Annotating Rhyme Judgments for a Complex Corpus of Manuscript Sources: Making Sense of the *Cang Jie pian* 蒼頡篇

Establishing a standardized annotation framework for communicating rhyme judgments identified in historical texts will both ease the use of computational tools for rhyme analysis, and hopefully inspire greater collaboration amongst scholars interested in historical linguistics. The framework we have proposed (List, Hill & Foster 2019), was designed with simplicity, exhaustiveness, and flexibility in mind (p.30), with the intension of eventual inclusion in the Cross-Linguistic Data Formats initiative (<https://cldf.clld.org>). Further testing of the framework is desired to demonstrate its utility and identify areas requiring refinement.

Although developed initially with a focus on the reconstruction of Old Chinese through an analysis of received ancient Chinese texts, such as the *Shijing* 詩經 (*Classic of Poetry*), the framework might be applied fruitfully to other types of Chinese texts and to different poetic traditions besides. One such example that we raised was the opening chapter of an important Chinese scribal primer, known as the *Cang Jie pian* 蒼頡篇, newly discovered among various fragments of Han period (206 BCE-220 CE) wood- and bamboo-strip manuscripts (Foster 2017; Fukuda 2004; Liang 2015). To continue the testing of our framework, I have compiled a dataset that expands upon this example by including the entirety of the *Cang Jie pian*’s content, as it is found on five of our major manuscript witnesses. This dataset is now available on Zenodo at: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. It cleans, updates and therefore supplants the data previously distributed for the *Cang Jie pian* on PoePY.

The *Cang Jie pian* 蒼頡篇 was an important scribal treatise (*shishu* 史書) in early imperial China. Through mastery of this text, students were able to attain coveted government positions. Although the *Cang Jie pian* failed to be transmitted among our received corpus of ancient Chinese texts, manuscript discoveries over the past century shed new light on the nature of its content. Indeed, the *Cang Jie pian* is ubiquitous among recently unearthed caches of Han period manuscripts, extant now in sixteen different collections. Furthermore, it appears in diverse archaeological contexts, from an aristocratic burial in Anhui, to the military installations in Gansu, and even among the remains of the Jingjue 精絕 kingdom in Niya 尼雅, Xinjiang. The Han dynasty witnessed the construction of China’s first enduring empire. Accompanying this political reform were linguistic changes and shifting norms in the manuscript culture. The central role played by the *Cang Jie pian* in scribal education, alongside its widespread distribution during the Han dynasty, recommend itas a potentially fruitful source for research into the standardization of Chinese language and writing at this time.

“CJP Rhyming Data” is offered as a first step towards the systematic analysis of the language of the *Cang Jie pian*, to aid in the reconstruction of Han Chinese and the articulation of linguistic changes that occurred during this period. In what follows is any explanation of the information included in this dataset and its presentation. The first sheet, “CJP Rhyming Data,” presents a sortable table filled in with the pertinent information from our manuscript sources. Additional explanations for the columns may be found below. The second and third sheets, “Rhyme ID Index” and “Pre-VT & VT Line Index” respectively, help the user to locate and compare data across rhyme IDs, editions and stanzas. The final sheet, “Bibliography,” provides standard references for the primary source publications and the secondary scholarship cited in the previous sheets.

In ancient China, texts were often written on scrolls of bound strips made from wood or bamboo strips (Tsien 2013). As mentioned before, content from the *Cang Jie pian* is prolific among recent manuscript discoveries. There are hundreds of individual strips or fragmented strip-pieces which relate to this text. Accounting for the entirety of this data will take a significant amount of time, and may necessitate multiple versions for “CJP Rhyming Data.” The eventual goal is to include \*all\* manuscript evidence in the database, along with every contended transcription and interpretation proposed by paleographers, to support a broader apparatus for the study of the *Cang Jie pian*. For the time being, however, “CJP Rhyming Data” prioritizes data relevant only to the structural rhymes around which the text is based.

The *Cang Jie pian* is a tightly structured text and is organized around rhymes. Every line is four characters in length, with a rhyme position falling at the conclusion of every second line (e.g., every eighth character). Each chapter, moreover, participates in a single overarching rhyme scheme. Knowing these rules greatly eases our adjudication of where rhyming positions should occur. “CJP Rhyming Data” presents these structural rhymes, and the variants found in those positions. It is, of course, possible that irregular internal rhyming exists in the *Cang Jie pian* as well, or that other interesting linguistic phenomena, such as alliteration, are present. While “CJP Rhyming Data” is not designed to highlight these features, it may be of service in their discovery and eventual analysis.

