



HAL
open science

A lexical and contrastive analysis of zì 字

Françoise Bottéro

► **To cite this version:**

Françoise Bottéro. A lexical and contrastive analysis of zì 字. *Journal of Chinese Writing Systems*, 2024, 8 (2), pp.103-114. <10.1177/25138502241242809>. <hal-04872206>

HAL Id: hal-04872206

<https://hal.science/hal-04872206v1>

Submitted on 7 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A lexical and contrastive analysis of 字

Abstract

Today, one of the most essential Chinese linguistic terms related to writing, 字, is translated into English and other western languages as “character”. This paper points out the terminological and conceptual problem related to this translation as soon as one is entering the field of precise linguistic analysis. By analysing the lexical development of the term from pre-Qin documents, *Shuō wén jiě zì* dictionary, to modern times and the introduction of *cí* “word”, it shows that 字 was never limited to graphs nor graphemes. It introduces the way 字 was used to describe the Indian writing systems around the 5th century, and how it was understood outside China: in Asia, by those who borrowed the Chinese writing, and in Europe, by the first missionaries.

Keywords

graphemes, characters, written words, *Shuō wén jiě zì*, Indian writing,

Introduction

Today the terms “pictograph” and “ideograph” are *almost* definitively avoided, at least in the field of linguistic. But the rendering of 字 as “character” (which as we shall see goes back to the 16th century) is nonetheless problematic, and seems to produce some rather unfortunate effects. In this paper, I would like to point out a terminological but also conceptual problem related to the translation into western languages of one of the most essential Chinese linguistic terms related to writing : 字. I shall try to show the substantial impact such a translation can have on the linguistic analysis of the Chinese writing system. I shall not redefine 字, but call attention to the dangers of the translation of 字 as “character” as soon as one is entering the field of precise linguistic analysis.

字 are the individual graphs one can see in Chinese printed texts, but 字 are not limited to graphic signs. They also refer to monosyllabic morphemes, words or linguistic units of the Chinese language. Therefore, 字 are today basically associated with morphemes by most western linguists (Arcodia and Basciano 2021, Handel 2017, 2019, Yang-Drocourt 2022). There again, it is striking to see that some linguists present 字 as corresponding to graphemes (See Arcodia and Basciano 2021).¹

Now, the definition of grapheme varies among linguists.² Yet, graphemes, like phonemes and morphemes in their own ways, represent basic elements. If phonemes are defined as the smallest units of sound which are significant in a language, and morphemes as the smallest meaningful constituent of a linguistic expression, graphemes should then be defined as the smallest graphic units which are significant in a script.

In the Chinese writing system, graphemes then should be construed as the constitutive elements within characters the combination of which creates the large set of characters that we find explained in dictionaries.³ Graphemes are necessarily limited in number. The 26 letters of the roman alphabet can be called graphemes, so can be the *hiragana* and *katakana* in the case of the syllabic writing of Japanese. But the Chinese case is a little more complicated.

Chinese graphemes are hardly ever mentioned nor listed. This fact contributes to make people unaware of their existence and simply focus on self-evident larger graphic units labelled *zì* “characters”. That explains partially why graphemes happen to be often confused with “characters”.

The association of Chinese characters with graphemes entails implications one needs to be aware of. It implies, for example, that the Chinese writing system, which counts a huge number of characters, should contain an identical huge number of basic elements. Put this way, one can immediately see how unfortunate this association is. It also goes against the linguistic typological analysis of writing systems which distinguishes phonetic writing systems including alphabetic and syllabic writings on the one hand, and morpho-syllabic writings such as the Chinese writing system, on the other hand. The number of basic elements is usually much more bigger in morpho-syllabic writings than in phonetic ones, but it cannot reach as many as 50 thousands. Thus this impossible association does not only make Chinese look special, compared to other linguistic systems, it also completely disregards the analysis of the script, which is at stake here.

In order to understand the deep meaning of *zì* 字 and realise the problems with its translation as “character”, I shall first look at the term from a historical perspective. I shall try to see how and when *zì* started to be used in connection with the script, by looking at its different occurrences in early transmitted and excavated texts, as well as in the first dictionary of “characters”. Then, I shall look at the way the Chinese described other writing systems, such as the Indian scripts, for example, when they first encountered them and which terms they employed for that. In a third part, we shall see how *zì* 字 was understood outside China by the Japanese, the Korean and the Vietnamese who borrowed the Chinese writing system, and by the European missionaries when they first described the Chinese script, before focusing on the first Chinese grammar and the introduction of the term *cí* 詞 “word” in China at the end of the 19th century.

1. A short history of the term *zì* 字

The historical perspective is useful to analyse the way *zì* 字 was understood in pre-Qín as well as Qín and Hàn transmitted and excavated texts.

1. 1. The use of *zì* in the transmitted texts

Among the oldest occurrences of *zì* 字 in ancient transmitted texts, we find 3 different uses:

1) In the *Book of Odes Shījīng* 詩經, for example, the graph *zì* 字 refers to the word “nurture, feed” :

牛羊腓字之。 (SSJZS, p. 528) (Ode n° 245 *Shēng mǐn* 生民)
 “The oxen and sheep between their legs nurtured him.” (Trad. Karlgren, 1945, p. 71)

2) In the *Book of Documents Shūjīng* 書經, *zì* 字 refers to the word “love, cherish”, which is related to the word “nurture, feed”:

于父，不能字厥子，乃疾厥子。 (SSJZS, p. 204b) (The “Announcement to [prince] Kang” *Kāng gào* 康誥)
 “In the case of the father, he cannot cherish his son, but hates his son.”

3) *Zì* could also refer to the word “style name, public name, to give a name” in ancient texts, as we can see, for example, in the *Book of Rites Lǐjì* 禮記 or in the *Lǎozǐ* 老子
 冠而字之。敬其名也。 (SSJZS, p. p. 958c) (*Lǐjì*, *Shì guànlǐ* 士冠禮)
 “The giving of the public name after the capping is to maintain the honor of the personal name” [which the young man has received from his parents].

4) 吾不知其名，強字之曰道。 (*Dàodējīng* 道德經·25·1)

“I do not know its [proper] name. When forced to find a given name, I call it ‘the Way’.”

In this last example, the two terms *míng* 名 and *zì* 字 which both refer to “names” are in contrast with each other. Liú Zhāo 劉釗 (2005, p. 17) explains: “In ancient times, people believed that *míng* 名 was the expression of a reality, reflecting the authentic nature of this reality. And that for all the things already named, there was a good correspondence between the name and the reality, as well as a necessary link between them. On the other hand, they considered *zì* 字 as external names without any necessary link with the realities [referred to].”

1. 2. The use of *zì* in the excavated documents

The excavated documents show a slightly different picture of the use of the relevant graphs.

- The document from Chǔ 楚

In the Guōdiàn 郭店 excavated document of Chǔ, dating from the first half of the third century BC, we find almost the same sentence as that from the transmitted text of *Lǎozǐ* 老子 mentioned above. But the word “give a name, public name” it refers to is written with a different graph. The bamboo slips 21 of the *Lǎozǐ* A 老子甲 writes:

未智（知）其名，𠄎之曰道⁴

“I still do not know its [proper] name, so to give it a name, I call it ‘the Way’.”

