

## Article

# Scribal Notation in Medieval Chinese Manuscripts: The *hewen* (Ligature) and the *chongwen* (Duplication) Marks

Imre Galambos | London

Early Chinese manuscripts and inscriptions often make use of two devices referred to by modern researchers as *hewen* 合文 (ligature) and *chongwen* 重文 (duplication). Both of them are signified with the same mark, comprising two small dashes which are placed below the lower right corner of the character. The mark resembles the character 二 written in a small script, similar to what we would today call a subscript.<sup>1</sup> Since the notation is identical in both cases, it is the context that determines whether it marks a joint character or a repetition.

The first examples of this notation date back to the oracle-bone records but their heyday was during the centuries BC 8th–3rd. While their use in inscriptional material up to the Han is relatively well-studied, there is almost no treatment of it with regard to paper manuscripts, especially ones from the post-Han period.<sup>2</sup> In this article, I would like to use the Chinese manuscripts from Dunhuang and Turfan to demonstrate the application of this notation during the medieval period. This has added relevance because, although the continuity of orthography and its transitions from early China to the medieval period has been fairly well researched, the secondary or peripheral aspects of writing, such as the marking of repetitions or the notation used in editing and correcting mistakes, have received little attention.

## *Hewen* (ligature)




*Hewen* is what modern researchers call a scribal device used on early manuscripts where two or more adjoining characters are united into a single composite graph. A parallel phenomenon in Western manuscript studies is the ligature, which is when ‘two consecutive letters are combined in such a manner

<sup>1</sup> Whether this notation is actually related to the character 二 is open to debate. Since such a connection is yet to be proven, I am, at this point, hesitant to make a definite identification of this mark with any particular character in the Chinese script.

<sup>2</sup> For examples of *hewen* on oracle-bone inscriptions, see Qiu 1992a and 1992b. On the same phenomenon on Warring States seals, see Wu 1989; on bronze inscriptions, see Shen 2002.

that one or both lose their normal form to a greater or lesser degree’.<sup>3</sup>

The joint graph appears in the text as a single entity and is ‘unpacked’ into its original components by the reader, who reads and pronounces it as a multisyllabic string. Strictly speaking, *hewen* is a graphical device without any direct indication of phonetic changes; it is read as the combination of its original component graphs, and is pronounced as if these were written out in full. Of course, it is also possible that, at least in some cases, the *hewen* also represented a phonetic abbreviation but we do not currently have any evidence for this.<sup>4</sup>



*Hewen* was relatively common in pre-Qin times but almost completely disappeared in later periods. It used to be marked with two short parallel strokes added below the lower right corner of the graph. Generally speaking, this device was used for characters that commonly occurred together, even if the words they represented did not form a grammatical unit. For example, the characters 之所 appear in the Houma covenant texts (ca. 490 BC) as , while the characters 之日 in the Baoshan bamboo strips (ca. 320 BC) as . In the Kongzi shilun 孔子詩論 manuscript (ca. 300 BC) in the Shanghai Museum collection, the characters 上下 are written together sharing their horizontal stroke as a single constellation of .

In each of these cases, the reader is alerted with the *hewen* mark at the lower right corner of the graph. Technically speaking, writing the two characters this way was not an abbreviation, since even if the scribe economized one stroke in the characters themselves, he still had to write two more to indicate the omission. Instead, in continuous text, it was perhaps more of an indication that these characters appeared together frequently, even if the words they stood for did not

<sup>3</sup> Lowe 1980, vol. 1:158.

<sup>4</sup> The use of ligatures in Western liturgical traditions (e.g. Church Slavic) also points to the predominantly graphical nature of this device.



Fig. 1: A fragment from the Houma covenant texts. Even within this short text, the *hewen* with the word 子孫 occurs three times, most clearly in the last line as  in the phrase 及子孫 ('down to his descendants'). Another *hewen* combination on the same fragment is the name of the city of Handan 邯鄲, appearing as .

