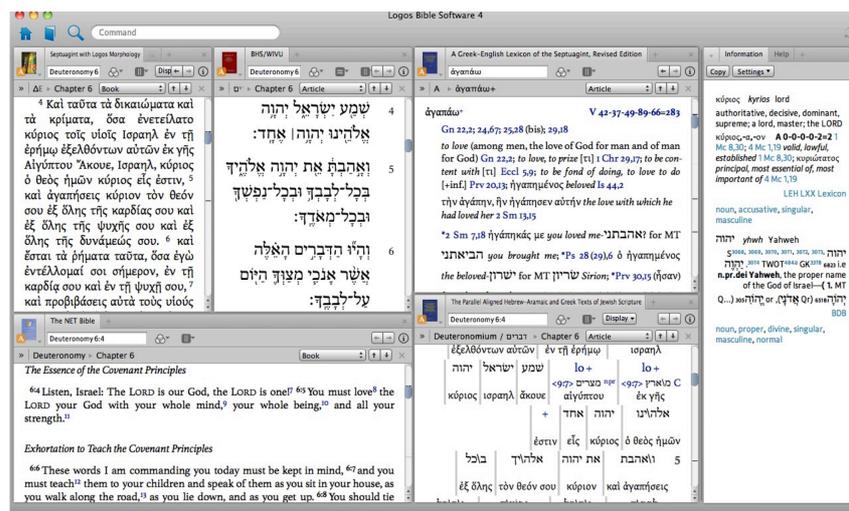
We have a choice: if we believe that a work underwent a unique redactional moment, we should be concerned to reconstruct the form with which it emerged from that redaction, and as a secondary task, we might want to investigate the sources which fed into and were redacted into that final form. But can we really speak of a final redaction for the Buddhist scriptural literature with which we are concerned? Is not the final redaction, as Schäfer asks,<sup>16</sup> “merely the more or less incidental discontinuation of the manuscript tradition?” Is it then the case, as Davila suggests for his materials, that “the ideal way to study ... would be ... to create a massive critical edition that reconstructed every level of development of the document in all MSS, from the earliest redactional levels to the forms in

<sup>16</sup> Schäfer 1986: 150.

the latest and most expanded MSS.”<sup>17</sup> Milikowsky strongly disagrees with this—or rather, he appears to. “[T]he idea,” he writes, “that *the* purpose, or even *a* purpose, of a critical edition should be the presentation of the independent lines of transmission must be rejected out of hand. The realization of this eminently worthwhile goal should be by means of a synoptic edition.”<sup>18</sup> Milikowsky’s apparent disagreement here seems not to concern the desirability of establishing an edition which provides historical information. Rather, he is saying that if we indeed have to do with a *Work*, then we are obliged to decide *what form* we wish to give that *Work*: are we aiming at a critical edition, which assumes we can establish a particular shape, or are we aiming to document the fluidity of a text? The renunciation of the critical, eclectic edition is the renunciation of the idea of an Ur-text, and vice versa, not the renunciation of any effort to document the fluidity of the textual tradition.

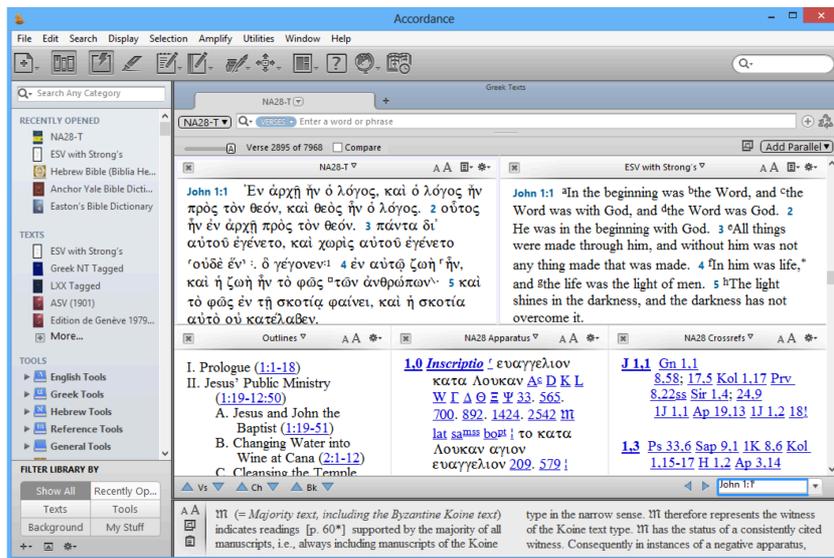
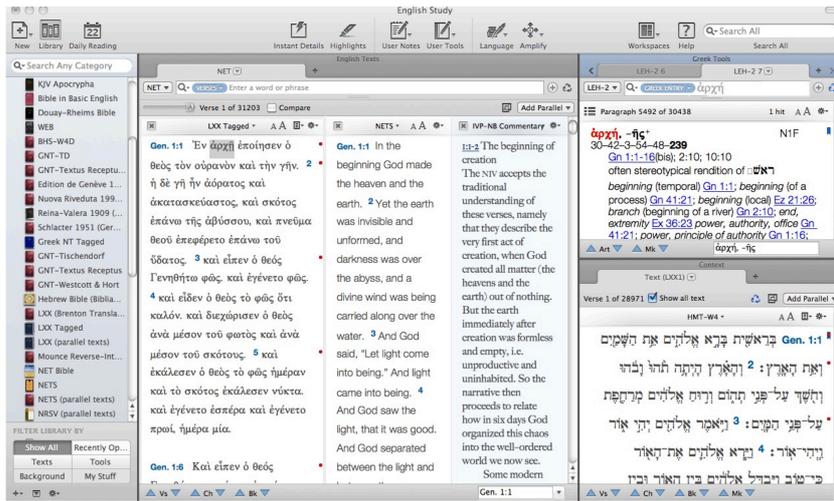
So, what can we do? We cannot do it on paper and in two dimensions, that much is clear. And so now we come to the purpose and the approach of the ERC-funded project Open Philology based at Leiden University. Our materials are Buddhist scriptures, and we begin with a relatively limited collection of 49 texts referred to as the Mahāratnakūṭa (大寶積經). The *Kāśyapaparivarta* mentioned a moment ago forms part of this collection.

