

**Feedback on L2/20-148 Collection of Un-encoded Chinese
Characters from Tsinghua Bamboo Slips**

Andrew West

10 July 2020

1. The Tsinghua collection of early bamboo slips is very important, and their texts have been the subject of intense study by scholars in China, USA and Europe, so it is very important to be able to represent the text of these bamboo slips in digital format.
2. There is no intention to encode bamboo script forms of characters (as is the case with oracle bone script), and scholars have shown no interest in representing the ancient brush-written bamboo script forms directly as encoded characters, but universally transcribe the bamboo script forms into modern character forms that reflect the structure of the bamboo script characters. In many cases these transcribed forms are not the same as (and not unifiable with) any existing encoded CJK ideographs because they differ in structure from the equivalent modern form of the character. It is therefore necessary to encode these artificial transcription forms of CJK ideographs in order for scholars to represent these texts digitally.
3. The UK already submitted 139 characters used in some of the Tsinghua bamboo texts for encoding as part of IRG Working Set 2017. This is because these characters were used in academic papers by several authors published in the journal *Early China* (ISSN 0362-5028). *Early China* uses Unicode encoding of CJK characters for all submitted papers, and currently needs to represent unencoded transcription forms of bamboo script characters (as well as transcription forms of characters used on bronze inscriptions) as PUA characters. As some of these characters reoccur in different papers by different authors, the editor of *Early China* is very keen to be able to represent them as encoded characters rather than PUA characters. This demonstrates the practical demand for encoding these characters.
4. There are several issues that need to be addressed when encoding transcriptional forms of Old Chinese characters. Firstly, different scholars may analyse the structure of the same bamboo script form differently, and therefore produce somewhat different modern character forms (perhaps with a different arrangement of components, perhaps with different identification of one or more components). Even if two scholars analyse the character the same, they may still produce slightly different modern character forms (e.g. with Kangxi Radical 66 written as 支 or 攴). Therefore, in some cases it may be necessary to seek scholarly consensus on the most appropriate modern character form. Secondly, some of the transcriptional forms may be unifiable with existing encoded characters, even though the scholars prefer the particular form they produce. For example, L2/20-148 no. 02-01-02 (𠄎𠄎單口戈) should be unifiable with U+210E3 𠄎 as the left side component (𠄎單口) is a literal transcription of the bamboo script form of U+563C 𠄎, and only varies by the extension of the central vertical stroke. Examples such as this could be dealt with using an IVS solution. Thirdly, because the original bamboo script forms are handwritten by different scribes, there is some variation in the glyph forms of individual bamboo script characters, which may result in two or more modern transcriptional forms of the same basic character (e.g. 04-01-23, 04-01-24, and 07-04-76; and 04-01-25 and 06-02-02). We recommend that such variant forms should be unified where appropriate.
5. We are willing to work with the author of L2/20-148 to produce a font and prepare the characters for submission to IRG for encoding.