# A Phylogenetic Approach to Textual Criticism:

The Case of the Xianyu jing in Tibetan

# ISHIDA Katsuyo

## 1. Introduction

Textual criticism of the Tibetan Kanjur has been performed for each individual sutra. As most of such criticism has been manually executed by researchers, the purpose of this research is to develop an algorithm to enable computer-assisted textual criticism.

In previous articles, I have discussed the use of phylogenetic estimation method used in biology to analyze variants of texts(Ishida 2011; 2013). This paper proposes an algorithm for textual criticism using this phylogenetic estimation method.

In textual criticism, the phylogenetic estimation method can clearly indicate the position of an arbitrary text in the phylogenetic tree using the property "an archetype is close to the root of the phylogenetic tree." To find a text close to an archetype, readings (variants or new readings) are selected by trial and error so that the text approaches the root of the phylogenetic tree. Searching by trial and error requires time and efforts; some heuristics are necessary. In this paper, I use the majority rule approach.

#### 2. Method

### 2.1. Work Procedure for Textual Criticism

Textual criticism is generally performed in the following steps (Yamano 2013):

(Step 1) Collect as many texts as possible.

(Step 2) Make a phylogenetic tree and classify texts into groups.

(Step 3) Select a base text that is a candidate critical text.

(Step 4) Reconstruct a critical text, especially select readings (variants or new readings).

#### 2.2. Algorithm

The algorithm for textual criticism is divided into two parts: Algorithm A and Algorithm B.

(168)

Algorithm A performs Step 2 and Step 3. Algorithm B performs Step 4. The details of Algorithm B are still under consideration and will be reported in the future.

In Step 2, phylogenetic estimation is performed with the tool SplitsTree4 and is based solely on variants. Although SplitsTree4 was designed for biology, it has been used for phylogenetic estimation of texts such as "Genji Monogatari," "The Canterbury Tales," and "Little Red Riding Hood." Phylogenetic estimation is not possible with cluster analysis, a type of multivariate analysis. Many previous textual criticism studies have used classification (which is based on similarity) instead of phylogenetic estimation (which is based on branches).

In Step 3, readings (variants or new readings) are selected according to "selection criteria." For example, the following can be considered:

- (1) Using the majority rule approach
- (2) By trial and error
- (3) Using qualitative judgment

In this paper, I adopted the majority rule approach, which is versatile and easy to implement. The most frequent variant is selected by the principle of majority rule.

While there may be opponents to the principle of majority rule, I decided that it is sufficient for the purposes of this research. The purpose of Step 3 is to select a candidate critical text, which serves as the input for Step 4 rather than the final result of the textual criticism. It should be noted that the principle of majority rule is not applied to all texts but to texts belonging to each group.

## 3. Data

#### 3.1. Texts Utilized

The twelve texts of Chapter 2 of the *Xianyu jing* in Tibetan (*Sutra of The Wise and the Foolish*, hereafter SWF) are utilized in this study. Although four groups are given here for explanation, this information is not used in the phylogenetic estimation.

Tshal pa group	Them spangs ma group	Independent group
C Cone	L London manuscript	F Phu brag manuscript
D Derge	S Tog Palace manuscript	Mixed group
J Lithang	T Tokyo manuscript	N Narthan
U Urga	Z Shey Palace manuscript	
Y Yongle		
P Beijing		

### 3.2. Variants

After collating 12 texts, 523 variants were obtained. The variants were edited into a table format, which was  $12 \times 523$ =6,276 in size. It is nearly impossible to manually process 6,276 data and computer-assisted analysis is required. Although previous studies have reduced the number of data using external information, in this paper, original data is input to the phylogenetic estimation tool without undergoing compression.

The variant used in this paper is a phrase-level difference, but it is not limited to a phrase-level variant. The input data for the phylogenetic estimation tool is "difference between texts" and is not limited to variants. Other representation forms may be more suitable, such as a storyline in narrative literature.

## 4. Results

## 4.1. Classification of texts

Fig. 1 shows the phylogenetic tree using the phylogenetic estimation tool SplitsTree4 based only on the variants. As there is contamination in the process of transferring various texts, the estimation result is represented by a network (unrooted tree) rather than a tree (rooted tree). For convenience, a network is referred to as a tree in this paper.



Fig. 1 A phylogenetic tree of SWF

It can be seen that the twelve texts are roughly divided into the following four groups:

- (1) LTSZN: Them spangs ma group
- (2) JCUD: Tshal pa group
- (3) PY: apart from JCUD
- (4) F: independent group

These results are almost the same as the results of the previous study of Kanjur; PY is often classified in Tshal pa group, but it is far away from JCUD. This point will be the subject of future research. It is not possible to determine which of the four groups is the oldest using the phylogenetic tree. Instead of assuming one ancestor for the four groups, I will seek an ancestor for each group.

#### 4.2. Selection of the Base Text in LTSZN

The base text in LTSZN will be found by introducing the principle of majority rule and se-

(172)

lecting the most frequent variants among the five texts for each variant. Fig. 2 shows the result of phylogenetic estimation using SplitsTree4 based on the variants of six texts including X and LTSZN, where X is the text selected by majority rule. X is nearly at the center of LTSZN. It can be determined that X should be used as a base text in the next Step 4.



Fig. 2 Selection of the base text in LTSZN

#### 4.3. Selection of the Base Text in JCUD

Likewise, the best base text in JCUD will be identified. Since there are four texts, two kinds of texts, Y1 (DU oriented) and Y2 (CJ oriented), will be identified. Fig. 3 shows the results of phylogenetic estimation using SplitsTree4 based on six texts with the addition of Y1 and Y2 to JCUD.

The texts Y1 and Y2 are located near the texts that are emphasized, and it can be determined that Y1 and Y2 should be used as the base text in the next Step 4. The text that will be selected (or merged) will be chosen in the next Step 4.



Fig. 3 Selection of the base text in JCUD

## 5. Conclusion

This paper shows that Algorithm A can be used for Step 2 and Step 3 in the textual criticism of SWF. Although Step 4 has not yet been executed, the base text selected by Algorithm A is considered to be very close to the target text (the ancestor of each group).

#### Bibliography

Ishida Katsuyo 石田勝世. 2011. "Tōkeikaiseki no seibutsugakuteki shuhō niyoru tekisuto no keitōju sakusei: Zōyaku Hannyashinkyō wo chūshin ni"統計解析の生物学的手法によるテキストの系統樹 作成: 蔵訳『般若心経』を中心に. Indogaku Bukkyōgaku kenkyū 印度学仏教学研究 60(1): 414-411. \_\_\_\_\_\_. 2013. "Seibutsu keitōgaku no tōkei shuhō wo riyōshita Zōyaku Kengukyō no keitō suitei" 生物系

2012: Solotada kenegaka no toket shaho we hytoshita Zoyaka kengakyo no kene shahi エレバ 統学の統計手法を利用した蔵訳『賢愚経』の系統推定. Indogaku Bukkyōgaku kenkyū 印度学仏教学 研究 62(1): 471–467.

Minaka Nobuhiro 三中信宏. 1997. Seibutsukeitougaku 生物系統学. Tokyo: University of Tokyo Press. Yamano Chieko 山野千恵子. 2013. "Tekisuto kōtei no riron: Bukkyō tekisuto no kōtei" テキスト校訂の 理論: 仏教テキストの校訂. Rengeji Bukkyō kenkyūjo kiyō 蓮花寺佛教研究所紀要 (6): 169–142.

Key words textual criticism, phylogenetic estimation, Kanjur, Xianyu jing

(Research Student, Kyushu University)

(173)