

## Issues in the Use of Electronic Markup for the Comparative Analysis of Āgama Literature

Marcus Bingenheimer  
M.A., University of Wuerzburg

### Summary

The present paper is one of the first steps taken towards the design of an advanced digital edition of the Pāli Nikaya and the Chinese Āgama sutras.

After presenting the two text corpora in digital form, and linking them by a comparative catalogue, it is possible to encode comparative research using TEI markup. Through the formalism of the markup language the resulting digital edition would tell the user in a consistent and precise way about certain aspects in the relationship between two sutras. It is the researcher alone who decides what aspects to encode and how precisely to do so.

The paper outlines (in 1.& 2.) the notion of markup and the reasons for its use in the present case. Then I offer two examples of how research can be encoded into markup. In 3.1 the *occurrence* of a word or phrase in the Pāli and / or Chinese is treated in a comparative fashion. Fully done, this kind of analysis would enable, e.g. a user of the Pāli text to know to what degree literal parallels to her text exist in Chinese. For an example, we look at a gatha from the *Za ahan jing* (T2.99.120b) and its parallel in the *Samyutta Nikāya* (PTS IV, 218).

In 3.2 the *content structure* of the Arrow sutra in Chinese and Pāli is analysed and compared. The results are encoded by way of one markup

line that describes the difference between the sutras according to a previously devised matrix. This perspective allows precise, quantifiable statements of differences between the two corpora as a whole, based on structural differences between sutra versions.

**Key words:** 1. Āgama Literature    2. Pāli Nikāyas  
                  3. Comparative Analysis    4. Encoding Markup  
                  5. TEI (Text Encoding Initiative)

**【Contents】**

- |   |
|---|
| 0. Āgama Literature                         |
| 1. Markup                                   |
| 2. Why use Markup for Comparative Analysis? |
| 3. What to encode?——Two examples            |
| 3.1. Occurrence                             |
| 3.2. Content Structure                      |

## 0. Āgama Literature

Of the two main collections of Āgama literature——in Pāli and Chinese——the Pāli Nikāyas have for various reasons attracted a lot more scholarly attention during the past century. Only the Pāli texts have been translated into English, and it is these that are generally held to be the “the earliest texts.” To this day, some are even convinced that the Chinese Ahan jing are mere translations of the Pāli as it is found in the PTS edition.

However, the differences between the two corpora do not suggest that one is a translation of the other. They are rather the result of different lines of transmission in the Buddhist order. Both are the product of long processes of canonization and editing that came to an end (of some sort) only after the 5<sup>th</sup> century CE. By that time the translation of most of the Chinese Ahan jing from the Prakrit Āgamas of Northern India had been completed and the Pāli Nikāyas had reached a new degree of stability due to Buddhaghosa’s commentaries.

Almost hundred years ago, in order to give a western reader an idea of how large the differences between the two collections are, Akanuma wrote: “if these deviations are not of the kind that we find in the four synoptic Gospels, or of such degree as those between the Gospels and the Apocrypha, they are nevertheless more than the various readings of Shakespeare in the Quarto and Folio present.”<sup>1</sup> Statements like this, and the assurance that the doctrines are basically the same, may have lead to the impression that a comparative study would yield few relevant new insights. Perhaps this is true. It all depends on what one believes to be “relevant.”

Relevancy is also a central feature of electronic markup, the meta-data

---

<sup>1</sup> Anesaki, Masaharu: *The Four Buddhist Āgamas in Chinese — A Concordance of their Parts and of the Corresponding Counterparts in the Pāli Nikāyas*. (1908), p. 2.

that is added to digital texts. This article is a first take on how we could use the advantages of the digital medium for a comparative analysis of the two traditions.

## 1. Markup

Markup is the data that can, or rather must, be added to an electronic text to ensure its usefulness in the digital medium. In order to transfer a text from paper, one needs to encode several features of the text-on-paper which are usually taken for granted. Merely to reproduce the words is not enough. Paragraphs are not there to make a page look nice, but are a conscious effort by an author or editor to encode information regarding the structure of the text. In a manuscript the hand of the scribe, the quality of paper, etc. might be worth mentioning. Therefore, standards have evolved that allow us to encode almost every kind of information concerning any level of a given text. HTML is one of these standards, XML (eXtensible Markup Language) another.

For the examples given here we will use the XML-conformant standard developed by the TEI (Text Encoding Initiative)-consortium (edition P4).

There are different standards, and even within the same standard there is always more than one way to markup a text. The designs used in the present paper are meant as illustrations and do not suggest a final markup solution for the problems in question. They are rather preliminary suggestions toward a digital comparative edition of the Chinese Ahan jing and the Pāli Nikāyas.