For our purposes here, we differentiate between three editions of the *Cang Jie pian*:

1. An early version in which chapters vary in length, but contain over 100 characters. This will be called the “Pre-VT” edition. Witnesses to the Pre-VT edition include the Peking University *Cang Jie pian* (Beijing daxue chutu wenxian yanjiusuo 2015; abbreviated PKU), and the Fuyang Shuanggudui 阜陽雙古堆 *Cang Jie pian* (Hu and Han 1983; Zhongguo jiandu jicheng bianji weiyuanhui 2001+; abbreviated FY).
2. A later version, said to have been edited by “village teachers 閭里書師,” that divides the content into 60 character chapters. This will be called the “VT” edition. The main witness to the VT edition is the so-called “Han board” *Cang Jie pian* (Liu 2019; abbreviated HB), for which we will have further comment below. Also included is JYX EPT 50.1, from among the “new Juyan strips 居延新簡” (Zhang 2016; cache abbreviated JYX). This single bamboo strip writes out a nearly complete version of the “opening chapter” to the *Cang Jie pian*.
3. An edition based on the VT text, but which adds rhyming commentary to each base line. This will be called the “SQZ” edition, after the only manuscript witness, known as the Shuiquanzi 水泉子 *Cang Jie pian* (Zhang 2015; abbreviated SQZ). Because the SQZ *Cang Jie pian* supplies additional testimony for the VT edition, it is entered into the database twice, as evidence for both the VT and SQZ editions.

These are our main sources, and each has been documented in full in “CJP Rhyming Database.” Other manuscripts are included, but only when they bear information relevant to the structural rhymes. The above categorization of our manuscript sources into three *Cang Jie pian* editions is speculative and meant primarily as an heuristic guide. Note, for example, the conflict between the PKU and FY mss in the 漢兼 stanza (lines 3-6), which may be edition-level variation; despite this conflict, we retain both as Pre-VT sources. Furthermore, due to the fragmented state of many of our *Cang Jie pian* manuscripts, it is not always feasible to determine from which edition their content derives. As a general working hypothesis, allmanuscripts besides PKU and FY are treated as VT sources.

Comparing the Pre-VT and VT editions of the *Cang Jie pian*, it appears that the VT edition rather mechanically divided the longer chapters of its predecessor into shorter 60-character long segments, without significant further alteration to the content (Foster forthcoming). In other words, the content and line order of Pre-VT and VT largely coincide, even though there are alternative divisions for larger textual units (i.e., chapters). For this reason, despite the fragmentary nature of our sources, we can propose reconstructions that draw across manuscripts witnesses and editions with some measure of confidence. For instance, the placement of FY C046 before PKU 1 in the Pre-VT edition of the text is justified in part because of the evidence found in the VT edition, which on HB 3 has parallel content to that of FY C046 and PKU 1 written consecutively together. A more extensive discussion of the textual history of the *Cang Jie pian* and the methodologies that have been employed in its reconstruction may be found in Foster 2017. Of course, a degree of caution still is warranted when drawing across different manuscript witnesses to reconstruct hypothetical base text and line breaks. (See the note to Pre-VT ?#2 line a, PKU 65, or to SQZ VT 3 line a, SQZ C072, for examples).

The columns employed in the “CJP Rhyming Data” table follow those proposed in List et al. 2019 (see especially p.31). Further specifications pertinent to this particular case study are as follows. These modifications highlight some of the issues and refinements needed to work with a complex manuscript corpus:

ID =

Every entry is given a unique ID, allowing for convenient location of data during discussions. If “CJP Rhyming Data” is updated in the future, this will impact the numbering of the IDs. It is therefore advised to include version information when citing IDs. The version current for this documentation is v.3.2 (7 October 2020).

POEM =

CJP stands for *Cang Jie pian* 蒼頡篇. Non-CJP is listed for materials that do not belong to the text proper, or for which there exists significant doubt. This includes a mirror inscription, that quotes a line from the *Cang Jie pian*, but has altered the content to fit its own unique rhyme scheme. Another example is content from the sexagenary cycle, found in the SQZ cache, which Zhang Cunliang 張存良 argues belongs to the *Cang Jie pian*, but this is suspect.