Zhāng Shǒuzhōng (2000, p. 73) and Zhāng Xiǎnchéng (2013, p. 1547) transcribe the graph 𠄎 as 【兹/才】， whereas Bái Yùlán 白於藍 (2012, p. 23) transcribes it as 【茲/才】. But the three of them take it to be an early graphic variant of *zì* 字 “name, give a public name”. As already mentioned above, Liú Zhāo (2005, p. 17), on this occasion, insisted on the fundamental distinction between *míng* and *zì*.

In the *Tài yī shēng shuǐ* 太一生水, another document from *Guōdiàn*, the word “name, give a public name” appears again, but written differently: 𠄎 【才/心】.⁵

道亦（其）𠄎（字）也 (太一 10)

“‘The Way’ is surely its given name.”

天 𠄎（地）名 𠄎（字）並立 (太一 12)

“The name and the given name are all established.”⁶

- The document from Qín 秦

In the Qín documents from *Shuǐhǔdì* 睡虎地, dating from around the end of the third century BC (the tomb was sealed in 217 BC), the graph 𠄎 is used to refer to the word “give birth, engender” (*shēngzǐ* 生子) or “be pregnant” (*rènshēn* 妊娠).

有(又)令隸妾數字者診甲前血出及癰狀 (*Fēng zhěn shì* 封診式 86)⁷

“One should also order that a woman who had already given birth a few times (/who had the experience of being pregnant several times) examine the Jia woman’s bleeding and soar pudenda.”

女子以巳字，不復字 *Rìshū jiǎ* 日書甲種 150 正 3⁸

“A woman who gives birth on a *sì* day, will not give birth a second time.”

The graph 字 also appears in the Lǐyé 里耶 documents dating from the end of the third century BC (222-208) with a different meaning in the context of a multiplication table. The table registers 25 multiplications, from $9 \times 9 = 81$ (九九八十一) to $1 \times 0,5 = 1$ (二半而一), and provides the following final sentence:

凡千一百一十三字 (里 J1 (16) 1 正).

“All together, it makes 1 113 *zì* 字”

This example offers an interesting use of *zì* 字 because 1 113 does not correspond to the total number of words or graphs written on the tablet. It gives instead the total number of the 25 multiplications listed. Some scholars like Hú Píngshēng take *zì* 字 as an erroneous addition, while Wáng Huànlín 王煥林 considers that *zì* 字 represents here counting units, such as bamboo token used in the *chóusuàn* 籌算 calculating method.⁹ But then it is not easy to make the link with the word “name”, “written name” / “character” and even less with the word “give birth”. In fact, it is as if these 11 113 *zì* 字 represented 11 113 items or rather signs one should visualise. If that was the case, this way of totalizing the calculations would offer an interesting abstraction, and relate *zì* 字 to written signs for the first time.¹⁰

- The Hàn documents

In the *Mǎwángduī* documents of the Western Hàn dynasty, dating from around 168 BC, the graph *zì* 字 still refers to the word “give birth”:

凡治字者以清【水】幹(澣)包(胞?) (Tāi chǎn shū 胎產書, column 14)

“As a matter of principle those who help with the birth use pure water to wash?...”

字而多男毋(無)女者而欲女... (Tāi chǎn shū 胎產書, column 18)¹¹

“Those who have given birth to many boys but no girls and would like a girl...”

But *zì* 字 also refers to the word “name, give a public name” in the *Lǎozǐ* B 老子乙:¹²

吾未知(其)名也,字之曰道。

“I still do not know its [proper] name, so to give it a name, I call it ‘the Way’.”

In the Warring States, the Qín and the Hàn excavated documents, the graph 字 could also be used to record other words that are today written *zǐ* 子 or *zhī* 支, for example.¹³ Since words were widely written phonetically in the excavated documents, we should not simply stick to graphs we are used to see associated with certain words. Looking for phonetic loans *tōngjiǎ zì* 通假字 helps getting a better picture of the use of words in ancient texts as well as the way they could be written.

The excavated documents show that the word “name, give a public name” was written differently in Chǔ: 𠄎 or 𠄎, and that the Qín seem to be the first ones to employ the graph *zì* 字 as also referring to written signs.

1. 3. The first occurrence of *zì* 字 as “character”

According to the 17th century scholar Gù Yánwǔ 顧炎武 (1613-1682), the use of *zì* 字 in connection with *wén* - that is in the sense of ‘character’ - did not appear before the Qín period. Gù Yánwǔ notes in his *Rìzhī lù* 日知錄 21:

以文為字乃始於《史記》。秦始皇瑯邪臺石刻曰：「同書文字。」

“The use of *wén* with the meaning of *zì* starts with the *Shǐjì*. The *Lángyá tái* stela of the First Emperor writes: ‘Unify the “characters” *wénzì* in the documents’.”

For Gù Yánwǔ, the first concrete example we have of *wén* together with *zì* is related to the First Emperor’s intention to unify the script. But it is not clear whether the famous sentence *Tóng shū wén zì* 同書文字¹⁴ he quotes should not rather be read as ‘Unify the graphs (figures) *wén*¹⁵ and the **names** *zì* in the documents’? In other words, the idea would be to unify the way the graphs and the words they correspond to should be written, without, for example, employing too many phonetic loans.

Two kinds of evidences tend to support this last interpretation:

1) In the second century AD, Zhèng Xuán 鄭玄 (127-200) reports a lexical shift between *míng* 名 and *zì* 字 in his commentary of the *Zhōulǐ* 周禮:

古曰名今日字。“What in ancient times was called *míng* “names” is today called *zì*.¹⁶

In other words, for an author of the second century, *zì* 字 was still related to (spoken) “names”.

2) If *zì* was newly associated with *wén* from the Qín dynasty on, *yán* 言 “word” was nevertheless still largely used to refer to the number of “spoken units/ words” in a text. In the *Shǐjì*, there are many examples of written words counted in terms of *yán* 言, but very few in terms of *zì* 字.¹⁷ On the other hand, with the Western Hàn bamboo slips documents, we have clear examples of total numbers of written words counted in terms of *zì* 字. This is the case of the bamboo slips n° S002, S003, S017, and S026 of the *Shījīng* 詩經, discovered in Fùyáng 阜陽 in 1977, which can be dated prior to 165 BC.¹⁸

The fact that *yán* 言 and then *zì* 字 were used to count the number of words in a text shows that *zì* 字 “names” still preserved its oral nature around that time. But, in truth, many texts recording the number of words or written words went without adding any term.¹⁹

At the same time, it is interesting to note that we have a rare example of units constituting a text counted in terms of *wén* 文 in the Wǔwēi 武威 documents dating from around the beginning of the 1st century, whereas three other texts employ *zì*. The end of the Yánlǐ 燕禮 chapter found in Wǔwēi writes:

記三百三文

“330 graphs recorded”²⁰

The change from *yán* 言 to *zì* 字 and then sometimes to *wén* 文 suggests a gradual shift from counting words to counting written words/ graphs and thus reveals a new approach in the analysis of the language: from an spoken to a written perspective.