## 2. Why use Markup for Comparative Analysis?

We are rapidly moving into an era where the printed text will become a snapshot version of a constantly evolving digital text. Texts, especially scholarly texts, will not anymore be produced with the primary aim of publishing them in print.



Markup is not only a *sine qua non* in the process of digitization, but also a useful research tool for textual studies. For comparative textual studies of the kind the Āgama literature demands, where we have to deal with large text corpora and a complex edition history, the use of markup for comparative analysis offers several advantages.

1. It is precise and consistent. Researchers are encouraged to follow through their own standards in a much more rigorous way than in the usual research paper. Other researchers can access, evaluate and expand the information easily by using a common standard (in our case TEI).

2. It is flexible. The researcher can encode comparative information in one of the corpora (e.g. encoding Pāli equivalents in the Chinese text), in both (by using various linking or alignment methods), or in a third place outside the texts (e.g. as in my digital comparative catalog of Pāli and Chinese sutras [ComCat]).

The researcher decides what and how to encode (in the framework of the standard) according to his/her aims. The use of a standard makes it easier for researchers to communicate with each other, work together, evaluate and streamline their different approaches.

Errors can easily be corrected, and different opinions and interpretations can be added to any one edition, all the while keeping track of all changes by various ways of version-control.

3. The results of one's research, encoded in the markup, can be easily interchanged and reformatted for presentations in print, on the web or further digital uses.

4. The resulting digital edition is cheap and accessible. Through the elimination of printing and distribution costs, scholarly editions can be used by researchers everywhere, independent of the financial means of their institution.

### 3. What to encode?—Two examples

Here, we will, in two examples, consider the encoding of two different areas in comparative analysis: occurrence and content structure. Other important fields, such as the markup of linguistic features, will have to wait until another occasion.

#### 3.1. Occurrence

Occurrence denotes the existence of a certain passage in a text. In the following example, we will consider a gāthā found in the *Za ahan jing* (雜阿含經, T02n0099) in volume two, page 120b of the Taishō edition. There is a closely corresponding gāthā in the Pāli canon in volume four, page 218 of the PTS (Pāli Text Society) edition. The topic is a metaphor that likens the arising of sensations (vedanā) in the body to winds in the sky.

We now want to encode information about which passages occur in both the Chinese and the Pāli version, and which occur only in one version. For the sake of simplicity, we will ignore the problem of variant readings as they are given in the apparatus of both editions and treat the passages as they appear in the main text.

Here is the Pāli text. The underlined passages do not occur in the Chinese:

Yathāpi vātā ākāse // vāyanti vividhā puthū //  
puratthimā pacchimā cāpi // uttarā atha dakkhiṇā //  
Sarajā arajā capi // sitā uṇhā ca ekadā //  
adhimattā parittā ca // puthu vāyanti mālutā //  
tathevimasmim kāyasmim // samuppajjanti vedanā //  
sukhadukkhasamuppatti // adukkhamasukhā ca yā //  
yato ca bhikkhu ātāpī // sampajaññaṃ nirūpadhi //  
tato so vedanā sabbā // parijānāti paṇḍito //

So vedanā pariññāya // diṭṭhe dhamme anāsavo //  
 kāyassa bheda dhammattho // saṅkhyam nopeti vedagūti //

In the XML file of the CBETA (Chinese Buddhist Electronic Text Association) CD we find the Chinese text together with some useful basic markup. Underlined passages do not exist in the Pāli.

```
<lb n= "0120b26"/><lg><l>譬如虛空中</l><l>種種狂風起</l>
<lb n= "0120b27"/><l>東西南北風</l><milestone n= "0120b271b"/><l>四
維亦如是</l>
<lb n= "0120b28"/><l>有塵及無塵</l><l>乃至風輪起</l>
<lb n= "0120b29"/><l>如是此身中</l><l>諸受起亦然</l>
<pb ed= "T" id= "T02.0099.0120c" n= "0120c"/>
<lb n= "0120c01"/><l>若樂若苦受</l><l>及不苦不樂</l>
<lb n= "0120c02"/><l>有食與無食</l><l>貪著不貪著</l>
<lb n= "0120c03"/><l>比丘勤方便</l><l>正智不傾動</l>
<lb n= "0120c04"/><l>於此一切受</l><l>點慧能了知</l>
<lb n= "0120c05"/><l>了知諸受故</l><l>現法盡諸漏</l>
<lb n= "0120c06"/><l>身死不墮數</l><milestone n= "0120c061b"/><l>永處
般涅槃</l></lg>
```

What has been encoded here are linebreaks <lb>s, one pagebreak <pb> (in fact a column break between 120b and 120c), lines <l> in the gāthā and the gāthā as a whole has been taken as a linegroup <lg>. I have added two <milestone> elements, which we will need later.