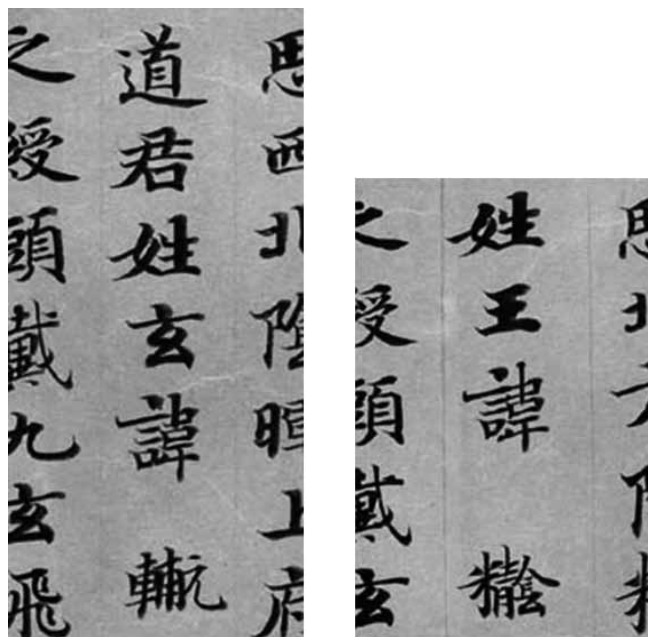






Fig. 2: Two sections from manuscript Or.8210/S.238, showing the composite characters and the empty space left before them. On the left image we can see the name 元輔 written as a single unit; and on the right one, the name 陰精

form a compound. In other words, the words written by joint characters do not always form a semantic unit and their relationship is simply that of collocations.


An interesting type of *hewen* was when one of the two original characters structurally already included the other. For example, in the Houma covenant texts (ca. 490 BC) we often see the form , which stood for the characters 子孫 (Fig. 1). From a structural point of view, this was only the character 孫 with the *hewen* sign underneath, alerting the reader that some sort of duplication was at play. Since the orthography of the character 孫 already incorporated the character 子 as a component, 孫 was enough to represent both of them. Examples of similar usage from the Warring States period are the characters 大夫 written as 夫<sub>二</sub> on seals, or 孔子 written as 孔<sub>二</sub> on so-called Chu manuscripts.

In Western Zhou and Warring States periods, when the use of *hewen* and *chongwen* was most common, the notation for both of these devices was identical: a small double-notch sign placed underneath the lower right corner of the character. In both cases, the mark indicated a doubling: either that two characters have been joined together, or that one was to be read twice.






In medieval Chinese manuscript culture, the use of *hewen* differed markedly from that seen in pre-Qin manuscripts. Although we can find a number of examples of joint characters, these always tend to have a semantic justification for being grouped together. On manuscript Or.8210/S.529, a series of letters of introduction dated from 9th-10th centuries, the name of monks Guiwen 歸文 and Dequan 德全 are joined into single graphs as  and , respectively. While the other characters in the manuscript are distinctly separated from each other, these names appear written together as single entities. The obvious reason behind writing names like this would be to treat them as a whole, lending them an emblematic quality.

Manuscript Or.8210/S.238<sup>5</sup> with a Taoist text called *Jin zhen yuguang bajingfei jing* 金真玉光八景飛經 bears a colophon dating the document to 692 AD. There are two cases of joint characters within the main text: the first one is the name Yuanfu 元輔 written as , appearing in the string 道君姓玄諱元輔. The interesting phenomenon is that an empty space stands before the composite character on the manuscript, as if indicating that, if not for the name taboo, part of the graph to follow would have actually been written in that space (Fig. 2). In other words, there is enough physical space left for 'unpacking' the joint character. The same text has also been preserved in the Taoist Canon, only there

<sup>5</sup> All manuscripts from the Stein collection in the British Library, beginning with 'Or.8210', are taken from the IDP website: <http://idp.bl.uk>.

the characters appear without observing the taboo. The second example within the same manuscript is very similar. In this case, it is the name Yinjing 陰精 in the string 道君姓王諱陰精, written as . In the new character, the character 精 is fully present but 陰 is missing its radical, thus in this case the fusion also involves an abbreviation.

It is clear from the context that, in this case, the composite character was used as a means of observing a name taboo for deceased masters. This was very similar in nature to the name taboo of imperial names during a given dynasty. The characters 世 and 民 in the Tang dynasty, for example, were routinely written without their last stroke due to the fact that they occurred in the personal name of Li Shimin 李世民, the founder of the dynasty. It was also a common practice to replace these two characters with 代 and 人 writing more or less synonymous words. In the case of the two composite characters seen in the Taoist manuscript Or.8210/S.238 above, the name taboo was observed by writing the names of the late masters together as single units, and by leaving an empty space before the joint character. Needless to say that while the joint graphs function as a *hewen*, their use and application are very unlike those seen on pre-Qin manuscripts.<sup>6</sup>