1. 4. *Zì* 字 in the *Shuō wén jiě zì* 說文解字

The oldest dictionary *Shuō wén jiě zì* 說文解字, written by Xǔ Shèn 許慎 in 100 AD, brings a concrete evidence of the use of *zì* as a “written word/ character” in the first century AD.

On the one hand, Xǔ Shèn 許慎 glosses *zì* in reference to the word “nurture, give breast” in the main part of his dictionary:

字乳也。從子在宀下子亦聲 (*Shuōwén* 14B 12a)

“*Zì* is to give breast. [The graphs] has *zǐ* ‘child’ under ‘roof’ as semantic constituents, *zǐ* is at the same time the phonetic constituent.”

This way of glossing is not only in accordance with the meaning of *zì* in ancient texts, as we have seen before with the *Book of Odes*, it is also in accordance with the graphic structure of *zì* which includes *zǐ* 子 “child” as a so-called semantic constituent. It is a way for Xǔ Shèn to

show the motivation at the basis of the graph and the link between the meaning of the semantic constituents and the meaning of the written word.

On the other hand, in the postface of his dictionary, Xǔ Shèn uses *zì* to refer to a completely different word, contrasting with *wén* 文, exactly like in the title of his dictionary. The analysis of all the sentences with the terms *wén* 文 and *zì* 字 in the postface shows that Xǔ Shèn employs *wén* 文 basically to refer to figures or graphs that can be explained (*shuō* 說) and *zì* 字 to refer to “names”, “words” or rather “written words” that can be sorted out or analysed (*jiě* 解).²¹ In fact, both *wén* and *zì* can refer to “characters” or “graphs”, but each seen from a different angle. *Wén* refer to graphs as visual signs or natural markings, whereas *zì* “written words” make the link between the graphs (*wén* 文) and the name or the pronunciation given to them. We can thus conveniently translate *wén* as “graphs” and *zì* as “written words” in order to represent the contrast between the two terms.

Now, the fact that Xǔ Shèn analyses *zì* 字 graphically into a certain number of semantic and phonetic constituents, through their visual figure (*wén*), shows that he is no longer simply dealing with names or spoken words, but also with their written counterparts.

The *Shuō wén jiě zì* played an important role in standardizing the shift from names/spoken words (*míng, yán, zì* 名, 言, 字) to “written names/words” *zì*, initiated by the Qín. The prestige of this dictionary which offered for the first time a systematic graphic analysis of the “written words” was such that it had a tremendous influence on scholars of later generations, even until today. It contributed to develop the analysis of the language through the written words, and to systematise the use of *zì* 字 in connection with the script as “written words”. As the comparison between the Bibliography of the Hàn and that of the Suí illustrates, after Xǔ Shèn’s dictionary, many authors started to use *zì* 字 or a combination of *wén* and *zì* 字 in the title of their work,²² attesting the importance of the new graphic perspective.

Now let’s have a look at the way the Chinese described other writing systems (such as the Indian scripts) when they first reflected on them, and which terms they employed for that.

2. The rendering of the basic elements of the Indian scripts in Chinese around the 5th C.

In order to figure out whether *zì* was ever simply understood as a character or not, it is useful to look at the way the Chinese described other writing systems.

Around the first century AD, China came into contact with Indian Buddhism, engendering a huge work of translations. It confronted Chinese scholars with a different world, while in the perspective of the script, it offered them an occasion to face a different writing system and to reflect on it.

It is not known which Indian script was first used to transmit the Buddhist scriptures, but we know from the Dunhuang manuscripts that several kinds of Indian and central Asian scripts were circulating in China.²³ The awareness of Kharoṣṭhi and Brāhmī scripts was for a long time mostly limited to Buddhist circles mentioning them occasionally in some translations.²⁴ But which terms did they use to describe the Indian writing system when they first talked about it in the 5th century?

Around the 5th century, alongside the important work of translation of the Buddhist Canon, two Chinese terms, *bànzì* 半字 and *mǎnzì* 滿字, were introduced to describe the Indian script. The first author to use these two terms is apparently the Chinese poet Xiè Língyùn 謝靈運 (385-433) who was brought up by a Buddhist monk and participated in the

translating work of the Buddhist canon. Xiè Língyùn is the author of the *Shísì yīnxùn xù* 十四音訓敘 (*Instruction to the 14 sounds*) which, according to Hirata (1994: 34 sq), was a kind of commentary of the *Sūtra of Nirvāna* meant to provide some explanations on the Indian language and script, for those who were translating the Buddhist Canon. The *Sūtra of (Mahāpari) Nirvāna, Dà bān Nièpān jīng* 大般涅槃經,²⁵ was the first text to introduce the Indian letters (Varṇamālā).²⁶ The original Sanskrit text is no longer extant except for a small number of fragments, but it survived in Chinese (and Tibetan) translations.

In the *Shísì yīnxùn xù* 十四音訓敘, Xiè Língyùn introduced for the first time the distinction between *bànzì* 半字 and *mǎnzì* 滿字. The text was lost, but it has been copied by various Buddhist monks. Hirata (1994) has conveniently gathered all the passages he could find in the Mahāyāna Buddhist Canon. One of these passages presents the two terms *bànzì* 半字 and *mǎnzì* 滿字 (Hirata, 1994: 61) in the following way:

凡小字皆曰半字[...]譬如諸字兩字合成, 名滿字。

“As a matter of principle, all the small *zì* are called *bànzì*” [...] For example the written word *zhū* 諸 ‘all’ which is a combination of two *zì* is called *mǎnzì*” (T. 84 n° 2702, ch. 5)

It is not clear what is exactly meant by these two terms, but other authors also used them. Among them, the Chinese Buddhist monk Sēngyòu 僧祐 (445-518) provides us some interesting explanation. In the chapter entitled “Notes on the differences between pronunciations and meanings in the translation of the Indian sacred texts into Chinese” 胡漢譯經音義同異記, Sēngyòu discussed for the first time the differences between the Indian and the Chinese scripts.²⁷ In this text, he explains the two expressions in terms of comparison with Chinese characters/written words:

又梵書製文有半字滿字。所以名半字者, 義未具足。故字體半偏。猶漢文月字虧其旁也。所以名滿字者, 理既究竟。故字體圓滿。猶漢文日字盈其形也。(…)
“Moreover, concerning the formation of graphs in Indian writing, there are “half written words/characters” and “full written words /characters”. The reason why they are called “half written words/ characters” is because [with them] the meaning is not completely provided, hence their incomplete body, just like the written word/character ‘moon’ in Chinese which lacks its half [= which only represents one crescent]. The reason why they are called “full written words/characters”, is because [with them] the reasoning (of the mind) is completely apprehended. Hence their complete body, just like the written word /character ‘sun’ in Chinese which represents its full form.”

In this passage, the written aspect of the names/words is put forth, but in close relation with the meaning or the absence of meaning attached to it. *Zì* is not taken as a simple graph.

