{Title}

Introduction

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 (Foster 2017). 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, appearing 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 (206 BCE-220 CE) witnessed the construction of China’s first enduring empire. Accompanying this political reform were linguistic changes and shifting manuscript culture norms. 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 subject 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. The data is presented using the annotation framework established by List, Hill, & Foster 2019, in order to facilitate access and allow for analysis with computational tools. 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 sheet, “Bibliography,” provides standard references for the primary source publications and the secondary scholarship cited in the previous table.

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.” Anticipating the need to gradually update data over time, this corpus likewise offers an ideal case study for testing methodologies on database construction. In particular, we will use “CJP Rhyming Data” to show… {MATTIS CONTENT}

--- --- ---

Documentation for “CJP Rhyming Data”

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” only presents these structural rhymes, and the variants found in those positions. It is possible that irregular internal rhyming exists in the *Cang Jie pian* as well. Proper analysis of this will require a listing of all variants in the text (e.g., JYX EPT 50.1 風 -> 諷 in 諷誦), which must await for a future version of the table.

For the first version of “CJP Rhyming Data” (v1), two manuscripts are taken to serve as the base sources for organizing the data. The first is JYX EPT 50.1, from among the “new Juyan strips 居延新簡” case. This single bamboo strip writes out a nearly complete version of the so-called “opening chapter” to the *Cang Jie pian* (Zhang 2016). The second is the Peking University *Cang Jie pian*, as published in *Beijing daxue cang Xi Han zhushu yi* 北京大學藏西漢竹書壹 (Beijing daxue chutu wenxian yanjiusuo 2015; abbreviated PKU). These two manuscripts cover a large proportion of the known content of the *Cang Jie pian*, but are not exhaustive.

For the second version of “CJP Rhyming Data” (v2), data from a third manuscript, known as the Shuiquanzi 水泉子 *Cang Jie pian,* is added to the table (Zhang 2015; abbreviated SQZ). This manuscript is important for two reasons. On the one hand, it preserves some additional content from the base-lines of the *Cang Jie pian*, and at times can extend that base text. On the other hand, this manuscript is unique, in that it appends three-character “extensions of meaning” (referred to for convenience as “commentary”) to the four-character base-lines of the *Cang Jie pian*, forming seven-character long lines. Each of these seven-character lines then also rhyme, establishing as a second rhyme group. The rhymes from the base text of the *Cang Jie pian* are thus listed in the R:1 column, those from the Shuiquanzi commentary are given in the R:2 column.

The JYX EPT 50.1, PKU and SQZ mss are each incomplete artifacts. Only the PKU “Zhuan Xu 顓頊 chapter” (strips #46-52) offers an intact chapter. Every other chapter in the JYX EPT 50.1, PKU and SQZ mss includes some measure of textual loss. At times, these gaps can be supplemented by other manuscript witnesses. These supplements are made in the “LINE” column, and explained in the “NOTES” column. For example, the final line on PKU 1 reads “□□獲得,” with the first two characters missing. These characters may be supplemented by the Fuyang Shuanggudui *Cang Jie pian* witness, strip C009, which has “益就獲得” (Hu & Han 1983; Zhongguo jiandu jicheng bianji weiyuanhui 2001+; abbreviated FY). In “CJP Rhyming Data,” we give the “LINE\_IN\_SOURCE” as “□□獲得” and the “LINE” as “益就獲得.” Since some of the content for the line is still found on PKU 1, it remains the representative source in “SOURCE.” All of this is documented in “NOTES.”

If the supplemented text extends beyond lines found in the PKU or JYX EPT 50.1 mss, then a representative witness is cited in the “SOURCE” column. The extension of *Cang Jie pian* content can be speculative, particularly when it relies on chaining together a number of short fragments. Caution is advised in such instances. For example, before PKU 1 we may surmise that there was a line written “高囂平夷” now lost on the PKU ms, but found on other manuscripts. Because none of this content is present in the PKU ms, we use FY C046 as a representative base source. FY C046 is then given in the “SOURCE” column.

Only the content for lines of the base text and SQZ commentary are recorded in “CJP Rhyming Data.” Paratextual features, such as the titles and character counts seen on the PKU ms, have been removed. As a primer, the *Cang Jie pian* was studied and often used in practice writing. The PKU ms is a model text, written in a careful and legible hand, and faithfully formatted. This is not always the case on other manuscripts. Numerous strips or strip fragments have been discovered where content from the *Cang Jie pian* is scribbled out in disjointed lines, mixed with other content, or full of repeated characters and mistaken orthography. Although JYX EPT 50.1 is composed in a confident hand, it is unclear if it is a model text, or was the result of practice writing. The content at the very end appears to be unrelated to the *Cang Jie pian*. It has thus been removed as well.