EDITION =

This column differentiates between various editions of CJP, as outlined above: Pre-VT for the edition prior to the village teachers’ editing; VT for the version produced by the village teachers; and SQZ for the edition based on VT that also appends a three-character rhyming commentary. Mirror and 干支 are given for the mirror inscription and sexagenary cycle respectively.

STANZA =

Chapters are taken as the “stanzas” for the *Cang Jie pian.* Each chapter participates in a single overarching rhyme scheme. Often times, however, these rhyme schemes continue across chapters as well. On this, please see the explanation for the “Rhyme ID Index” sheet under RHYMEIDS. None of our manuscripts offer a complete text of the *Cang Jie pian*, making reconstruction necessary when determining units of textual division.

For the Pre-VT edition, our longest and most complete witness is the PKU ms. It explicitly titles a number of chapters, and at times also records character counts summarizing chapter lengths. The structure for the Pre-VT *Cang Jie pian* therefore is derived primarily from the PKU ms. When a chapter title is written on the PKU ms, this is used for the stanza name (e.g., 漢兼). When a partial title appears on the PKU ms, if the missing character can be supplemented by comparison to other sources, this is added inside [] (e.g., [賞]祿); if the missing character cannot be supplemented, it is left blank with a □ (e.g., □輪). When a title is not extant on the PKU ms, but is explicitly mentioned in other sources, this is also given within [] (e.g., [爰歷], a title mentioned in the *Hanshu* 漢書 and *Shuowen jiezi* 說文解字). On a few occasions, a title both missing in the PKU ms and not mentioned in other sources may be suggested based on a comparison to the VT edition and our knowledge of the title conventions governing the PKU ms. This is signaled by placing the title within [] and adding an \* beforehand (e.g., [\*窒竅]).

There are nine chapters in the Pre-VT edition for which titles are missing completely. These are labelled “Pre-VT ?#1,” “Pre-VT ?#2,” etc. At times, content found on Pre-VT sources (namely the PKU and FY mss) cannot be located definitively within a known Pre-VT chapter. In such cases, the content is assigned to “Pre-VT Unknown” and then consecutively numbered. If we can hypothesize possible chapters to which the content may belong, these options are given in parenthesis: “Pre-VT Unknown #1 (爰歷, ?#4, or 機杼),” “Pre-VT Unknown #2 (爰歷, ?#4, or 機杼),” and “Pre-VT Unknown #3 (齎購 or ?#5).” If not, it is left with just the number: “Pre-VT Unknown #4,” “Pre-VT Unknown #5,” etc.

For the VT edition, the *Hanshu* and *Shuowen jiezi* argue that village teachers divided the text into 55 chapters, each 60 characters in length. Based on this description, stanzas are labelled as VT 1, VT 2, VT 3… VT 55. Our longest and most complete witness is the HB ms. This manuscript consists of wooden boards, which each bear 60-characters of text written in three columns. Our assumption is that a single board corresponds to one chapter of the VT edition; this is supported by a comparison of the textual divisions across the HB board to our prior understanding of VT chapter divisions (see for instance JY 9.1, and the discussion in Foster forthcoming). On the top of the HB boards, a numerical label is written (e.g., 第一 or “1st”). This presumably numbers the VT chapter for the content on the board, and serves as our overall guide for placing content in a given chapter. Note however that these labels are often difficult or impossible to discern in the published photographs, and the proposed transcriptions given by Liu Huan 劉桓 can be erroneous. We judiciously adopt different arrangements for the VT edition, proposed by other scholars (e.g., the board Liu labels as 10 is treated as content for VT 20; the one Liu labels as 53乙 is treated as VT 55). With further research, our arrangement is liable to change.

Again there are times when certain content found on VT sources (namely the HB and SQZ mss) cannot be located definitively within a known VT chapter. In such cases, the content is assigned to “VT Unknown” and then consecutively numbered. If we can hypothesize possible chapters to which the content may belong, these options are given in parenthesis: “VT Unknown #1 (22, 28 or 32),” “Pre-VT Unknown #2 (22, 28 or 32),” etc. If not, it is left with just the number: “VT Unknown #4,” “VT Unknown #5,” etc. One unique situation requires explanation: the content on HB 42 likely connects precedes the content on HB 43甲, but we cannot determine if this pair corresponds to VT chapters 41+42 or VT chapters 43+44. We therefore title both chapters as #3, but further designate them as “a” versus “b” to communicate relative order: “VT Unknown #3a (41 or 43),” and “VT Unknown #3b (42 or 44).”