Sēngyòu then goes on and writes:

又半字為體, 如漢文言字。滿字為體, 如漢文諸字。以者配言方成諸字。諸字兩合即滿之例也。言字單立即半之例也。半字雖單為字根本, 緣有半字得成滿字。
“Moreover, as for a “half written word/character” to constitute a body [independent structure], it is like the written word/character *yán* 言 in Chinese; and for a “full written word/character” to constitute a body [independent structure], it is like the Chinese written word/character *zhū* 諸 ‘all’. When one joins *zhě* 者 and *yán* 言 side by side (方?), one forms the written word/character *zhū* 諸 ‘all’. As a combination of two [bodies], the written word/character *zhū* 諸 ‘all’ is an example of fullness. As standing alone, the written word/character *yán* 言 is an example of the half type. Even if “half

written words/characters” stand alone, they constitute the base of the written words 字. Availing oneself of having “half written words/characters”, one gets to form “full written words /characters”.”

In this passage, Sēngyòu shows how the combination of smaller elements can form bigger meaningful written words.

According to *The Great dictionary of Buddhist studies*, compiled by Dīng Fúbǎo, “*Bànzì* 半字 represent the Sanskrit letters or, in other words, the 12 vowels and 35 consonants” 半字:梵語悉曇章之生字根本, 謂半字, 摩多十二字, 體文三十五字是也. Whereas “*mǎnzì* 滿字 represent the combination of vowels and consonants forming complete written words” 摩多體文相合而完成全字, 謂之滿字。²⁸

Hirata, on the other hand, considers that “‘Half-characters’ represent consonants, whereas ‘full characters’ represent the combination of consonants and vowels.” “半字” は子音文字を, “滿字” は子音文字と母音文字との合わせをさす。²⁹

But, as mentioned above, Xiè Língyùn, wrote that “As a matter of principle, all the small *zì* are called *bànzì*”. In other words, in Xiè Língyùn’s mind, *bànzì* 半字 rather corresponded to “letters” or basic written signs, without any distinction between vowel and consonants.

For the Chinese who had no knowledge of the study of the sounds, “half characters” probably represented the basic elements of the Indian script: the letters. But for the Indians who had a very good knowledge of the study of sounds, “half characters” probably represented something phonetically more precise such as “consonants”, in the way Hirata understood it. Here again the opposition between the phonetic analysis of the Indians and the graphic analysis of the Chinese should be pointed out.

In any case, we can see that in order to explain how the Indian script functioned, the terms *bàn zì* 半字 and *mǎn zì* 滿字 were respectively coined to refer to non-meaningful “letters” and to meaningful “syllables or written words”. From the Chinese point of view, letters or syllabic signs deprived of meaning could only be felt as “incomplete” compared to Chinese characters. Since there is a correspondence in Chinese between a written sign, a syllable and a linguistic unit, the Indian written signs that did not correspond to a syllabic word or a meaningful unit were described as *bànzì* 半字 “incomplete syllables, half written words” or “half characters”; whereas those which corresponded to a word or a meaningful syllable were presented as *mǎnzì* 滿字 “complete syllables, full words” or “full characters.”

If the example of the written word *zhū* 諸 ‘all’ could help understand how *bànzì* 半字 combined to form written words, the choice of *yán* 言 or *yuè* 月 as examples of “half characters/ incomplete syllables” does not really work, since both of them also happen to correspond to “meaningful syllables”. They were more specifically meant to express the Buddhist philosophical and theological idea of incompleteness or deficiency in this world.³⁰

All in all, the terms *bànzì* 半字 and *mǎnzì* 滿字 show that around the 5th century AD, *zì* was essentially associated with a meaningful syllable or a meaningful linguistic unit, and was not simply a graphic sign.

3. The understanding of the term *zì* 字 outside China

It is now useful to have a look at the way *zì* was understood outside China and see whether it was taken as a simple graphic sign or not. I shall first briefly confront the example of those who borrowed the Chinese script in Asia during the first half of the first millennium AD,

before turning to the description of the Chinese script by the Western missionaries in the 16th century.

3. 1. The use of *zì* by the Korean, Japanese, and Vietnamese who borrowed the Chinese writing system.

When the Vietnamese, the Korean or the Japanese “borrowed” the Chinese writing system, during the first half of the first millennium AD³¹, at first they did not record their own language, they wrote Chinese. They thus took *zì* 字 the way the Chinese did, that is as a graph specifically representing a monosyllabic meaningful unit in Chinese (a written word / morpheme). When they decided later to record their own proper names and words, they had to adapt the Chinese script to their own language. They could then use the Chinese graphs phonetically as a syllabary and write down the words in as many graphs as syllables counted. They could also use the Chinese graphs semantically to translate the same words, but with a pronunciation in their own language. In this case, depending on the language, a Chinese character could transcribe a polysyllabic word, especially in Japanese, and Korean but not in Vietnamese.³²

“Fish” is pronounced *sakana* in Japanese and counts 3 syllables. Thus *sakana* could either have been written phonetically with three graphs 左加奈, for example, or simply with the character *yú* 魚 which refers to “fish” in Chinese, but then pronounced *sakana*.³³ The Late Middle Korean word *petul-wos* 버들웃, which refers to a kind of spurge, could be written with three characters playing different roles 楊等柒 ([literally] willow-lacquer),³⁴ whereas the ancient word for “winter” 겨울 *kyezul* could be written with the Chinese character 冬 “winter”.³⁵ In the Vietnamese case, Chinese characters wrote essentially monosyllables or sesquisyllables of the ancient languages.³⁶

In a second step, the Vietnamese, the Korean and the Japanese developed their own national or autochthonous characters³⁷ to write down names and words in their own language.³⁸ One of the particularity of non-Chinese ‘characters’ is that they recorded autochthonous words which could be polysyllabic in Japanese and in Korean, but not in Vietnamese.³⁹ The term *hànzì* 漢字 “Chinese written words/ characters” was then used in each of these three languages⁴⁰ to differentiate Chinese from non-Chinese written linguistic units. *Hànzì* 漢字 always represented monosyllabic words/morphemes in all these three languages.

A distance eventually appeared with the original meaning of *zì*, as it came to refer to a simple written sign or a non-meaningful “letter” especially in Korean.⁴¹ We can read, for example, in the preface of the *Hunminjeong'eum* 訓民正音, presenting the new script said to have been invented in the 15th century by the King Sejong the Great of the Joseon dynasty of Korea, the following sentence: 予(...)新制二十八字 “I (...) have newly created 28 letters”, which are then described, and will later be known as the Hangul alphabet.⁴²

Even if *zì* 字 can refer to simple graphic signs or “letters”, usually more specifically written 字母 in Korean and Japanese, they are still basically taken as referring to monosyllabic meaningful units in the four languages at stake.

3. 2. The translation of the term *zì* by the European missionaries *et al.* around the 16th C.

When first confronted with the Chinese writing system, the western missionaries and first authors of books on China could only analyse it in comparison with their own script (the

alphabet). They thus talked about “letters”. Yet, impressed by the fact that the Chinese script counted a vast number of graphic signs, they very soon introduced the term “character”.