For our purposes here, we may differentiate between three editions of the *Cang Jie pian*: (1) An early version in which chapters vary in length, but contain over 100 characters; (2) a later “Village Teachers” version in which chapters have precisely 60 characters in length; and (3) the Shuiquanzi edition which adds commentary to the “Village Teachers” edition. The PKU ms is of the first variety. We suspect that the JYX EPT 50.1 ms belongs to the second variety. The SQZ ms accounts for the third. Due to the fragmented state of many of our *Cang Jie pian* manuscripts, it is not always feasible to determine from which edition the content derives. Since edition-level variation in content is possible, a degree of caution is warranted when drawing across different manuscript witnesses to supplement or reconstruct hypothetical base text. (See especially PKU 65 for this issue, which will be discussed in more detail shortly).

Comparing these editions, however, it appears that the text of the Village Teachers’ edition rather mechanically divides the longer chapters of its predecessor into shorter 60-character long segments, without significant further alteration to the content (Foster forthcoming). For this reason, despite the fragmentary nature of our sources, we can propose supplements across manuscript witnesses with some confidence. 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.

The transcriptions given in “CJPRhyming Data” follow those given in Zhang 2016 (for JYX EPT 50.1), Beijing daxue chutu wenxian yanjiusuo 2015 (for PKU), 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. Be aware that these are strict transcriptions, not interpretative. For example, PKU 1 writes 㝡, but this is best read as 冣 instead. 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.

Similarly, there are times when the published transcription for a character is difficult or impossible to type. In these cases, we revert to the more 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]() , which is not yet encoded. The Peking University editors identify this as the word 勦, which is what is then used for “CJP Rhyming Data.” Alternatively, if it is important to retain the spelling of the strict transcription, or when the writing is only partial, we describe the character with ⿰ symbols and regular *kaiti* 楷體 form components, but try to relegate this to the “NOTES” column. These details are discussed only when they pertain to rhyming positions, otherwise once again no comment is given.

The columns employed in the “CJP Rhyming Data” table follow those proposed in List et al. 2019 (see especially p.31), with the addition of RHYME, RHYME\_WORD, SOURCE, and NOTES. Further specifications pertinent to this particular case study are as follows:

ID =

Every entry is given a unique ID, allowing for convenient location of data during discussions. The IDs, however, change across different versions of “CJP Rhyming Data.” It is therefore important to include version information when mentioning IDs. On this, see also STANZA below.

POEM =

In “CJP Rhyming Data” (v1 and v2), CJP stands for *Cang Jie pian* 蒼頡篇.

EDITION =

This column differentiates between different editions of CJP. In “CJP Rhyming Data” (v1), there is VT for the Village Teachers’ edition, and PKU for the Peking University manuscript. In “CJP Rhyming Data” (v2), SQZ is added for the Shuiquanzi manuscript edition, which appends a three-character rhyming commentary to each four-character base line.

STANZA =

None of our manuscripts offer complete texts, and (with a single exception) each stanza bears some loss or demands reconstruction. VT signals the Village Teachers edition. JYX EPT 50.1 is probably the “opening chapter” of this edition. The base text of CJP on PKU ms, however, is from an earlier edition with longer chapters than the VT editions. On the PKU ms, the titles for some of these chapters are known in full or in part; for others, we can only give a range of strips which the editors have identified as a cohesive unit. When titles are known, the chapter is labelled with that title, followed by the PKU strip range in parenthesis: 顓頊 (PKU46-52). When the title cannot be determined, only the strip range is given: PKU14-15.

With “CJP Rhyming Data” v2, the Shuiquanzi manuscript is added. This witness employs the VT edition of the CJP, which has a shorter chapter length and at times different chapter divisions than the earlier PKU edition. When content overlaps with the PKU ms, but the specific VT chapter number is unknown, we use capitalized letters to differentiate stanzas: VT A, VT B, VT C, etc. Because the PKU CJP chapters are longer than those in the VT edition, sometimes one PKU chapter may parallel multiple VT chapters. To designate such divides, and their relative order, numerals follow after the capitalized letter. For example, the CJP base text found in the single PKU ms chapter of 顓頊 (PKU46-52) constitutes three different VT edition chapters: VT A1, VT A2, and VT A3. For novel content in the SQZ edition that does not overlap with the base text for CJP in the JYX EPT 50.1 or PKU mss, divisions are given as lowercase letters following a question mark: VT ?a, VT ?b, etc. In STANZA, all lettering is used only to signal divisions in the data, not relative order; numerals, however, do give the specific order of the chapter in the theoretical complete text.