Especially useful for us are the identifying numbers <n> that enable us to address each line. It saves some work if there is already a basic markup structure with addressable identifiers that one can refer to. If there is none, one can always create one's own (TEI-conformant) structure. In English the text reads:

Like the winds in the sky, different winds blow  
From east and west and north and south [(only in the Chinese:) from the  
other four directions as well]<sup>2</sup>  
Some with dust and some without [(only in the Pāli:) Hot and cool,  
Fierce and easy] many winds there blow  
Thus, in the body arise the various sensations  
Pleasant, unpleasant and neutral [(only in the Chinese:) Defiled or  
undefiled,<sup>3</sup> with attachment or without]  
So does the bhikkhu strive ardently, with right understanding, free from  
passions<sup>4</sup>  
All the sensations the wise will be able to understand  
And by knowing the various sensations he realizes the Dhamma, is without  
impurities  
After he dies, [(only in Pāli:) established in Dhamma], he will not be  
reborn, [(only in Pāli:) he has reached the final goal] [(only in Chinese:)  
forever abiding in Nirvana]<sup>5</sup>

---

<sup>2</sup> I am grateful to Douglas Gildow for suggesting this translation.

<sup>3</sup> I take 有食與無食 to be *sāmisā nirāmisā* ca (lit. with flesh and without flesh) in Pāli. The term *sāmisā* when applied to *vedanā* denotes the *vedanā* that arise through contact with sensual objects, thereby being defiled. The sensations experienced in the *jhānas*, however, are said to be *nirāmisā*, undefiled. See the paper by the Vipassana Research Institute: “Sāmisā and Nirāmisā in Meditation” in *The Importance of Vedanā and Sampajañña*. Nashik (Mahārashtra, India): VRI, 1990 [Reprint 2002], 53~55.

<sup>4</sup> Here the PTS main text Pāli version does not match as well with the Chinese as the variant “*sampajaññaṃ na riñcati*” given in footnote 7.

<sup>5</sup> The last lines do not match well. 身死不墮數 translates *kāyassa bhedā saṅkhyam nopeti*. The Chinese omits *dhammaṭṭho* and *vedagū*. 永處般涅槃 “abiding forever in Nirvana” is not there in the Pāli, but does not look like a addition to fill the meter either. The original must have been different.

Now that the material is ready, let us consider: what are the logical possibilities of “occurrence” for the purpose of comparative analysis? Clearly, there are three.

1. A passage appears in both Chinese and Pāli. (Leaving aside varying degrees of correspondance for now.)
2. A passage appears only in Pāli.
3. A passage appears only in Chinese.

Translated into TEI this set of possibilities could be expressed like this:

```
<interpgrp type= “occurrence”>
<interp id= “oc” resp= “MB” value= “Only in Chinese”>
<interp id= “op” resp= “MB” value= “Only in Pali”>
<interp id= “chpa” resp= “MB” value= “In Chinese and Pali”>
</interpgrp>
```

This is a group of interpretive tools <interpgrp>. It assigns an identifier <id> to each possibility. It also says that someone with the initials of Mister Bean is responsible for the interpretation.

Now we can start with the markup. Here, we will insert the markup in the Pāli text, but we could as well put it in the Chinese text or in an altogether different location.

```
<seg ana= “chpa”> Yathāpi vātā ākāse, vāyanti vividhā puthū
puratthimā pacchimā cāpi, uttarā atha dakkhiṇā.
<xptr ana= “oc” doc= “T02n0099.xml” from= “0120b271b” to= “0120b28”/>
sarajā arajā capi,</seg> <seg ana= “op”>sitā uṇhā ca ekadā
adhimattā parittā ca,</seg> <seg ana= “chpa”>puthū vāyanti māluta.
tathevimasmiṃ kāyasmim, samuppajjanti vedanā
```

sukhadukkhasamuppatti, adukkhamasukhā ca yā  
<xptr ana= “oc” doc= “T02n0099.xml” from= “0120c02” to= “0120c03”/>  
yato ca bhikkhu ātāpī sampajaññaṃ na riñcati.  
tato so vedanā sabbā parijānāti paṇḍito.  
So vedanā pariññāya, diṭṭhe dhamme anāsavo;  
kāyassa bhedaṃ </seg><seg ana= “op”>dhammaṭṭho</seg>, <seg ana=  
“chpa”>saṅkhyāṃ nopeti </seg> <seg ana= “op”>vedagūti</seg>  
<xptr ana= “oc” doc= “T02n0099.xml” from= “0120c061b” to= “0120c07”/>

Here we are saying that the segment <seg> from “Yathāpi” to “arajā capi” has the analytical value (ana) “chpa”. In our <interpgrp> above we have defined “chpa” as denoting a passage that exists in both Chinese and Pāli. The following segment *sitā uphā ca ekadā adhimattā parittā ca* appears only in the Pāli version, therefore it has the analytical value “op.”