The Spanish soldier Bernardino de Escalante (1537-1605), who had never been to China, was apparently the first one to use the term “character” in the 16th century. Discussing the Chinese writing and education in the chapter 11 of his book, he wrote: “The Chinese do not have a limited number of alphabetic letters”. “They use more than 5 000 different characters”.⁴³

A few years later, the Spanish Augustinian Juan González de Mendoza (1545-1618) who also had never been to China presented the Chinese script saying: “There is no limited number of letters (like us), they [the Chinese] write with figures, learned with difficulties, because each word has its character”.⁴⁴

Matteo Ricci (1552-1610), the Italian author of the first Portuguese - Chinese dictionary with Michele Ruggieri (1543-1607), lived for 27 years until his death in China and learned to speak and write Chinese. He thus provided more information on the Chinese script and language in his book *Histoire de l'expédition chrestienne*. In the chapter “Des arts et sciences libérales entre les Chinois, et des degrés des hommes de lettres”, he wrote: “The way they write and compose resembles the hieroglyphic figures of the Egyptians...” He then mentioned that all the words in any Chinese language consisted in one syllable, that every word and every thing had its “hieroglyphic character”, that there were no fewer letters than words and that the number of characters did not exceed 70 to 80 000.⁴⁵

Deeply influenced by their own alphabetic writing system, the western missionaries and authors of books on China distinguished clearly graphs from words. They thus focused on the graphic aspect of *zì*. That is the reason why they chose the term “character”. In their languages, a character corresponded to a letter, a sign, a mark or a symbol used in writing or in printing.⁴⁶ The translation did not quite suit the meaning and the use of *zì* in Chinese, but the term “character” stayed until today.

While *zì* was directly translated as “character” i.e. “written sign”, by western missionaries, due to the influence of their own graphic system, its lexical meaning remained much broader in Korean, Japanese and Vietnamese. It referred to a written word/ morpheme, even if it sometime acquired the meaning “letter”, “character”.

4. The use of *zì* 字 in the first Chinese grammar before the introduction of *cí* 詞 “word”

At the end of the 19th century, Mǎ Jiànzhōng 馬建忠 (1845-1900) wrote the first grammar book on Classical Chinese: the *Mǎshìwéntōng* 馬氏文通 which was published in 1898. Mǎ was inspired by Western grammar⁴⁷, but deeply influenced by Chinese traditional philology and categories⁴⁸. He adapted ancient terms and also introduced some new ones to render western grammatical categories, such as “syntactic functions” and “cases”, for example. He is known as the first author to use the term *cí* 詞 which today translates the term “word”. But in his work, *cí* 詞 referred to “syntactic functions”, whereas “word” was rendered by *zì* 字. He then called “noun” *míngzì* 名字, “pronoun” *dàizì* 代字, “verbs” *dòngzì* 動字, “adverbs”, *zhuàngzì* 狀字, etc. We can see that Mǎ Jiànzhōng did not feel the need to invent a new way to translate “word”, since *zì* 字 corresponded sufficiently in his mind. He thus kept the traditional term *zì* 字 to refer to “(written) words”.⁴⁹

It was Yán Fù’s 嚴復 (1854 - 1921), the author of the first grammar presenting the categories of English grammar to Chinese, *Yīngwén Hàngǔ* 英文漢詁, that employed *cí* 詞 to refer to “words” for the first time, in 1904. Some twenty years later, in 1924, Lí Jīnxī 黎錦熙

(1890 - 1978) wrote and published the first comprehensive and systematic description of the modern Chinese grammar: *Xīnzhù guóyǔ fēnfǎ* 新著国语文法. He then made a clear distinction between *zì* and *cí* saying: 字就是一個一個單字, 詞就是說話的時候表示思想中一個觀念的詞語。有時一個字就是一個詞, 如“人”, “馬”, “紅”, “來”, 等。有時要兩個字以上組合起來才成功一個詞, 如“鸚鵡”, “老頭子”, “便宜”, “吩咐”, 等。文法中分別詞類, 是把詞作單位, 不問他是一個字或是幾個字, 只要是表示一個觀念的, 便叫做詞。 “Zì is a single written word/character, whereas *cí* is a word that expresses an idea in the mind when one speaks. Sometimes a “a written word/character” correspond to a “spoken word” like, for example, *rén* 人 “man”, *mǎ* 馬 “horse”, *hóng* 紅, “red”, *lái* 來 “to come”, etc. Sometimes it takes more than two “characters” to make a word like, for example, *yīngwǔ* 鸚鵡 “parrot”, *lǎotóuzi* 老頭子 “old man”, *piányi* 便宜 “cheap”, *fēnfù* 吩咐 “instruct”, etc. In grammar, parts of speech are distinguished, regardless of whether they are one “character” or several “characters”, as long as they express an idea, they are called words.”⁵⁰

The notion of *cí* “word” gradually anchored and the distinction between spoken words and written words was eventually made in China. A kind of dichotomy appeared then between the graphic and the oral forms of *zì* 字. But in reality, for a Chinese speaker *zì* 字 kept its meaning “written monosyllabic meaningful unit”. It is thus quite unfortunate that modern linguists who are used to make a clear distinction between language and writing tend to ignore this basic use of the concept as they leave aside the script.

5. Conclusion

To the contrary of what the English translation “character” suggests, the history and lexical development of *zì* 字 shows that *zì* 字 was never a term exclusively used for “graphs” and even less for graphemes. For those who borrowed the Chinese writing system, *zì* 字 perfectly suited the correspondence between graphs and monosyllabic words/ morphemes in Chinese, but not necessarily in their own language.

Since *zì* 字 first referred to names before widening its lexical meaning to written names, it came to stand between two different encoding systems: the language and the script.

The Qín new use of *zì* followed by Xǔ Shèn’s comprehensive and systematic dictionary of *zì* in 100 AD contributed to engender an important shift from a spoken to a written perspective in the analysis of the Chinese language. From the second century AD on, the Chinese linguistic analysis came to be based on the script, with Xǔ Shèn’s *Shuō wén jiě zì* 說文解字, and thus differs profoundly from Modern linguistics solely based on the spoken language. Thus *zì* 字 introduces the challenge of not only confronting two very different types of writing systems, but also two different linguistic traditions.

When confronting linguistic traditions, the problems and the complexity in the use of terms pops out. Western linguists who show little interest in the script and in the way it functions seem to be unaware of this terminological and conceptual problem related to the translation of *zì* in western languages as “character”. That becomes evident when one sees that they associate characters with graphemes.

As a matter of fact modern Chinese scholars are not spared. The rendering of *zì* as “character”, and the introduction of *cí* 詞 to refer to “words” exclusively, seem to have influenced the apparition of a new expression: *zìcí* 字詞. Wáng Guìyuán 王贵元 and Lǐ Jiéqióng 李洁琼 (2022), for example, largely use this expression *zìcí*. In their article *Hànyǔ zìcí fāzhǎn lùngāng* 汉语字词发展论纲, they translate as ‘An outline of Chinese word (*zìcí*) development’, they insist on the need to integrate the graphs (written words) into the analysis

of the language. Interestingly, they could have used the term *zìxíng* 字形 to refer to graphs, but they took *zì* as “characters”/ “graphs” and translated *zìcí* as *word* in their title!

The long and now widely spread use of the term “character” makes it difficult to introduce a new translation of *zì* 字 such as, for example, “written morphemes” or “written linguistic units”, but it is important to be aware of the shortcoming of this translation, especially in the field of linguistics.