At the same time, there is a small number of words that occasionally appear in Buddhist manuscripts from Dunhuang in the form of *hewen*. The most common of these is the graph  standing for the word *pusa* 菩薩 (*bodhisattva*).<sup>7</sup> Another joint graph is for the word *puti* 菩提 (*bodhi*), written as  or . The last form here overlaps in structure with how the word *pusa* 菩薩 was abbreviated and can be distinguished only with the help of the context. A somewhat less frequent example of *hewen* in the Dunhuang material is the word *niepan* 涅槃 (*nirvāṇa*) which was sometimes written as  or .<sup>8</sup> Now it is apparent that all three examples are Buddhist technical terms and in this sense their usage is closely reminiscent of Western ligatures. It is perhaps significant that each of these three words was a transliteration of a Sanskrit term and because of this their individual component characters had no semantic significance. Another important aspect is that these forms never appear in sutras but only in non-canonical

texts, such as commentaries or transformation texts.<sup>9</sup> This shows that the *hewen* forms were not accepted as standard forms and were banned from canonical usage.

Beside the above examples, there are also the cases of the graphs 廿 (twenty), 卅 (thirty), and 卌 (forty) which were commonly used in medieval Chinese manuscripts, although not limited to them. While some researchers believe that these were read as two-syllable words in medieval times, and thus represented true cases of *hewen*, there is also evidence to the contrary. For example, the celebrated Song dynasty scholar Hong Mai 洪邁 (1123–1202) described how, in transmitted sources, the odes on the First Emperor's steles were composed in four-character units, except when a date was involved, when these would become five-character units (e.g. 維二十六年, 維二十九年, 三十有七年).<sup>10</sup> When a fragment of one of the steles was discovered, it became clear that the numbers in the dates had originally been written with the joint form (e.g. 廿有六年) and thus did not violate the tetrasyllabic principle. Of course, this also means that these joint characters were read as a single syllable, at least during the Qin, and because of this they should be considered characters in their own right, rather than *hewen* combinations.


The above cases are the types of *hewen* that occur in medieval Chinese manuscripts. An important difference from early usage is that *hewen* in Dunhuang and Turfan is never marked. Although there are many different kinds of notations for repetition, deletion or insertion of characters, it was not considered necessary to indicate composite characters in writing. The obvious reason for this was that, in contrast with the use of *hewen* in early Chinese manuscripts, combined characters in medieval practice made up meaningful units (e.g. words, names). Accordingly, there seems to be no direct evolutionary connection between the use of *hewen* in pre-Qin and in medieval times.


### Chongwen

In pre-Qin manuscripts, the *hewen* mark was identical to that of *chongwen*, but while in the first case it meant that two characters were fused into a single unit, in the second it indicated that a character or a string of characters was to be read twice. The reader had to rely on the context to determine how to interpret the mark. Unlike the case with the *hewen*, the *chongwen* device in medieval manuscripts suggests a direct connection with the pre-Qin one. In both its function and ap-

<sup>6</sup> It is also worth mentioning here the Daoist tradition of combining characters into elaborate talismans (*fu* 符). Such graphic constellations, however, are strictly speaking not part of writing habits, and shall not be considered here.

<sup>7</sup> The following examples are from Huang 2005.

<sup>8</sup> Interestingly, the late 10th century dictionary *Longkan shoujian* 龍龕手鑑, compiled on the basis of Tang-Song Buddhist manuscripts, identified these *hewen* graphs as consisting of two separate sounds. For the graph , for example, it says, 'pronounced *mang*, meaning the abundant growth of vegetation; also pronounced as the two characters *pusa*' 莫朗反, 草木多生不死也; 又音菩薩二字, revealing that it identifies the same graph both as a variant form of the character 莽, and as a *hewen* for the word *pusa*.

<sup>9</sup> A similar pattern can be observed with regard to the use of the graph , a non-standard form of the character 佛 (Buddha). This form, coinciding with the modern Japanese way of writing the same character, was never used in sutras, only in non-canonical Buddhist texts.

<sup>10</sup> Hong 1978:69–70.



Fig. 3: Section of manuscript Or.8210/S.1547 showing the use of the 𠄎 repetition mark. The characters 薪 in the first line and 滅 in the second are repeated