We indicate the existence of a passage that appears only in Chinese by using an extended pointer <xptr> that identifies a passage in another document (doc) by using the (from) and (to) attributes. With the (ana) attribute we express that this passage exists only in Chinese “oc.”<sup>6</sup>

If we do not want to add the information into the texts themselves, we can use the <span> element to analyse the two versions. This results in a list of <span> elements that can be attached to the text documents or stored in a different place. The items on the list would look like this:

<span value= “chpa” from= “IDREF for *yathāpi*” to= “IDREF for *arajā capi*”/>  
<span value= “op” from= “IDREF for *sitā*” to= “IDREF for *parittā ca*”/>  
<span value= “oc” from= “0120c02” to= “0120c03”/>

---

<sup>6</sup> We have to tweak the DTD to allow the ana attribute into the <xptr>, but I think this is the best solution. For this kind of comparative analysis it is desirable to have links that can associate interpretative values.

The above example was to illustrate how the occurrence of shorter passages in both or either version of our texts can be encoded. While occurrence can be encoded on the level of single words and passages it is also possible to state occurrence for larger units. To say that a certain sutra exists in both Pāli and Chinese can already be taken as an instance of “chpa.” In many sutras large, distinct parts exist only in either Chinese or Pāli. These could be marked with the same mechanism. But there is clearly more to do in a comparison of the two corpora.

### 3.2. Content Structure

Similarity and difference in the content structure of related sutras is a relevant phenomenon for comparative analysis. We need a mechanism to describe differences on this level, which is considerably more complex than that of occurrence. In order to accomplish this, one has to define a set of structural components. Again, what kind of structures are encoded depends on one’s interests and research aims.

In the following a tripartite content structure is assumed. The beginning of a sutra often outlines a “setting” that describes the where, who and why of the sutra (this could be further divided into “location”, “protagonists” and “speech instance” components). Also, most Āgama sutras have a “topic,” which is usually, but not necessarily, a question answered by the Buddha.

Finally, let us call “argument” the argumentation as it is developed in its logical structure and the ideas and concepts employed.

Comparing a Pāli and a Chinese sutra<sup>7</sup> one might want to say that the one or the other of these structural parts are the same or different. Expressed in TEI as <interpgrp> this becomes:

---

<sup>7</sup> It is well possible to compare more than two sutras in this way. For the sake of simplicity here we will use only two.

```
<interpgrp type= "content similarity">
<interp id= "sameset" resp= "MB" value= "same setting"/>
<interp id= "sametop" resp= "MB" value= "same topic"/>
<interp id= "samearg" resp= "MB" value= "same argument"/>

<interp id= "differentset" resp= "MB" value= "different setting"/>
<interp id= "differenttop" resp= "MB" value= "different topic"/>
<interp id= "differentarg" resp= "MB" value= "different argument"/>
</interpgrp>
```

With these tools, however, we can so far only produce statements of *occurrence* on a content structure level. This does not take us much beyond the previous example, where we could say about a <seg> that it existed in both versions or only in one.

If we want to gain a new take on structural sameness and difference of content we will have to explore further what possible differences between sutras exist.

What then, can happen, assuming there are setting, topic and argument? The logical possibilities are that two sutras have:

“same setting, same topic, same argument”

“same setting, same topic, different argument”

“same setting, different topic, same argument”

“same setting, different topic, different argument” [not useful when one compares Pāli and Chinese]<sup>8</sup>

---

<sup>8</sup> Comparing the Pāli with the Chinese “same setting” does not mean much, because there is only a limited number of settings. Thus, the constellation “same setting, different topic and argument,” is true for too many sutra pairs, e.g. all the sutras that Buddha told Ananda in Sāvathī. It can be a useful category, however, if one first searches for structures inside one tradition and then compares the results. E.g. How many sutras were told by Sāriputta in the Pāli and the Chinese tradition respectively?



“different setting, same topic, same argument”  
 “different setting, same topic, different argument”  
 “different setting, different topic, same argument”  
 “different setting, different topic, different argument” [not useful, this is  
 tantamount to saying the sutras are not related; comparison requires at  
 least some form of similarity]

If we include only the useful combinations into our <interpgrp> we get a more sophisticated scheme. The <id>s can be chosen freely by the encoder.

```

<interpgrp type= “content similarity”>
  <interp id= “sameset” value= “same setting”/>
  <interp id= “sametop” value= “same topic”/>
  <interp id= “samearg” value= “same argument”/>

  <interp id= “differentset” value= “different setting”/>
  <interp id= “differenttop” value= “different topic”/>
  <interp id= “differentarg” value= “different argument”/>

  <interp id= “CS1” value= “same setting, same topic, same argument”/>
  <interp id= “CS2” value= “same setting, same topic, different argument”/>
  <interp id= “CS3” value= “same setting, different topic, same argument”/>
  <interp id= “CS4” value= “different setting, same topic, same argument”/>
  <interp id= “CS5” value= “different setting, same topic, different argument”/>
  <interp id= “CS6” value= “different setting, different topic, same argument”/>
</interpgrp>
  
```

We can now not only encode *if* the topic of two (or more) sutras is the same or different. We can also express which out of six types of content

structure-differences exists between these sutras.