For the SQZ edition, because the base text follows the VT edition, chapter titling likewise will correspond to that of the VT edition: VT 1, VT 2, VT 3, etc. The only exceptions are when unknown content on the SQZ ms is too fragmentary to propose line breaks. In such cases, because it is uncertain if the content is from the VT base text or the SQZ commentary, the chapter is labelled “SQZ Unknown” and consecutively numbered: “SQZ Unknown #1,” “SQZ Unknown #2,” etc.

LINE\_IN\_SOURCE =

This gives the text found on the manuscript cited in the corresponding SOURCE column. No paratextual features are included, such as chapter titles or character counts. Any punctuation on the manuscript is removed as well. Unless otherwise noted, the transcriptions given in “CJPRhyming Data” follow those given in: Beijing daxue chutu wenxian yanjiusuo 2015 (for PKU), Hu and Han 1983 (for FY), Liu 2019 (for HB), Zhang 2016 (for JYX EPT 50.1), and Zhang 2015 (for SQZ). Future versions of “CJP Rhyming Data” will take into account *all* proposed transcriptions for each word, as debated in the scholarly literature. Our preference is to record the strict transcriptions, not interpretative, especially for the rhyme words. Strict transcriptions do not communicate when scribal errors, alternative forms, or loaning effect our reading. When this information potentially impacts the pronunciation of a word in a rhyming position, it is discussed in the “NOTES” column. For content outside of rhyming positions, no comment is given.

Occasionally the published transcription for a character is difficult or impossible to type. If the character falls outside a structural rhyme, we revert to the interpretative readings given in the publication’s annotations. For example, the second character on PKU 2 is given a strict transcription of ![A picture containing light, sitting, traffic, city

Description automatically generated](). The Peking University editors identify this as the word 勦, which is what is then used for “CJP Rhyming Data” (see ID 119). If no interpretative reading exists, or if it is important to retain the spelling of the strict transcription (especially for rhyme words), we describe the character with ⿰ symbols and regular *kaiti* 楷體 form components. Thus ⿰歹易 spells the rhyme word for the base text of VT 12 line 7 on SQZ C021, even though this is read by Zhang Cunliang as 殤 (see ID 452). This can be used to describe partial characters as well. For instance, FY C066 has ⿰□翏 as the partial remains of the rhyme word for Pre-VT ?#7, line n (see ID 1336).

LINE\_ORDER =

As described in List et al. 2019, LINE\_ORDER is “A numerical value that provides the order of the lines of a poem in a given stanza.” For “CJP Rhyming Data,” numerical values designate absolute line number: 1 is the first line of the chapter, 2 is the second, and so forth. When only relative line order can be determined, this is represented by consecutive letters in the alphabet: a, b, c, etc. Consider for instance PKU 45. We may speculate that it belongs to the chapter Pre-VT ?#6. A character count of 144 is found after this final line on PKU 45, which moreover tells us that it was the thirty-sixth line of the chapter. Thus the LINE\_ORDER value is 36 (see ID 1252). With PKU 14, while it is likely that the content on this strip belongs to Pre-VT ?#8, we do not know where precisely it fits within this chapter. The five lines written on PKU 14 and therefore labelled a-e for LINE\_ORDER, as only their relative order is secure (see IDs 1373, 1375, 1377, 1380, and 1383).

We can exploit LINE\_ORDER to document variants found on other manuscripts (List et al. 2019, p.39). This is accomplished by designating the same LINE\_ORDER to multiple entries under the same STANZA, with each given a different SOURCE. For example, the database gives five different sources for the rhyme word on VT 1 line 2, showing 嗣, 子 and 生 as variants (see IDs 6-11). This allows for direct comparison of variants, but only from the same edition (e.g., Pre-VT versus VT and SQZ). Often it is important compare parallel text across editions. For example, Pre-VT 顓頊 line 4 is only attested on PKU 46, where the rhyme word is 襄; but there is parallel text for this line in VT 11 line 4, found on HB 11乙, which gives the rhyme word as 鑲. To help preserve these relationships, two approaches have been adopted. First, “CJP Rhyming Data” lists parallel or connected content close to one another on the sheet, allowing for easy visual reference and grouping together IDs for like material. “Pre-VT & VT Line Index” directs the user to where a given Pre-VT or VT chapter begins and ends relative to the other edition, offering a rough guide for where to consult when looking for content across editions. Second, the numbering of RHYME IDs also suggests content relationships.