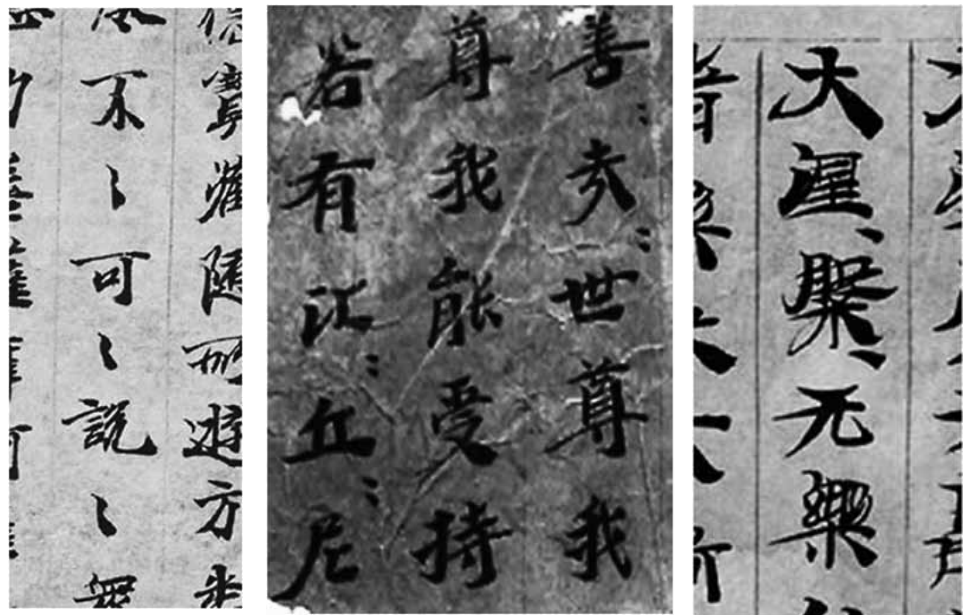


Fig. 4: Repetition marks for multi-character *chongwen*

A: Or.8210/S.2067

B: 80TBI:009

C: Or.8210/S.116

pearance, it remained practically unchanged, as abundantly manifested in the Dunhuang and Turfan corpora.

Principally speaking, there are two kinds of *chongwen*: single and multiple ones. In the first type, only one character is repeated, whereas in the latter two or more. While this may seem a trivial distinction, the notation for these in actual usage was somewhat different. The single character repetition is simply marked by a small 𠄎 mark in place of the second character. This mark was sometimes written as 𠄎 or 𠄎, and probably derived from the pre-Qin *chongwen* mark. Nevertheless, the form 𠄎 is by far the most common in Dunhuang and Turfan. An important difference between early and medieval usage was that, in the latter, the mark was placed within the main text, in place of the omitted second character. In this way, the repetition mark occupied a full character space.

Or.8210/S.1547, for example, is a manuscript of the *Chengshilun* 成實論 (\**Tattvasiddhi-śāstra*) dated to 512 AD. At the very end of the scroll, we find the following two sentences (given below in modern punctuation):

如火燒薪，薪盡則滅，是人亦爾，以不受故滅。滅三心故於一切諸苦永得解脫。

Such a man is like a fire burning the firewood: once the firewood is exhausted, it will become extinguished; this man is

also like this; because he receives no more, he becomes extinguished. If he extinguishes the three minds, he will attain eternal liberation from all sufferings.

Underlined in the translation are two cases of *chongwen*: first the character 薪 and then a bit later the character 滅. In each case, the second character is omitted and a 𠄎 mark is placed in its stead (Fig. 3). Although the characters do not form a single unit in the text grammatically (i.e. 薪薪, or 滅滅) and, in a modern punctuated transcription, are separated from each other by a comma or a period, this did not stop the medieval scribe from applying the *chongwen* device purely based on their physical adjacency.

In multi-character repetition, the *chongwen* mark is placed either underneath the character or at its lower right corner. An example of the former usage is Or.8210/S.2067 (Fig. 4/A) where the characters 不可說 ('indescribable' or 'unspeakable') are repeated in the phrase 'indescribable and indescribable myriads of sentient beings' 不可說不可說眾生. What makes this case different from the single *chongwen* seen above is that the three characters are to be read together and only then repeated as a string. At least theoretically, it would be possible to read them repeated one by one as 不不可可說說 but this would produce a meaningless string of characters. Therefore, the context is used by the reader for disambiguation.

The other way of marking multi-character *chongwen* can be seen on manuscript 80TBI:009 discovered in Turfan (Fig. 4/B), where the device is marked with a slanted double dash underneath the lower right corner of the character. The section shown on the picture contains two such cases: in the first line, we find Buddha's habitual exclamation *shanzai shanzai* 善哉善哉 ('Excellent, excellent!'); and in the third line, the words *biqiu biqiuni* 比丘比丘尼 ('monks and nuns') are written with the characters 比丘 marked as having to be read twice. This latter case is a wonderful example to show that the *chongwen* device is completely unrelated to the grammatical structure of the text and it relies solely on the physical position (i.e. adjacency) of characters. In other words, *chongwen* appears to be concerned only with characters, not words or sentences.