Why would anyone want to know this? Simply, because this kind of structural comparison would lead to objective, quantitative data that could be used to gain new insights in the formation of the early Buddhist canon. If it can be shown, for instance, that in a sizeable number of Pāli-Chinese sutra pairs that have the same setting and topic there are differences in argumentation (“CS2”), this would constitute further evidence that during the centuries of oral transmission the connection between topic and argument was especially susceptible to change.

To illustrate this the *Arrow Sutta*<sup>9</sup> will serve as example. All my comments in [ ] could be expressed with some form of markup. Merely for the sake of simplicity I will keep the focus on content-structure. This does not require any markup in the text (except <id>s). I will show at the end how to encode our conclusion about the content structure in a simple way.

## [1. Part: Setting and Topic]

### 1.1 如是我聞一時佛住王舍城迦蘭陀竹園

爾時世尊告諸比丘愚癡無聞凡夫生苦樂受不苦不樂受多聞聖弟子亦生  
苦樂受不苦不樂受諸比丘凡夫聖人有何差別

Thus have I heard once the Buddha was staying at Rājagaha in the Karaṇḍa Bamboo Grove.

At that time he told a number of bhikkhus: In an ignorant, untaught average person there arise pleasant, unpleasant and neutral vedanā [sensations]. Also in a well taught saintly follower, there arise pleasant, unpleasant and neutral vedanā. What, bhikkhus, is the difference between

---

<sup>9</sup> *Sallatena Sutta*: PTS SN IV, 207; English IV, 139. Chinese version in Taishō vol. 2, p. 120a.

the average person and the saint?

1.2 諸比丘白佛世尊是法根法眼法依善哉世尊唯願廣說諸比丘聞已當受奉行

The Bhikkhus said: The World-honored One is the root, the eye, the garment of the Dhamma, verily, if only the World-honored One were to elaborate on this, we would listen, remember and revere [the teaching].

1.3 佛告諸比丘愚癡無聞凡夫身觸生諸受苦痛逼迫乃至奪命憂愁啼哭稱怨號呼

The Buddha told the bhikkhus: When sensations arise in the body of the ignorant, untaught average person, when they are harrowed by pain and suffering, as if it were going to kill them. They sob and cry sorrowfully, and loudly exclaim their anguish.

[1.3 might be a mistake in the text. The passage is the same as the next section, but without the introduction (“Listen carefully.....”). It looks like an unintentional reduplication of the following, perhaps a scribal error. (I could encode this interpretation. If later someone finds reason to the contrary he / she could correct me, by making use of the same markup mechanism.)]

Assutavā, bhikkhave, puthujjano sukhampi vedanaṃ vediyati, dukkhampi vedanaṃ vediyati, adukkhamasukhampi vedanaṃ vediyati. Sutavā, bhikkhave, ariyasāvako sukhampi vedanaṃ vediyati, dukkhampi vedanaṃ vediyati, adukkhamasukhampi vedanaṃ vediyati. Tatra, bhikkhave, ko viseso ko adhippāyoso kiṃ nānākaraṇaṃ sutavato ariyasāvakassa assutavatā puthujjanenāti? Bhagavaṃmūlakā no, bhante, dhammā ..... la .....

[As we see, 1.3 does not appear in Pāli. The rest corresponds exactly

to the Chinese, except that the Pāli lacks the introductory formula that describes the location. It is for the encoder / interpreter to decide, if the absence of the stereotypical formula “while the Buddha stayed in Rājagaha.....” in the Pāli version is enough to say that the setting is different. I consider the location-part of the setting to be of less importance than protagonists, and speech instance. And these are the same in both sutras: the Buddha speaks to the bhikkhus and he does so unasked. Since the location is not explicitly different in the Pāli, but only absent, I would in this case opt for “same setting.” The topic regarding the arising of vedanas in the wise and the average person is the same in Pāli and Chinese therefore “same topic.”]

[2. Part: Sensations in the untrained person. Unique passages in bold font.]