The first problem which we attempt to solve is a basic one: what belongs together, what parts of one source correlate with what parts of another? When one deals with the bible, the work of scholars of long ago already aligned the Hebrew text with its Greek translation, just to offer the most basic situation. So the task of putting biblical texts into a digital environment is in some sense trivial. With a nice interface, it is possible easily and elegantly to examine the correlations between the Hebrew text, even to examine multiple Hebrew or Greek sources, to see the grammatical analysis of every word, and so on, but this is possible because all of this work of correlation and identification was done manually.



17. Davila 1994: 220.

18. Milikowsky 2006: 101.



Now, the alignments from the *Kāśyapaparivarta* were also made manually, but this is not practical for the whole of the Buddhist scriptural canon, which is vast. Moreover, the corpora we have, while digitized in standard versions, are not Part-of-Speech tagged. What is worse, there is no good way yet even to identify “words” in Classical Chinese or Tibetan text.

Now, not being able to see what is parallel is very problematic if one wants to study scriptures in detail. Studying a Hebrew passage while looking at the Greek is easy, because the texts have already been aligned. But ours have not. The situation we find ourselves in, however, is not entirely dire, since we know—thanks to catalogues and other manually performed work—mostly what corresponds to what, in the sense that we know that a given text in Tibetan corresponds to something in Chinese, or several things in Chinese. So we do not face the problem that Google faces in trying to determine what English sentence is equivalent to a random sentence in, say, Hungarian: we know where to look, more or less, for the parallels, but it is still not easy to align the sources sentence by sentence. So we are developing algorithms to align the Tibetan and Chinese sources of the sūtras in the Mahāratnakūṭa, and when that is successful we will extend this effort to the remainder of the

canon. In fact, even our alignment algorithms do not need to be perfect (and they will never be), because although large, our corpora are finite and in the end can be hand checked. One of our ideal goals, then, is to align the entirety of the Chinese and Tibetan Buddhist canonical corpora, but that remains a task for the future.

That part of our project is, in a sense, then, easy. But how is this relevant for the whole discussion of microforms? The connection is that the algorithm which connects one Tibetan sentence, defined in our system as a sequence of syllables called an n-gram, to a corresponding n-gram in the Chinese version, can also be used to search for the same or similar n-grams across entire canons. Our alignment engine can be a microform detector. And the power of this is that it does not require identity; it could not, since it functions both within languages (that is, Chinese–Chinese), but also crosslinguistically.

Since we are making the entire source code for our project freely available online, this textual model can be utilized for other projects and even other languages. The complete documentation for our project will be released via our website: [openphilology.eu](http://openphilology.eu).

But this is not all. Because, as with any transmitted text not presented as a *codex unicus*, we do not actually have “a Chinese text,” but rather we have multiple witnesses, which are evidence for the diversity of the Chinese transmission of a particular translation, and the same for Tibetan, for Sanskrit when available, and so on. In the case of the Tibetan and Chinese translations, it may be the case that the witnesses we have can in fact enable us to trace the history of the text stemmatically, such that we can obtain something close to the form the translation had when it left the translator’s pen, that is, the unique redactorial moment. And this may be of great interest to many scholars.

In addition to the philological considerations here, there is also what could even be termed a moral question, so to speak, namely: why would we privilege a reconstruction over real world exemplars which actually belonged to real communities? In other words, a reconstructed form, an Ur-text or a hyparchetype, may be of use, for instance, in reflecting, in a theoretical way at least, the form of a text available to a Tibetan translator at a given moment in time, and that may be of great interest, particularly to those who would use it to retrovert an Indic Vorlage. But for others, the lived existence of the text in other moments of its history and in other places is also, if not even more, important. By focusing only on the hyparchetype, we ignore the historical reality of the text over time. Is it possible to provide access *both* to the diversity of the textual tradition, in its particularity, *and* to editions which offer, for example, reconstructions of transmissional lineages?