RHYMEIDS =

Note that RHYMEIDS is different from ID. RHYMEIDS describes where in a given line a rhyme word is present. For each character position in the line, a numerical digit is assigned. When the word is not part of a rhyming relationship, it is assigned a 0. For example, the RHYMEIDS for the four characters in VT 1 line 1, 蒼頡作書, is 0 0 0 0, telling us that none of the words are rhyming (see ID 2). When the word is part of a rhyming relationship, it is assigned a number (1, 2, 3…), with each word participating in that rhyme scheme sharing in the same number. Thus VT 1 line 2, 以教後嗣, has 0 0 0 1 for the base text, which describes the fourth word, 嗣, as rhyming (see ID 6). Compared to VT 1 line 4, 謹慎敬戒, which also has 0 0 0 1 (see ID 16). This means that the fourth word of this line, 戒, is a rhyme word; furthermore, it participates in the same rhyme scheme as 嗣, since both have the Rhyme ID of 1. If it is uncertain which words participate in a rhyme scheme, we write a question mark instead of a numeral. Take for instance VT Unknown #31, line b, 被衾襖絝, found on SQZ 104 (see ID 1717). Because we do not know if 絝 is a rhyme word, the RHYMEID for the line appears as: 0 0 0 ?.

Because the Pre-VT, VT, and SQZ editions of the *Cang Jie pian* derive from similar content and include parallel text, often rhyme schemes are repeated or extended through comparison across different editions. We have attempted to preserve this information in the logic behind assigning Rhyme IDs. VT edition rhyme schemes run from 1-55 (representing each of its fifty-five chapters). For VT content that rhymes, but for which we do not know the corresponding chapter, Rhyme IDs then count up from 56.

The Pre-VT edition of the *Cang Jie* *pian* has longer chapters, which often incorporate multiple VT chapter rhymes schemes per single Pre-VT chapter. For example, the content of the Pre-VT 顓頊 chapter parallels that of VT 11-13. To help communicate this association, Pre-VT Rhyme IDs add a 0 to the Rhyme ID for that of the first corresponding VT chapter. Because the Pre-VT 顓頊 chapter begins with parallel content to VT 11, its Rhyme ID is 110. Note that a 0 is added in front of Pre-VT Rhyme IDs which parallel VT 1-9 content: e.g., Pre-VT [賞]祿 has 030, because it begins with content seen on VT 3 as well. For Pre-VT content that rhymes, but for which we do not know the corresponding chapter, Rhyme IDs then count up from 1000. Note that the relationship between VT Rhyme ID 10 and Pre-VT Rhyme ID 100 is unique, in that scholars have proposed different line breaks for the parallel content found here; this means that these two sections are textually related but potentially offer conflicting rhyme schemes. This is a problematic section of the *Cang Jie pian* which I will treat in more depth in a future article.

The SQZ ms is based on the VT edition, but adds three-character rhyming commentary to the four-character base text. This creates two connected yet separate rhymes schemes, given as R:1 and R:2, for the base text and commentary respectively. Since the SQZ edition is based on the VT edition, the R:1 Rhyme IDs correspond to those of the given VT chapter. The R:2 Rhyme IDs, however, will add a 1 to the end of that same VT edition Rhyme ID. For example, the strip SQZ C052 bears content related to VT 20 line 5, but with additional commentary: 偃鼂運糧 (base text) + 載穀行 (commentary) (see IDs 791-793). Its Rhyme ID therefore is: 0 0 0 20 0 0 201. This tells us that the fourth character, 糧, is the first rhyme word (R:1) and that it participates in the rhyme scheme for VT 20; the seventh character, 行, is the second rhyme word (R:2), and participates in the rhyme scheme for the SQZ commentary to VT 20. Note that again a 0 is added in front of SQZ R:2 Rhyme IDs based on VT 1-9 content: e.g., SQZ commentary to VT 1 is 011. For SQZ content that rhymes, but for which we do not know the corresponding VT chapter, Rhyme IDs then count up from 2000.

Finally, the Han mirror inscription quotes a line from the *Cang Jie pian*, but alters the rhyme word to fit a rhyme scheme unique to its own content. Its Rhyme ID is assigned to 9999.