2.1. 佛告諸比丘諦聽善思當爲汝說諸比丘愚癡無聞凡夫身觸生諸受增  
諸苦痛乃至奪命愁憂稱怨啼哭號呼心生狂亂當於爾時增長二受若身受  
若心受

The Buddha told the bhikkhus: Listen carefully and consider well. I will explain it to you, oh bhikkhus. When ignorant, untaught average persons come in touch with the arising of the various sensations, they add to their pains and sufferings, as if they were going to kill them. They sob and cry sorrowfully, they loudly exclaim their anguish, and their minds get confused. At these times their sensations become twofold, physical and mental.

Assutavā, bhikkhave, puthujjano dukkhāya vedanāya phuṭṭho samāno socati kilamati paridevati urattāḷīkandati sammohaṃ āpajjati. So dve vedanā vediyati — kāyikañca, cetasikañca.

[For the level of this analysis, this is the same.]

2.2. 譬如士夫身被雙毒箭極生苦痛愚癡無聞凡夫亦復如是增長二受身  
受心受極生苦痛

所以者何以彼愚癡無聞凡夫不了知故於諸五欲生樂受觸受五欲樂受五  
欲樂故爲貪使所使苦受觸故則生瞋恚生瞋恚故爲恚使所使

於此二受若集若滅若味若患若離不如實知不如實知故生不苦不樂受爲  
癡使所使爲樂受所繫終不離苦受所繫終不離不苦不樂受所繫終不離云  
何繫謂爲貪恚癡所繫爲生老病死憂悲惱苦所繫

This is like in a man whose body is stuck with two poisoned arrows in whom there arises intense pain. An ignorant, untaught average person, having compounded his sensations to two layers, suffers intense pain on both the level of physical and mental sensations.

Why is that? It is because the ignorant, untaught average person does not understand that out of the five sense-desires [五欲 *pañcakāma*?] he comes in touch with pleasant sensations. By these pleasant sensations he experiences the pleasures of the five senses [五欲樂 *pañcakāmasukhā*?]. And because he experiences these pleasures one becomes afflicted by the tendency to crave [爲貪使所使]. When he experiences unpleasant sensations, then because of this there arises aversion. With this aversion one becomes afflicted by the tendency to hate [爲恚使所使].

He does not well understand the formation and the dissolution, the sweetness<sup>10</sup>, the misery and the abandonment of these two kinds of sensation. Because he does not well understand this, when neutral sensations arise, one becomes afflicted by the tendency towards ignorance [爲癡使所使 *avijjānusayo so anuseti*]. Thus he is bound by pleasant sensations, not able to let go of them; bound by unpleasant sensations, not

---

<sup>10</sup> 味 (“taste”) here probably renders *assādam*.

able to let go of them; and bound by the neutral sensations, not able to let go of them. Thus, bound by craving, aversion and ignorance, he is bound to birth, old age, sickness, death, worry, grief and all the painful trouble.

[The Pāli version is somewhat different:]

Seyyathāpi, bhikkhave, purisaṃ sallena vijjheyyaṃ. Tam enaṃ dutiyena sallena anuvedhaṃ vijjheyyaṃ. Evaṃ hi so, bhikkhave, puriso dve sallena vedanaṃ vediyati. Evaṃ eva kho, bhikkhave, assutavā puthujjano dukkhāya vedanāya phuttho samāno socati kilamati paridevati urattāḷiṃ kandati sammohaṃ āpajjati. So dve vedanā vediyati — — kāyikañca, cetasikañca. [This says the same as the Chinese. Not so the following:]

Tassāyeva kho pana dukkhāya vedanāya phuttho samāno paṭighavā hoti. Tam enaṃ dukkhāya vedanāya paṭighavantam, yo dukkhāya vedanāya paṭighānusayo, so anuseti. So dukkhāya vedanāya phuttho samāno kāmasukhaṃ abhinandati. Tam kissa hetu? Na hi so, bhikkhave, pajānāti assutavā puthujjano aññatra kāmasukhā dukkhāya vedanāya nissaraṇam, tassa kāmasukhaṃ abhinandato yo sukhāya vedanāya rāgānusayo so anuseti.

[PTS - Translation:]

Touched by the painful feeling he feels repugnance for it. Feeling that repugnance for the painful feeling, the lurking tendency to repugnance fastens on him [*paṭighānusayo so anuseti*<sup>11</sup>]. Touched by the painful feeling, he delights in pleasant feeling. Why so? The untaught manyfolk,

---

<sup>11</sup> There are seven anusaya (inclinations, tendencies). One can have anusaya towards kāma-rāgā (sensual passion), paṭigha (grudge), diṭṭhi (views), vicikicchā (doubt), māna (conceit), bhavarāga (craving for continued existence), avijjā (ignorance).

brethren, knows no refuge from painful feeling save sensual pleasure. Delighting in that sensual pleasure, the lurking tendency to sensual pleasure [*rāgānusayo so anuseti*] fastens on him.