Can we create a system which allows users to access a text from different perspectives, with different parameters? We can line up transcriptions of different witnesses, which largely but not entirely agree. It will be possible for a user of the system to select which aspects of the data to employ. Most users may be interested in the first place, and often only, in the results of an editor’s work, which will usually indeed seek to discover the shared core of a given version, and therefore probably its earliest recoverable state. But we know—and here we should recall our cloud diagram—that the earliest recoverable states of the various available versions of a work will almost certainly *not*

allow us to reconstruct some archetype tying them all together. We can go back to a set of related but not commonly descended sources or versions. All of these are “versions” of a Work, but none of them is *the* Work.

There is another advantage to such an open system. Normally, when we edit a text, we carefully collect what we call “variants,” and then in a very defensive manner we equally carefully hide these variants in an apparatus that is for practical purposes inaccessible. They are, after all, mostly mistakes, which is why we consider them variants. But consider: the linguist interested in the phonological history of a language, for instance, might be above all else interested in the errors that scribes make, the places where there is, from a normative point of view, incorrect voicing or aspiration or the like. Or one may be studying not the scripture itself but the work of a scholar or thinker who cited the text. Perhaps the readings under which he cited the text are not those an editor determined to be the most basic and historically valid readings. In a conventional edition, it will be possible only with tremendous difficulty, if at all, for a user of an edition to access the richness of the actual transmissional history of any of the witnesses which have been so carefully and lovingly collated by the editor, but then hidden in an apparatus. It is thus no small victory to have a way to present the results of an editor’s work that allows others to actually make use of it, particularly in ways that the editor him- or herself did not anticipate.

Therefore, the goal is to allow the user to determine what of the available data he or she sees, and how it appears. The environment from the point of view of software development is constructed modularly, to allow anyone to add to the sources provided by us. The structure of our software environment is based on Milikowsky’s model of Works, Texts and Witnesses, so that the philologists on our project, and eventual users of our environment do not require specialist knowledge in software design.

The Open Philology project starts with the belief that we should do all we can to make the artifacts of the past accessible and available to all. By creating an environment in which the fluidity of texts can be maintained, in a manner that each individual user controls, we hope to make a significant contribution toward this end.

---

\*

This work was supported by the European Research Council (ERC) under the Horizon 2020 program (Advanced Grant agreement No 741884).

See the website <http://www.OpenPhilology.eu>

## A Very Short Bibliography of Recent Works on Rabbinic and Biblical Text Criticism

- Boustan, Ra'Anan S. 2007. "The Study of Heikhalot Literature: Between Mystical Experience and Textual Artifact." *Currents in Biblical Research* 6/1: 130-160.
- Davila, James R. 1994. "Prolegomena to a Critical Edition of the Hekhalot Rabbati." *Journal of Jewish Studies* 45: 208-226.
- . 2005. "(How) Can We Tell if a Greek Apocryphon or Pseudepigraphon has been Translated from Hebrew or Aramaic?" *Journal for the Study of the Pseudepigrapha* 15: 3-61.
- Fraade, Steven D. 1999. "Literary Composition and Oral Performance in Early Midrashim." *Oral Tradition* 14/1: 33-51.
- Hayman, A. P. 1995. "The 'Original Text': A Scholarly Illusion?" In J. Davis, G. Harvey, and G. Watson, eds., *Words Remembered, Texts Renewed: Essays in Honour of John F.A. Sawyer* (Sheffield: Sheffield Academic Press): 434-449.
- Jaffee, Martin S. Jaffee. 1999. "Oral Tradition in the Writings of Rabbinic Oral Torah: On Theorizing Rabbinic Orality." *Oral Tradition* 14/1: 3-32
- Milikowsky, Chaim. 1988. "The Status Quaestionis of Research in Rabbinic Literature." *Journal of Jewish Studies* 39/2: 201-211.
- . 1996. "On Editing Rabbinic Texts." *Jewish Quarterly Review* 86/3-4: 409-417.
- . 1999. "Further on Editing Rabbinic Texts." *Jewish Quarterly Review* 90/1-2: 137-149.
- . 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.
- . 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." Carol Bakhos, ed., *Current Trends in the Study of Midrash*. Supplements to the Journal for the Study of Judaism 106. (Leiden and Boston: Brill).
- Schäfer, Peter. 1986. "Research into Rabbinic Literature: An Attempt to Define the Status Quaestionis." *Journal of Jewish Studies*, 37:139-152
- . 1989. "Once Again the Status Quaestionis of Research in Rabbinic Literature: An Answer to Chaim Milikowsky." *Journal of Jewish Studies* 40/1: 89-94.
- . 1992. *The Hidden and Manifest God: Some Major Themes in Early Jewish Mysticism* (Albany: State University of New York Press).
- Tov, Emanuel. 1992. *Textual Criticism of the Hebrew Bible* (Minneapolis: Fortress Press/Assen and Maastricht: Van Gorcum).
- . 1997. *The Text-Critical Use of the Septuagint in Biblical Research*. 2<sup>nd</sup> ed. Jerusalem Biblical Studies 3 (Jerusalem: Simor).