ALIGNMENT =

For lines with a rhyme present, the Chinese character writing the rhymed word is replaced with a reconstruction of that word’s pronunciation. This reconstruction follows William Baxter and Laurent Sagart’s 2016 Old Chinese (https://ocbaxtersagart.lsait.lsa.umich.edu). When multiple reconstructions are possible, each is recorded in NOTES. In such cases, the pronunciation given in ALIGNMENT is a preliminary judgment about the word intended in the linguistic context of the line (including the possibility of loaning), but this can be ambiguous and awaits final analysis. It should not be taken as our definitive statement on the text’s meaning.

When a reconstruction is not available for the rhyme word in Baxter and Sagart, an alternative reconstruction is found. As an expedient, this often entails substituting in a reconstruction given by Baxter and Sagart for a word that is both from the same *xiesheng* 諧聲 series and has the same Middle Chinese pronunciation. If such a substitute is not available, then the reconstruction is for the character’s phonetic component or based on some other phonetic information. This is of course methodologically problematic, and only serves as a placeholder until further analysis. If there is no or only partial evidence for the rhyme word, making a reconstruction impossible, a question mark is written instead.

R:1 and R:2 =

These columns provide a convenient way to identity the rhyme schemes present in the *Cang Jie pian* manuscripts. R:1 is based on the formulaic rhyming in CJP; R:2 is based on the formulaic rhyming in CJP SQZ. Internal rhymes or other rhyme schemes can be added in future versions with the addition of new columns (R:3, R:4, etc.). If no evidence for a word exists in a known rhyming position on our manuscript source, then we write a ? in the column. Partial evidence is documented by spelling out the orthography via ⿰ symbols.

SOURCE =

When dealing with multiple manuscript witnesses, it is necessary to reference specific sources for the content. Labels are provided for individual strips bearing representative text (recorded in LINE\_IN\_SOURCE). A key for these labels may be found at the top of the “Bibliography” sheet. If multiple strips are re-pieced together as testimony to single line, this is documented in SOURCE with a + sign. For example, the SOURCE for ID 24 is SQZ C003 + SQZ C004, since the final two characters of the commentary for this line are missing on SQZ C003, but supplemented by SQZ C004.

“CJP Rhyming Data” incorporates PKU, FY, HB, JYX EPT50.1, and SQZ in full. When variants for rhyme words appear on other sources, we add entries for them in the database. Other witnesses that write rhyme words, but do not offer a variant, are documented in the NOTES column, but not given an unique entry. We do not document strips and fragments from other sources if they lack information about rhyme words; eventually we hope to include all *Cang Jie pian* manuscript finds, but this must await a future version. Currently (September 2020), only a brief report is available for the 2018 discovery of *Cang Jie pian* material at the Chengba 城壩 site in Quxian 渠縣, Sichuan (Sichuan sheng wenwu kaogu yanjiuyuan et al. 2019). We will include this find in “CJP Rhyming Data” once the data is published in full.

It must be emphasized that the PKU and HB witnesses are purchased manuscripts. They were not secured through scientific archaeological excavation, and lack proper provenience. Foster 2017b argues for the authenticity of the PKU ms. In the fall of 2019, it was announced that a private collector possessed another *Cang Jie pian* manuscript, written across numbered wooden boards, labelled as the HB witness in “CJY Rhyming Data” (Liu 2019). HB has not yet been properly authenticated, although a number of scholars assert that it is indeed a genuine artifact (Bai 2019, Fukuda 2020, Zhuang 2019, etc.). The data for HB was published in Liu 2019, but unfortunately Zhonghua shuju 中華書局 has recalled the book and it is no longer available for purchase. We have decided to include the HB ms in “CJP Rhyming Data” in part to make transcriptions available for interested scholars. Due caution is still warranted, however, when using this source for our research.

SOURCE\_COLLECTION:

Finally, we have added in another column called SOURCE\_COLLECTION to enable users to filter between different manuscript sources (FY, PKU, HB, SQZ, JYX, etc.). In this way, a user may see the entire content from a single manuscript. This is necessary especially for multi-piece manuscripts, where the documentation in SOURCE alone would give the user only one piece of a larger manuscript. EDITION is likewise unsuitable as filter, as it usually will incorporate multiple manuscript witnesses.