So tāsam vedanānaṃ samudayañca atthaṅgamañca assādañca ādīnavañca nissaraṇaṃ yathābhūtaṃ nappajānāti. Tassa tāsam vedanānaṃ samudayañca atthaṅgamañca assādañca ādīnavañca nissaraṇaṃ yathābhūtaṃ appajānato, yo adukkhamasukhāya vedanāya avijjānusayo so anuseti. So sukhaṃ ce vedanaṃ vediyati, saññutto naṃ vediyati. Dukkhaṃ ce vedanaṃ vediyati, saññutto naṃ vediyati. Adukkhamasukhaṃ ce vedanaṃ vediyati, saññutto naṃ vediyati. Ayaṃ vuccati, bhikkhave, assutavā puthujjano saññutto jātiyā jarāya maraṇena sokehi paridevehi dukkhehi domanassehi upāyāsehi, saññutto dukkhasmāti vadāmi.

[The argumentation in this last part is again identical with the Chinese.]

[Here the Chinese and the Pāli both try to explain of how the experience of vedanā by an average person leads to bondage. They differ, however, in their argumentation. The way craving and aversion arise is explained differently. The Chinese states it relatively straightforward:

Sensual Desire -> pleasant sensations -> craving

Unpleasant sensations -> aversion

Neutral sensations -> ignorance

The Pāli makes an elegant little loop:

Unpleasant sensations -> aversion & delight in sensual pleasure

Delight in sensual pleasure -> craving

Neutral sensations -> ignorance

Though both versions converge again in the conclusion, we take this to be an instance of “different argument”.]

### [3. Part: Sensations in the noble disciple]

3.1. 多聞聖弟子身觸生苦受大苦逼迫乃至奪命不起憂悲稱怨啼哭號呼  
心亂發狂當於爾時唯生一受所謂身受不生心受

Now if there arises pain in body of the learned, noble disciple, if he is harrowed by great pain, as if it were going to kill him, he does not give rise to exclaim his anguish full of grief and sorrow, does not sob and cry, and does not become mad and wild. Because at this time there only arises one [kind of] sensation, namely physical sensation and not mental sensation.

Sutavā ca kho, bhikkhave, ariyasāvako dukkhāya vedanāya phuṭṭho samāno na socati, na kilamati, na paridevati, na urattāḷiṃ kandati, na sammohaṃ āpajjati. So ekaṃ vedanaṃ vediyati – kāyikaṃ, na cetasikaṃ. [This is the same.]

3.2. 譬如士夫被一毒箭不被第二毒箭當於爾時唯生一受所謂身受不生  
心受

爲樂受觸不染欲樂不染欲樂故於彼樂受貪使不使於苦觸受不生瞋恚不  
生瞋恚故恚使不使於彼二使<sup>12</sup>集滅昧患離如實知如實知故不苦不樂受  
癡使不使於彼樂受解脫不繫苦受不苦不樂受解脫不繫於何不繫謂貪恚  
癡不繫生老病死憂悲惱苦不繫

Like a man struck by only one arrow, not by two. He at that time has only one [kind of] sensation. What is called the physical sensation does not give rise to mental sensations.

When he comes in touch with pleasant sensations, he is not defiled by the desire for pleasure. Because he is not defiled by the desire for pleasure one does not become afflicted by the tendency to crave for these pleasant

---

<sup>12</sup> 使 here probably a mistake for 受.



sensations. When there are unpleasant sensations he does not give rise to aversion. Therefore one does not become afflicted by the tendency of aversion [against unpleasant sensations]. He understands the formation, dissolution, taste, misery and abandonment of these two kinds of sensations as they are. Because he understands them as they are, he does not become afflicted by the tendency of ignorance in regard to neutral sensations. He is released, not bound by his pleasant sensations; he is released, not bound by unpleasant and neutral sensations. How is this? Because he is not bound by craving, aversion and ignorance, he is not bound to birth, old age, sickness, death, worry, grief and all the painful trouble.

Seyyathāpi, bhikkhave, purisaṃ sallena vijjheyya. Tamenam dutiyena sallena anuvedhaṃ na vijjheyya. Evañhi so, bhikkhave, puriso ekasallena vedanaṃ vediyati. Evameva kho, bhikkhave, sutavā ariyasāvako dukkhāya vedanāya phuṭṭho samāno na socati, na kilamati, na paridevati, na urattāḷiṃ kandati, na sammohaṃ āpajjati. So ekaṃ vedanaṃ vediyati – kāyikaṃ, na cetasikaṃ.

