Developing Database of the Pāli Canon from the Selected Palm-leaf Manuscripts:
Method of Reading and Transliterating the Dīghanikāya in Khom and Tham Scripts

Suchada SRISETTHAWORAKUL

1. Database Development by the Dhammachai Tipitaka Project (DTP)

Since 2010, the Dhammachai Tipitaka Project (DTP) has launched for creating a database of palm-leaf manuscripts with an aim to publish a critical edition of the Pāli Canon. The DTP makes use of available palm-leaf manuscripts in several scripts through selective and systematic process. The intended outputs will be in forms of a printed edition and an electronic version, both of which will provide readers with the text database and its linked images of the palm-leaf manuscripts. The DTP intends to pioneer this work as the creation of database of the Pāli Canon from selective palm-leaf manuscripts has never been made possible for scholars. This work is relatively new, but highly useful for the entire field of Pāli and Buddhist studies.

Focusing on the process of creating the text database and studying manuscripts, the DTP currently has four teams working on each of the four scripts: Sinhalese, Burmese, Khom and Tham. Each team reads and transliterates manuscripts through an online system for entering manuscript texts: ODEM (Online Data Entry System of Manuscripts). The ODEM provides the readers—the project staff working on transliteration—with the manuscript images and the preliminary text so-called “a base text” taken from Chaṭṭha Saṅgāyana Tipiṭaka (Be). The readers need to read, check and alter the base text so that it matches with the given text in the manuscript image. The preliminary text or the base text enables readers to save time and help facilitate readers’ work.
2. Reading and Transliterating Palm-leaf Manuscripts

2.1. Background Knowledge for Reading and Transliterating Manuscripts

Text in the manuscripts is not in the standard format as in the printed edition. For example, it usually has no space, less punctuations than any printed edition, consonants occasionally departing from vowels, varied styles of correction and deletion, and non-standard handwriting style.

A good reader should have background knowledge about grammars and languages as used in the manuscripts such as Pāli and Sanskrit. Moreover, they should know about paleography of a script inscribed in the manuscripts, for example, Khom or Tham script. Manuscript backgrounds, lineage, scribes, etc. are also useful information, especially when readers encounter problematic issues and have to make a decision in some difficult reading.

2.2. The DTP’s Policy on Reading and Transliteration

The DTP’s policy concerning reading and transliteration is to assure that all readers transliterate the text as close to the selected palm-leaf manuscript as possible; record all information as appeared on the manuscripts, and follow the data management protocol. The readers would read and alter a base text on a basis of “syllable by syllable” against each manuscript. For a conjunct consonant, the readers shall see a conjunct consonant as one big consonant, which is inseparable. When a syllable is broken and a vowel comes at the first syllable, the readers must keep a vowel along with the preceding consonant.

Figure 1. When a vowel comes at the first syllable, it must be kept a vowel along with the consonant. 1)
For symbols and IT codes, they are used to record all information in the manuscripts, such as deletion or correction. For example, in the case of any correction found in the manuscript, the readers would see the text syllable-by-syllable, not letter by letter and use the symbol <<–, +>> to record it. In the case of text corrected by ink, the parentheses ((    )) will be used instead of <<    >>.

### Figure 2. Samples of the correction and deletion

![Sample of correction and deletion](image)

3. The Method of Reading and Transliterating Ambiguous Letters

Normally, although most letters are not ambiguous to readers, a few are and need a further clarification. Under this circumstance, readers would need to take into consideration the overall content, context and grammar in order to understand any ambiguous letter. Whenever one makes a comparison between the Pāli Canon in manuscripts from different traditions, one usually finds slight variations. In most cases, such variations are usually limited to variant readings, texts at the beginning and the end of the manuscript including uddāna, sutta titles and peyyala (...pe...). Generally, a scribe tends to be so conservative that he/she tries to keep every single word as close to the original manuscript as possible, even though in some cases he/she may come across the incorrect use of grammatical words or obvious mistakes. Moreover, the texts in the manuscript are often handwritten, and in some cases they are not always readable.

#### 3.1. Policy and Ambiguous Letters

Our policy is mainly to do with a reader’s decision making. It is necessary to ensure quality control among our working group. Previously, the policy “Read by what you see” had been strictly used in reading the manuscripts. However, this policy made the readers try to read by what they see and read the letters as a picture. They read a letter from part by part and over-focus on the shape of a letter so much that they cannot read
a manuscript smoothly. One of the consequences is that what they read differs from the actual content in the manuscripts especially when they come across ambiguous letters. Therefore, we have become to use our policy more flexibly. This enables the readers try to read the manuscripts by taking into consideration the context and read letters as a language. They read the whole contents continuously and focus more on the overall content than single letters. The readers can then gain a good understanding of the texts close to the original and so the reading is effective and usable.

3.2. Readers/Scribes and Ambiguous Letters

Those readers who are familiar with the script that they use in their daily life tend to have more capability to deal with the ambiguous letters than those who are not. The former have more experience of variations in handwriting than the latter. For example, Cambodian readers reading Khom script tend to deal with the ambiguous letters better than Thai readers reading Khom script.

The quality of the actual manuscripts depends on the experience and skill of scribes. Some unclear letters or mistakes found in the manuscripts reflect a lack of the experience and skill of the scribes. These unclear letters sometimes also indicate the poor condition of the source document of that manuscript.