“CJP Rhyming Data” is an initial attempt to test our standardized rhyming framework on a larger and complex corpus of manuscript data. Already, in the input of this data, certain modification have been made, anticipating users’ needs. The inclusion of EDITION, SOURCE and SOURCE\_COLLECTION offer an example, where we attempt to preserve textual and material relationships in the dataset that are difficult to express in the prior framework. Inevitably, with “CJP Rhyming Data” now compiled, unforeseen user’s needs and missing functionality will be laid bare. To this end, we treat this case study as an ongoing project and anticipate continuous updates to the dataset. Documenting these changes, and their reasons,

References

Baxter, William H., and Laurent Sagart. *Old Chinese: A New Reconstruction*. New York: Oxford University Press, 2016.

Beijing daxue chutu wenxian yanjiusuo 北京大學出土文獻研究所, ed. *Beijing daxue cang Xi Han zhushu* (*yi*) 北京大學藏西漢竹書(壹). Shanghai: Shanghai guji chubanshe, 2015. (PKU).

Foster, Christopher J. "Study of the *Cang Jie pian*: Past and Present." PhD Dissertation, Harvard University, 2017.

Foster, Christopher J. “Introduction to the Peking University Han Bamboo Strips: On the Authentication and Study of Purchased Manuscripts.” *Early China* 40 (2017): 167-239.

Foster, Christopher J. "The Shape of the Text: *Gu* Prisms and Han Primers." In Susan Blader, Constance Cook, and Christopher J. Foster, eds. *Thinking About Early China with Sarah Allan*, forthcoming.

Fukuda Tetsuyuki. *Setsubun izen shogakusho no kenkyo* 說文以前小学書の研究􀈺􀉏. Tokyo: Sobunsha, 2004.

Hu Pingsheng 胡平生 and Han Ziqiang 韓自強. "*Cang Jie pian* de chubu yanjiu 蒼頡篇的初步研究." *Wenwu* 文物 2 (1983): 35-40. (FY).

Jiandu zhengli xiaozu 簡牘整理小組, ed. *Juyan Han jian* 居延漢簡. 4 vols. Taipei: Zhongying yanjiuyuan lishi yanjiusuo, 2014-17. (JY)

Liang Jing 梁靜. Chutu Cangjie pian yanjiu 出土蒼頡篇研究. Beijing Kexue, 2015.

List, Johann-Mattis, Nathan. W. Hill, and Christopher J. Foster. “Towards a Standardized Annotation of Rhyme Judgments in Chinese Historical Phonology (and Beyond).” Journal of Language Relationship 17 (2019): 26-43, https://doi.org/10.31826/jlr-2019-171-207.

Liu Huan 劉桓. *Xinjian Han du Cang Jie pian Shi pian jiaoshi* 新簡漢牘蒼頡篇史篇校釋. Beijing: Zhonghua shuju, 2019.

Schuessler, Axel. *Minimal Old Chinese and Later Han Chinese: A Companion to* Grammata Serica Recensa*.* ABC Chinese Dictionary. Honolulu: University of Hawai’i Press, 2009.

Sichuan sheng wenwu kaogu yanjiuyuan 四川省文物考古研究院 and Quxian lishi bowuguan 渠縣歷史博物館, ed. “Sichuan Quxian Chengba yizhi 四川渠縣城壩遺址.” *Kaogu* 考古 7 (2019): 60-76.

Tsien, Tsuen-hsuin. *Written on Bamboo & Silk: The Beginnings of Chinese Books & Inscriptions*. 2nd ed., with afterword by Edward Shaughnessy. Chicago: University of Chicago, 2013.

Zhang Chuanguan 張傳官. "Tantan xinjian mudu *Cang Jie pian* de xueshu jiazhi 談談新⾒⽊牘蒼頡篇的學術價值." Fudan University Center for Research on Chinese Excavated Classics and Paleography 復旦大學出土文獻與古文字研究中心 Article Database. 25 December 2019, http://www.gwz.fudan.edu.cn/Web/Show/4510.

Zhang Cunliang 張存良. "Shuiquanzi Han jian Cang Jie pian zhengli yu yanjiu 水泉子漢簡蒼頡篇整理與研究." PhD Dissertation, Lanzhou University, 2015.

Zhang Defang 張德芳, et al. *Juyan xin jian jishi* 居延新簡集釋. Lanzhou: Gansu wenhua chubanshe, 2016, 7 vols. (JYX).

Zhongguo jiandu jicheng bianji weiyuanhui 中國簡牘集成編輯委員會, ed. *Zhongguo jiandu jicheng* 中國簡牘集成. Lanzhou: Dunhuang wenyi chubanshe, 2001+, vol.14 & 18. (FY).