Tassāyeva kho pana dukkhāya vedanāya phuṭṭho samāno paṭighavā na hoti. Tamenam dukkhāya vedanāya appaṭighavantam, yo dukkhāya vedanāya paṭighānusayo, so nānuseti. So dukkhāya vedanāya phuṭṭho samāno kāmasukhaṃ nābhinandati. Taṃ kissa hetu? Pajānāti hi so, bhikkhave, sutavā ariyasāvako aññatra kāmasukhā dukkhāya vedanāya nissaraṇaṃ. Tassa kāmasukhaṃ nābhinandato yo sukhāya vedanāya rāgānusayo, so nānuseti. So tāsam vedanānaṃ samudayañca atthaṅgamañca assādañca ādīnavaṃ ca nissaraṇaṃ yathābhūtaṃ pajānāti. Tassa tāsam vedanānaṃ samudayañca atthaṅgamañca assādañca ādīnavañca nissaraṇaṃ yathābhūtaṃ pajānato, yo adukkhamasukhāya vedanāya avijjānusayo, so nānuseti. So sukhañce vedanaṃ vediyati, visaññutto naṃ vediyati. Dukkhañce vedanaṃ vediyati, visaññutto naṃ vediyati. Adukkhamasukhañce vedanaṃ vediyati, visaññutto naṃ vediyati. Ayam vuccati, bhikkhave, sutavā ariyasāvako visaññutto jātiyā jarāya maraṇena

sokehi paridevehi dukkhehi domanassehi upāyāsehi, visaññutto dukkhasmāti vadāmi.

[Both the Pāli and the Chinese version follow their respective lines of argumentation.]

[4. Part: The fourth part in both sutras consists in a gāthā. Some passages are related, but the original of the Chinese must have differed considerably from the Pāli. There are only two stanzas in Chinese versus three in Pāli. Gāthās in their sūtra contexts pose a number of special problems for comparative research, which I will skip here.]

The analysis of content structure, as we have approached it above, concerns the text as a whole. It is therefore not necessary to split the texts in segments <seg>. What we can do here is to simply link both texts with a <link> element and allow for an analytical attribute like (ana) to carry our interpretation: “same setting, same topic, different argument” (“CS2” in our <interpgrp> above).

<link type= “content structure” targets= “ID chinese sutra \_ID pali sutra” ana= “CS2” resp= “we”>

A number of these links would be an effective way to refine the existing catalogues<sup>13</sup> that relate the Pāli and the Chinese Tripitaka. The use of markup allows us to bring structural analysis to new levels of precision and flexibility. Especially when large text corpora are concerned, the advantages of the digital medium are substantial. Where there is a

---

<sup>13</sup> Anesaki (1908), Akanuma, Chizen 赤沼智善: *Kanpa shibu shiagon goshōroku* 《漢巴四部四阿含互照錄》 (The Comparative Catalogue of Chinese Āgamas and Pāli Nikāyas) . Nagoya: Hajinkaku shobō 破塵閣書房, (1929), and my digital “ComCat” (trial version 2002) available at the DBLM website.

team of scholars working on the same corpus, the use of markup can provide a standard that allows for easy communication and interchange of results.

Using markup may not be everybody's cup of tea, but it has a number of solid advantages over the usual way to encode one's research in the form of annotated essays or translations. In an annotated, printed treatment of a text the author's strategy of what and why s/he annotates and if consistently so, is rarely transparent. The formalism of markup grammar will of course never replace the free flow of the essay, but it can certainly provide us with new helpful tools for textual research, especially in the field of Buddhist Studies.

## 漢巴經典之比較性分析 ——略說電子標記的問題

馬德偉  
德國烏茲堡大學碩士

### 提要

本文是邁向創造一個先進的巴利經典與漢文阿含經電子版的第一步。以比較性目錄連接巴漢兩個文獻匯編之後，研究者可以 TEI 標記符號（markup）記錄其研究結果。透過型式的標記語言，電子版可以標準化且精確的方式告訴讀者某兩個經典的關係。由研究者自己決定記錄的內容、使用的標記語言及標記語言的表達方式。

本文 1., 2. 節敘述用標記符號的原因。接著以兩個例子在 3.1 及 3.2 說明比較性研究如何運用標記符號。

3.1 說明如何使用標記讓所謂的「出現」（occurrence）概念形式化。「出現」代表某一個字、詞或句子是否僅在巴利文或僅在漢文或在兩種文獻皆出現。這種分析，舉例而言，可讓巴利文的讀者知道某一句巴利文在漢文有沒有平行的內容。說明此方法的例子源於《雜阿含經》（大 2.99.120b）及相應部（PTS IV, 218）。

3.2 節介紹「內容結構」（content structure）的概念。設計一個敘述內容結構差異的矩陣後，我們可以簡單的以一行標記符號來表現某兩個經典有何種差異。此研究方法可協助研究者從兩經結構性的差異獲得較具體的標誌，以探討巴漢文獻匯編的相同與相異之處。

**關鍵詞：**1. 阿含經 2. 巴利文阿含經 3. 比較性分析 4. 電子佛典  
5. 標記