Declaration of conflicting interests

The Author declares that there is no conflict of interest.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Acknowledgment

I want to express my gratitude to O. Venture, I. Tamba, C. Harbsmeier, Y. Mghari, H-L Dao, J. Phan for the information they provided me when writing this article. It goes without saying that any errors are my own.

Notes

¹ In chapter 3 “Chinese Writing” of their book *Chinese Linguistics. An Introduction*, Arcodia and Basciano make this association no less than 8 times (p. 62, 64, 65, 68 (3 times), 70, and 71).

² See Kohrt 1986, and Meletis 2019.

³ As proposed by Meletis (2019), we should look for a definition of grapheme as a universal basic unit of writing. But in order to embrace all writing systems and avoid the limitation of graphemes to phonographic writings, graphemes should not be *treated as signs of phonemes* nor as *lexical units*. Thus among the three criteria proposed by Meletis for a general definition of graphemes: 1. lexical distinctiveness, 2. linguistic value, 3. minimality, the second criteria which encouraged him to count characters as graphemes poses problem as it maintains a confusion between the characters and the smallest elements of the Chinese script.

⁴ Chén Wěi & Péng Hào 2011, vol 1, p. 3, slips 21.

⁵ According to Qiū Xīguī, the graph 【才/心】 should be read *zì* 字 “name, give a public name”, see Chén Wěi & Péng Hào 2011, vol. 1, p. 24, note 19. This graph also appears two times on a bamboo slip from the Shànghǎi manuscripts in the *Bào shū yá yǔ xí péng zhī jiàn* 鮑叔牙與隰朋之諫, but it is understood as referring to the word *zāi* 栽 “calamities”; cf. Lǐ Shǒukuí 李守奎 *et al*, 2007, *Shànghǎi bówùguǎn cǎng Zhànguó Chǔ zhū shū* 上海博物館藏戰國楚竹書, Běijīng: Zuòjiā chūbǎnshè, p. 897 (slips 5.2.6 and 5.2.8).

⁶ Chén Wěi & Péng Hào 2011, vol 1, p. 22, slips 10 and 12. Zhāng Shǒuzhòng 2000, p. 147.

⁷ Shuǐhǔdì Qín mù zhújiǎn zhènglì xiǎozǔ 1990/2001, p. 76 and 161, slip n° 86.

⁸ *Ibidem*, p. 101 and 206, slip n°150 正 3.

⁹ Wáng Huànlín 2007, p. 26 sq and 179 sq. Interestingly, one would not expect *yán* 言 here; see 1.3 above.

¹⁰ It is worth noting that among the Dunhuang manuscripts, a similar table of multiplication including 45 multiplications indicates the total sum at the end in terms of *wén* 文 instead of *zì* 字: 一千一百五十五文. See Lǐ Yán 李严 1963 (p. 39), *Zhōngguó gǔdài shùxué shǐliào* 中国古代数学史料 (Shànghǎi: Shànghǎi kēxué jìshù chūbǎnshè 上海科技出版社), quoted by Wáng Huànlín 2007, p. 180.

¹¹ Cf. Qiū Xīguī 2014, vol. 6, p. 95, 96 and vol. 7, p. 261, columns 14 and 18.

¹² *Ibidem*, vol. 4, p. 206, slip 240.

¹³ See Bái Yùlán 2012, p. 23, 2017, p. 36.

¹⁴ *Shǐjì* 6, p. 245.

¹⁵ see Bottéro 2004b, p. 28.

¹⁶ See “Chūnguān, Wàishǐ”, SSSJZS, p. 820c.

¹⁷ See, for example, the biographies of Lǎozǐ, Hán Fēi, Mencius, Lǚ Bùwéi, Dǒng Zhòngshū, etc., which all mention texts written with a large number of *yán* 言. *Zì* 字, on the other hand, was very often used for “public name” in the *Shiji*.

¹⁸ Cf. Hú Píngshēng 胡平生 & Hán Ziqiáng 韓自強 1988, p. I & X.

¹⁹ This is the case, for example, for at least five chapters discovered in *Mǎwángduī* which simply provide a single number for the words included: 570 for 須賈說穰侯章五百七十, 858 for 朱已謂魏王章八百五十八, 563 for 謂起賈章五百六十三, 569 for 觸龍見趙太后章五百六十九, 300 for 秦客卿造謂穰侯章三百, as well as the total number of written words they all gather at the end: 2 860 大凡二千八百七, cf. Qiú Xīguī 裘錫圭 (ed), 2014, vol 3, p. 244.

²⁰ *Wūwēi Hàn jiǎn*, 2005, p. 120 (slip n° 53).

²¹ Bottéro 2004b. Note that the simple versus the complex structure of the characters is never touched upon in *Shuō wén jiě zì*.

²² The *Hànshū* 漢書 30 *Yìwénzhì* (1992, Zhōnghuá shūjú p. 1720) records a single title with *zì*: *Bié zì* 別字, whereas the *Suǐshū* 隋書 32 *Jīngjìzhì* (1997, p. 942 sq) counts more than 30 titles with *zì*, and 7 with *wénzì*.

²³ Van Gulik 1980, p. 46-47.

²⁴ Kotyk 2021, p. 242.

²⁵ *Taishō shinshū daizōkyō* 12, n° 375.

²⁶ Chaudhuri 1998, p. 14.

²⁷ See his “Collected notes from the Tripitaka” 出三藏記集, T. 84, n° 2145.

²⁸ Dīng 1984, p. 452.1 and p. 1248.2.

²⁹ Hirata 1994, p. 54 note 13.

³⁰ Dīng 1984, p. 1248.2, and the explanations provided, for example, by the Japanese Buddhist monk Annen 安然 (9th C.) in his *Shittanzō* 悉曇藏 (880), T. 84, n° 2702, ch. 7.

³¹ See Coulmas 1991, p. 113 for Vietnamese; Seely 1991, p. 6 sq for Japanese; and Lee and Ramsey 2011, p. 1-2 for Korean.

³² Cf. Phan 2019.

³³ This example does not give the ancient Japanese pronunciation, it is simply meant to explain the methods.

³⁴ With 楊 standing for the word *petul* ‘willow’, 等 transcribing the phonogram *tul*, and 柒 for the word *wos* ‘laquer’. Thus the method could be more complicated as the Chinese graphs could combine phonetical and semantical uses in a single word. see Lee and Ramsey 2011, p. 88.

³⁵ *Ibid*, p. 87.

³⁶ As illustrated, for example, by the title of the 15th century Buddhist “false canon” : *Phật thuyết đại báo phụ mẫu ân trọng kinh* 佛說大報父母恩重經, in which 8 characters out of 9 represent monosyllables, see Phan 2019, p. 5.

³⁷ In Vietnamese, they are called “southern characters” *chữ nôm* 字喃, in Japanese and Korean “national characters” *kokuji* 国字 and *Kukcha* 国字. Lim and Lee (2019, note 1) call them Korean-coined characters and provide other names for them.

³⁸ The Vietnamese created much more autochthonous characters than the Korean and the Japanese did. While the Japanese invented quite a few characters referring to animals (especially fish) and to plants (see a list of 174 such graphs under the link below “origamijapan”), the Korean mostly used them for toponyms as well as for some other specific words. See, for example, the *Great dictionary of chữ nôm Đại từ điển chữ Nôm* 大字典字喃 by Vũ Văn Kính 1998 (or the link <http://www.hannom-rcv.org/wi/index.php/%F0%A1%A8%B8%E5%96%83> for *chữ nôm*).