In addition, the doubled *chongwen* mark is sometimes written as a single slanted stroke, as can be seen in manuscript Or.8210/S.116 (Fig. 4/C), where the word *niepan* 涅槃 (*nirvāṇa*) is marked in this way. In this particular case, the repeated word occurs at the end of one sentence and the beginning of another: 'This is why it is called the Great *Nirvāṇa*. In *Nirvāṇa* there is no pleasure...' 故名大涅槃。涅槃無樂。 In the first sentence, the word is actually 'Great *Nirvāṇa*' (i.e. *Mahānirvāṇa*), thus the second use of the word is semantically not completely parallel. In this respect, this usage is similar to that of the words *biqiu biqiuni* 比丘比丘尼 ('monks and nuns').

Looking through concrete examples of *chongwen*, it is apparent that the device was optional in medieval manuscripts. Even in documents where it occurs, there are places of often identical context where it is not used and the characters are 'spelled out' in their full form. In fact, there are fewer cases where *chongwen* is used than where it is not—it is rather an exception than the norm.

## Summary

The use of *hewen* and *chongwen* devices in medieval manuscripts is interesting from the point of view of the transmission of scribal practices in Chinese history. While the *hewen* in the medieval corpus shows no similarity to Warring States usage, *chongwen* remains a common phenomenon and is marked in a similar way as it was fifteen hundred years earlier. This observation is significant because we do not have any evidence that such scribal techniques would have been taught. They are certainly absent from the linguistic treatises, dictionaries and primers describing some of the more obvious features of the script (correct character forms, variants, etc). Therefore, the analogous use of the *chongwen* device and its similar notation arguably demonstrate a direct continuity between pre-Qin and Tang-Song manuscript culture.

## REFERENCES

- Hong Mai 洪邁 (1978), *Rongzhai suibi* 容齋隨筆 (Shanghai: Shanghai guji chubanshe).
- Huang Zheng 黃征 (2005), *Dunhuang suzidian* 敦煌俗字典 (Shanghai: Shanghai jiaoyu chubanshe).
- Lowe, Elias Avery (1980), *The Beneventan script: a history of the South Italian miniscule* (Roma: Edizioni di storia e letteratura).
- Qiu Xigui 裘錫圭 (1992a), 'Jiaguwen zhong chongwen he hewen chongfu pianpang de shenglu' '甲骨文中重文和合文重复偏旁'的省略', in Qiu Xigui, *Guwenzi lunji* 古文字論集. (Beijing: Zhonghua shuju), 141–146.
- (1992b), 'Zai tan jiaguwen zhong chongwen de shenglu' '再談甲骨文中重文的省略', in Qiu Xigui, *Guwenzi lunji* 古文字論集 (Beijing: Zhonghua shuju), 147–150.
- Shanxi sheng wenwu gongzuo weiyuanhui 山西省文物工作委員會 (1976) (ed.), *Houma mengshu* 侯馬盟書 (Shanghai: Wenwu).
- Shen, Baochun 沈寶春 (2002), 'Xi Zhou jinwen chongwen xianxiang tanjiu: yi "Yinxu jinwen jicheng" guilei chongwen wei li' 西周金文重文現象探究: 以“殷周金文集成”爲類重文爲例', *Guwenzi yanjiu* 古文字研究 24: 307–11.
- Wu Zhenwu 吳振武 (1989), 'Guxi hewen kao (shiba pian)' '古璽合文考 (十八篇)', *Guwenzi yanjiu* 古文字研究 17: 268–281.
- Xinjiang Weiwu'er Zizhiqu Tulufanxue yanjiuyuan 新疆維吾爾自治區吐魯番學研究院, Wuhan daxue zhongguo san zhi jiu shiji yanjiusuo 武漢大學中國三至九世紀研究所 (2007) (eds.), *Tulufan Bozikelike shiku chutu Hanwen fojiao dianji* 吐魯番柏孜克里克石窟出土漢文佛教典籍 (Beijing: Wenwu).

## Picture credits:

- Fig. 1: Tracing from Shanxi 1976: 264.
- Fig. 2: International Dunhuang Project (<http://idp.bl.uk>), Or.8210/S.238. Copyright of the British Library.
- Fig. 3: International Dunhuang Project (<http://idp.bl.uk>), Or.8210/S.1547. Copyright of the British Library.
- Fig. 4: Images A and C are from International Dunhuang Project (<http://idp.bl.uk>), shelfmarks Or.8210/S.2067 and Or.8210/S.116. Copyright of the British Library. Image C is from the publication Xinjiang 2007: 11.