For instance, for the terms “porisadhammo” and “herisadhammo” in Khom script manuscripts, normally a skillful scribe does not confuse “po” with “he.” When the consonant “pa” becomes “pā,” the form of the letter would be transformed into the special letter to avoid the confusion with “ha.” However, for an inexperienced scribe, this becomes a problematic issue. As shown in the figure 3, the scribe who inscribed “porisadhammo” was more skillful than the other scribe who inscribed “herisadhammo.”

\[
\begin{align*}
\text{porisadhammo} & \Rightarrow \text{ha} \\
\text{herisadhammo} & \Rightarrow \text{he}
\end{align*}
\]

Figure 3. Sample of confusion between “po” and “he” in Khom script
3.3. Ambiguous Letters and Case Studies

Whenever the reader finds it difficult to read any letter or word, i.e., either that he/she cannot read or that it can possibly be read and/or understood in more than one way; he/she needs to check whether the word itself is ambiguous or if it is just the handwriting style that differs from the standard shape. To verify this, the readers need to compare each ambiguous letter or word with the same letter or word as appeared elsewhere in the text. If the unusual shape of the letter or the unusual spelling of the word appears many times consistently, it is likely to be just a writing style of the scribe, i.e., that he/she writes that letter or word in his own way which is different from the standard shape. In this case, it is just an unusual style of writing that causes difficulties in understanding.

One important way to deal with the difficulties in reading is to try to become familiar with scribes’ handwriting as much as possible. The readers who are more familiar with particular handwriting will have an ability to read and understand what the scribe inscribed. Collecting a sample of letters or words in several texts/manuscripts will also help the readers accumulate experience with variant handwriting styles. Moreover, the readers have to consider choices of possible letters or words, and such consideration is based on the paleography of the script.

3.3.1. Case Studies: ṭa or ja

![Figure 4. “ṭa” or “ja” in Khom script](image)

In this manuscript, confusion between “ja” and “ṭa” randomly appears several times. It is plausible that the shapes of “ja” and “ṭa” in the source document of this manuscript were either similar or unclear. The scribe perhaps did not know Pāli very well. As a result, the scribe was using “ja” and “ṭa” interchangeably. In this case, using “ja” and...
“ṭa” randomly and interchangeably seems to be just the style of the scribe. Therefore, with an understanding that the letters “ja” and “ṭa” are interchangeable in this manuscript, the reader should be able to read according to the overall content.

3.3.2. Case Studies: Group of Similar Letters

Ambiguous letter also relates to a group of similar letters. Sometimes, the reader cannot read the text clearly because of the letter of the group of similar letters. The reader may read in more than one way and may not be able to decide immediately what it should be. The readers may limit the possible ways of reading that letters by having the knowledge about the group of similar letters. In this case, understanding an overall content or grammar is helpful for a decision making.

|  |
|---|---|---|
| cā |  |
| thā |  |
| pā |  |
| dā |  |

paricāreti
tathāgato
pākārasandhim
udāhu

**Figure 5.** Samples of similar letters in Khom script

3.3.3. Case Studies: ekadimāhaṃ/ekadāmihaṃ/ekamidāhaṃ

Other information such as inscribed tradition, local culture, script development, etc. should also be considered.

**Figure 6.** Scribal error “ekadimāhaṃ/ekadāmihaṃ/ekamidāhaṃ” in Tham script

Here, in the Tham script manuscript, this word can be read in many ways as follows: “ekadimāhaṃ”, “ekadāmihaṃ” or “ekamidāhaṃ.” Considering the order of the letters, it can be read as “ekadimāhaṃ” or “ekamidāhaṃ.” However, focusing on the content, it should be read as “ekamidāhaṃ.” In the tradition of Tham script, various and flexible
patterns are found more than other traditions. Moreover, the letter “ma” should be in the conjunct consonant form by its position but it is still in the full form. In this case, it seems to be a mistake by the scribe. Taking into consideration the overall content, the readers should then read “ekamidāham.”

4. Conclusion

In conclusion, in dealing with difficulties due to the ambiguity of words and the unusual shapes of letters, etc., the readers must take into consideration the overall content, context and/or grammar. Furthermore, readers need to be aware that each ambiguous letter could be read in several ways according various points of view, so to decide what each ambiguous letter is it would be different case by case.

Creating and developing the database of the Pāli Canon from the selected palm-leaf manuscripts are challenging and complicated. The creator and developer have encountered many difficulties in every single process. Focusing on reading and transliterating the manuscripts, the crucial concept is to read and transliterate as close to the original text as possible and with the concern on the context and the scribe’s handwriting style. Although we can never know habit and the handwriting style of the scribe for sure, to some extent we can still understand what the scribe means as the text in the manuscripts would gives us some clues. To comprehend the scribes’ style, skill and/or experience, the readers must read and consider the texts skillfully and contextually. To produce a satisfied result, the readers must make an effort to find out the scribal errors while trying to collect all data and deal with interesting issues. Learning from various case studies is one of the best ways to improve efficiency in terms of reading and transliteration.

Notes
1) ODEM Reference for readers (Handout for reading and transliterating of the DTP).
2) ODEM Reference for readers (Handout for reading and transliterating of the DTP).
3) Appendixes (Khom) ODEM Reference for readers (Handout for reading and transliterating of the DTP).

Key words Pāli Canon, Palm-leaf manuscript, Khom, Tham, Transliteration
(Researcher, Dhammachai Tipitaka Project, PhD)