For Japanese national characters see: <https://origamijapan.net/origami/2018/05/28/kokuji/>; and for some Korean ones, see: <https://zh.wikipedia.org/zh-hant/%E9%9F%A9%E5%9B%BD%E5%9B%BD%E5%AD%97>; as well as Lim and Lee 2019.

³⁹ As shown by Vũ Văn Kính’s *chữ nôm* dictionary *Đại từ điển chữ Nôm* (1998).

⁴⁰ Under the pronunciation *Hán tự* in Vietnamese, *Kanji* in Japanese and *Hanja* 한자 in Korean.

⁴¹ According to the online Vietnamese dictionary *Từ điển Hán Nôm*, 字 can refer to 1° “letter” “character”, “word”, 2° “document”, 3° an honorific title, 4° “an engaged person”; cf. <https://hvdic.thivien.net/why/%E5%AD%97>

In Japanese dictionaries, 字 can also refer to a letter, character, etc. (*Dictionnaire standard Japonais-français* Tokyo, Taishukan, 1970).

⁴² See in Lee and Ramsey (2011), the reproduction of this passage in the first pages. I am thankful to Younès Mghari for pointing out this source to me, and for more information on the Korean writing system.

⁴³ Bernardino de Escalante 1577, *Discurso de la navegación que los portugueses hazen a los reinos y provincias del oriente, y de la noticia que se tiene del reino de China*, chapter 11 “De las letras y caracteres de los Chinas, y de sus estudios generales” fol. 62. Published in Spanish in Seville, the book was translated in English under the title *Discourse of the Navigation which the Portugales doe Make to the Realms and Provinces of the East Partes of the World, and of the knowledge that growes by them of the great thinges, which are in the Dominion of China.* by John Frampton (London 1579). According to Lachs (1965, p. 742) “it was the second book on China published in Europe” after Gaspar Da Cruz.

⁴⁴ Cf. Juan González de Mendoza (Roma 1585), *Historia de las cosas más notables, ritos y costumbres del gran reyno de la China*, chapter 13 “De los Caracteres y letras que usan los Chinos, y de los estudios y escuelas que en este Reyno ay, y de otras cosas curiosas” (Libro III, p. 104). Lachs (1965 p. 743), notes that Mendoza’s book was the “the most influential and detailed work on China prepared in the sixteenth century.”

⁴⁵ Matteo Ricci & Nicolas Trigault, 1617, *Histoire de l’expédition chrétienne au royaume de la Chine: 1582-1610*. (Paris : Desclée, 1978). See Book 1 chapter V, p. 20-21, text available at: <https://gallica.bnf.fr/ark:/12148/bpt6k8704822w.r= Ricci%20Histoire%20de%20l%27exp%C3%A9dition%20chrestienne?rk=42918;4>

⁴⁶ *Cambridge dictionary*: “A letter, number, or other mark or sign used in writing or printing, or the space one of these takes”; *Littre*: Signe tracé ou écrit. Les lettres de l’alphabet, les signes de punctuation, les chiffres sont des caractères.”

⁴⁷ Cf. Peyraube 1999, 2001.

⁴⁸ Cf. Peyraube 2001, p. 349, Zádrapa 2016, 2020.

⁴⁹ Note that Victor Mair (1997, p. 15) in his translation of the prefaces of the *Mǎshìwéntōng* translates *zì* as “graph” instead of “written word”.

⁵⁰ Lí Jīnxī 1945, p. 2 : ch. 2. 2 *Zì yǔ cí* 字與詞.

References

- Arcodia G and Basciano B (2021) *Chinese Linguistics: An Introduction*. Oxford: Oxford University Press.
- Bái YL 白於藍 (2012) *Zhànguó Qín Hàn jiǎn bó gǔshū tōngjiǎzì huì zuǎn* 戰國秦漢簡帛古書通假字彙纂. Fúzhōu: Fújiàn rénmin.
- Bái YL (2017) *Jiǎn bó gǔshū tōngjiǎzì dàxì* 簡帛古書通假字大系. Fúzhōu: Fújiàn rénmin.
- Bottéro F (2004a) Chinese Characters Versus Other Writing Systems: The Song Origins of the Distinction Between ‘Non-compound Characters’ (*wen*) and ‘Compound Characters’ (*zì*). In: Takashima K and Jiang SY (eds) *Meaning and Form: Essays in Pre-Modern Chinese Grammar*. LINCOM Europa, pp. 1-17.
- Bottéro F (2004b) Revisiting the *wen* and the *zì*: The Great Chinese Characters Hoax. *Bulletin of the Museum of Far Eastern Antiquities* 74 (2002): 14-33.
- Chaudhuri SK (1998) Siddham in China and Japan. *Sino-Platonic Papers* n° 88.
- Chén W 陳偉 and Péng H 彭浩 (eds) (2011) *Chǔ dì chūtǔ Zhànguó jiǎncè héjí* 楚地出土戰國簡冊合集. Wūhàn: Wūhàn dàxué chūbǎnshè.
- Chén W et al. (ed) (2014) *Qín jiǎndú héjí* 秦簡牘合集. Wūhàn: Wūhàn dàxué chūbǎnshè (6 vols.).
- Coulmas F (1991) *The Writing Systems of the World*. Oxford: Blackwell.
- Dān XW 單曉偉 (2017) *Qín wénzì zìxíng biǎo* 秦文字字形表. Shànghǎi: Shànghǎi Gùjí chūbǎnshè.
- Dīng FB 丁福保 (1984) *Fóxué dà cídiǎn* 佛學大辭典. Běijīng: Wénwù chūbǎnshè.
- Fāng Y 方勇 (2012) *Qín jiǎndú wénzìbiān* 秦簡牘文字編. Fúzhōu: Fújiàn rénmin chūbǎnshè.
- Handel Z (2017) Chinese Characters. In: R. Sybesma and al (eds), *Encyclopedia of Chinese Language and Linguistics*. Leiden : Brill (1) pp. 435-438.
- Handel, Zev, 2019, *Sinography: The Borrowing and Adaptation of the Chinese Script*, Brill.
- Hirata S 平田昌司 (1994) 謝靈運「十四音訓敘」の系譜 [The genealogy of Xie Lingyun’s *Shisi yinxun xu*]. In: Tokio T 高田時雄 (ed) *Chūgokugoshino shiryō to hōhō* 中國語史の資料と方法. Kyōto: Institute of Research in Humanities, pp. 33-80.
- Hodge S (2006) [On the Eschatology of the Mahāparinirvāna Sūtra and Related Matters](#) (PDF), lecture delivered at the University of London, SOAS, archived from [the original](#) (PDF) on June 14, 2013