The chapter divisions given in “CJP Rhyming Data” are based on those presented by the original editors. Modifications are justified in NOTES. These divisions are often based on content, rhyming, textual formatting, verso lines, or other features, but it is possible that alternative arrangements to the manuscript may be proposed. Similarly, future discoveries could necessitate a reorganization of the stanza delineations. These are changes that can be reflected in future version of “CJP Rhyming Data.”

Note that when the SQZ manuscript can help extend the base text for CJP given on the PKU ms, this content is labelled according to the PKU chapter convention. For now, separate VT entries for CJP base text are not provided when SQZ materials overlap with the PKU chapters; this will be part of a future update to “CJP Rhyming Data.” The only exception for now is when edition-level variation may impact the rhyme scheme. This is the case, for instance, with the following base text of CJP: …堯舜禹湯顡𠨐… On PKU 65, the line break appears to fall after 舜, but on SQZ C058 (representative of VT), the line break is instead after 湯. To document this, multiple entries with different stanza labels exist for the same CJP content. In “CJP Rhyming Data” (v.1), the labels are PKU 65 and either VT A respectively. The latter label was changed to VT F in “CJP Rhyming Data” (v.2), to accommodate the order of other VT chapters with PKU ms overlaps. This is the only change in STANZA name between the two versions of the data set.

LINE\_IN\_SOURCE =

This gives only the base text found on the manuscript cited in the corresponding SOURCE column. Supplements to that base text drawn from parallels on other manuscripts are not given.

LINE =

This gives the reconstructed text for the line, including any supplements to the base text drawn from parallels on other manuscripts. In cases where there is loss in the SOURCE manuscript, the line is hypothetical.

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.” 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, entries with the ID of 27 and 28 in “CJP Rhyming Data” v1 both belong to STANZA “賞祿 (PKU1-7),” and have the same LINE\_ORDER of 5. Yet they differ in their LINE content: “寬 + 恵 + 善 + 志” vs. “寬 + 恵 + 善 + 記” respectively. This tells us that 志 and 記 are potentially variants for the same line of CJP*,* as found on different manuscripts.

Numerals are used to designate the specific line number for given text. Owing to the fragmentary nature of these manuscripts, there are certain lines for which we may speculate a relative order, but cannot determine their specific number. To reflect this, we use lowercase lettering.

RHYMEIDS =

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. 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. Because data entry entailed returning to various parts of the CJP and CJP SQZ texts, the numbering is not always consecutive.

[SHOULD I CREATE A RHYMEID INDEX?]

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 rhymed word in Baxter and Sagart, an alternative reconstruction is found, often for its phonetic component. This is of course 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 provides a convenient identity for the rhyme schemes present in the CJP and CJP SQZ 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 exists for a known rhyming position, then no character is written 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 (found in LINE\_IN\_SOURCE). A key for these labels may be found at the top of the “Bibliography” sheet.

Finally, a brief word about sources not used. At the time that these two versions of “CJP Rhyming Data” was first compiled (June 2020), only a brief report was 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). The first volume of the Xuanquanzhi 懸泉置 finds has also been released (Gansu sheng jiandu bowuguan 2019), but we have not yet been able to consult it for additional *Cang Jie pian* pieces. Once available to us, these data will be incorporated into new versions of “CJP Rhyming Data.” In the fall of 2019, it was announced that a private collector owned another *Cang Jie pian* manuscript, written across numbered wooden boards (Liu 2019). If genuine, it is the longest witness of the *Cang Jie pian* currently extant, and it moreover clarifies the chapter divisions for the Village Teachers edition. Until the manuscript is properly authenticated, however, it will not be included in “CJP Rhyming Data.”

Initial Observations?

皮 correction

Understanding VT chapter divisions

Checking for Consistency Across Version

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. "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.

Gansu sheng jiandu bowuguan 甘肅省簡牘博物館, ed. *Xuanquan Han jian* (*yi*) 懸泉漢簡(壹). Shanghai: Zhongxi shuju, 2019.

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).

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 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).