- Hodge S (2012) *The Mahayana Mahaparinirvana Sutra. The text & its Transmission* (PDF), corrected and revised version of a paper presented in July 2010 at the Second International Workshop on the Mahaparinirvana Sutra held at Munich University, archived from [the original](#) (PDF) on December 19, 2013
- Hodge S (2004) Textual History of the MAHĀYĀNA-MAHĀPARINIRVĀNA-SŪTRA
https://buddhanature.tsadra.org/index.php/Articles/Textual_History_of_the_Mah%C4%81y%C4%81na-Mah%C4%81parinirv%C4%81%E1%B9%87a-S%C5%ABtra_by_Stephen_Hodge
- Hú PS 胡平生 and Hán ZQ 韓自強 (1988) *Fùyáng Hàn jiǎn Shījīng yánjiū* 阜陽漢簡詩經研究. Shànghǎi: Gùjī chūbǎnshè.
- Huáng DK 黃德寬, Xú ZG 徐在國, Chéng Y 程燕, et al. (2017, 2020) *Zhànguó wénzì zìxíng biǎo* 戰國文字字形表. Shànghǎi: Shànghǎi Gùjī (3 vols.).
- Karlgren B (1944) The Book of Odes: *Ta Ya* and *Sung*. *Bulletin of the Museum of Far Eastern Antiquities* (17): 65-99.
- Kotyk J (2021) The Study of Sanskrit in Medieval East Asia: China and Japan. *Hualin International Journal of Buddhist Studies* (4)2: 240-273.
- Kohrt M (1986) The term ‘grapheme’ in the history and theory of Linguistics. In: Augst G (ed) *New Trends in Graphemics and Orthography*. Berlin, Boston: De Gruyter, pp. 80-96.
- Lachs DF (1965, 1971, 1994) *Asia in the Making of Europe*, vol. 1, *The Century of discovery*, Book 2. Chicago and London: The University of Chicago Press.
- Lee KM and Ramsey RS (2011) *A History of the Korean Language*. Cambridge: Cambridge university Press.
- Lí JX 黎錦熙 (1945) *Xīnzhù guóyǔ fénfǎ* 新著國語文法. Shànghǎi: Shāngwù yìnshūguǎn.
- Lim Y and Lee JY (2019) Korean-coined characters and syllable structure. *Journal of Chinese Writing Systems* 2019 (3) 3: 205-220.
- Liu Z 劉釗 (2005) *Guōdiàn Chǔ jiǎn jiàoshì* 郭店楚簡文校釋. Fúzhōu: Fújiàn rénmin.
- Mǎ JZ 馬建忠 (2000) *Mǎshìwéntōng* 馬氏文通. Běijīng: Shāngwù yìnshūguǎn.
- Mair V (1997) Ma Jianzhong and the Invention of Chinese Grammar. In: Sun CF (ed) *Studies on the History of Chinese Syntax*. Berkeley: Journal of Chinese Linguistics (Monographs Series 10), pp. 5-26.
- Meletis D (2019) The grapheme as a universal basic unit of writing. *Writing Systems Research* (11)1: 26-49.
- Peyraube A (1999) Sur les sources du *Ma shi wen tong*. *Histoire et épistémologie du langage*. HEL (21)2: 65-78.
- Peyraube A (2001) On the sources of *Mashi wentong*. In: Lackner M, Amelung I and J. Kurtz (eds) *New Terms for New Ideas. Western Knowledge and Lexical Change in Late Imperial China*. Leiden: Brill, pp. 341-356.
- Phan JD (2019) Sesquisyllabicity, Chữ Nôm, and the Early Modern embrace of vernacular writing in Vietnam. *Journal of Chinese Writing System* (4)1: 1-14.
- Qiú XG 裘錫圭 (ed) (2014) *Chángshā Mǎwángdū Hàn mù jiǎn bó jíchéng* 長沙馬王堆漢墓簡帛集成. Běijīng: Zhōnghuá shūjú, (7. vols.).
- Salomon R (1998) *Indian Epigraphy, A Guide to the Study of Inscription in Sanskrit, Prakrit, and Other Indo-Aryan Languages*. New York, Oxford: Oxford University Press.
- Seely C (1991) *A History of Writing in Japan*. Leiden: Brill.
- Shísānjīng zhùshū* 十三經注疏 (abbreviated as SSJZS) (1980). Běijīng: Zhōnghuá shūjú.
- Shuǐhǔdì Qín mù zhǔjiǎn zhěnglì xiǎozǔ 睡虎地秦墓竹簡整理小組 (1990, 2001) *Shuǐhǔdì Qín mù zhǔjiǎn* 睡虎地秦墓竹簡. Běijīng: Wénwù chūbǎnshè.
- Takakusu J (1896) *A record of the Buddhist Religion as Practised in India and the Malay Archipelago (AD 671-695)*, By I-Tsing, Translated by J. Takakusu, B.A., Ph. D. Oxford: Clarendon Press.
- Van Gulik RH (1980) *Siddham. An Essay on the History of Sanskrit Studies in China and Japan*. Dehli: The Jayyed Press.
- Wáng GY 王貴元 and Lí JQ 李潔琼 (2022) *Hànyǔ zìcí fāzhǎn lùngāng* 汉语字词发展论纲 [An outline of Chinese word development]. *Xuéshù yánjiū* 学术研究 (7): 166-176.

-
- Wáng HL 王煥林 (2007) *Lǐyē Qín jiǎn jiào gǔ* 里耶秦簡校詁. Běijīng: Zhōngguó wénlián.
- Vũ VK (1998) *Đại từ điển chữ Nôm* 大字典喃 [Great Nom Dictionary]. Hồ Chí Minh-ville: Éditions Văn Nghệ.
- Wúwēi Hàn jiǎn 武威漢簡 (2005). Běijīng: Zhōnghuá shūjú.
- Xú X 徐鉉 (1963) *Shuō wén jiě zì* 說文解字. Běijīng: Zhōnghuá shūjú.
- Xú ZG 徐在國, Chéng Y 程燕 and Zhāng ZQ 張振謙 (2017, 2020) *Zhànguó wénzì zìxíng biǎo* 戰國文字字形表. Shànghǎi: Shànghǎi Gǔjí (3 vols.).
- Yang-Drocourt Z (2022) *L'écriture chinoise. Au-delà du mythe idéographique*. Paris: Armand Colin.
- Zádrapa L (2016) Mǎshì wéntōng 馬氏文通. In: Rint S (ed) *Encyclopedia of Chinese Language and Linguistics*. Leiden: Brill. Consulted online: http://dx.doi.org/prex.num.bulac.fr/10.1163/2210-7363_ecll_COM_000286
- Zádrapa L (2020) The First Chinese Grammar between the Domestic Tradition and Western Modernity: Neglected Aspects of Mǎ Jiànzhōng's Prefaces to Mǎshì wéntōng (1898)". *Archiv Orientalní* (88): 57-93.
- Zhāng SZ 张守中 et al. (2000) *Guōdiàn Chǔ jiǎn wénzì biān* 郭店楚簡文字编. Běijīng: Wénwù chūbǎnshè.
- Zhāng XC 張顯成 (2010) *Qín jiǎn zhúzi suǒyǐn* 秦簡逐字索引. Sìchuān dàxué chūbǎnshè.
- Zhāng XC (2013) *Chǔ jiǎn bó zhúzi suǒyǐn* 楚簡帛逐字索引. Sìchuān dàxué chūbǎnshè.