

# Proposal for a New Classification System for Modern Chinese Characters\*

Tereza Slaměňíková

*Návrh na nový systém klasifikácie moderných čínskych znakov*

*Resumé* Príspevok predstavuje nový model kategorizácie moderných čínskych znakov. Navrhovaný dvojrozmerný model zahŕňa päť skupín rozdelených do 20 kategórií. Príslušnosť k skupine je podmienená kompozičnými prvkami, príslušnosť ku kategórii odzrkadľuje povahu vzťahu medzi celým znakom a jeho komponentmi z hľadiska sémantickej alebo fonetickej motivácie.

*Abstract* The paper introduces a new categorization model of Chinese characters. The suggested two-dimensional model includes five groups subdivided into 20 categories. The group status reflects the decomposition specification, and the category status reflects the nature of the relationship between the whole character and its components in terms of semantic or phonetic motivation.

*Key words* modern Chinese grammar · modern Chinese characters · categorization of Chinese characters, semantic component, phonetic component

## *Introduction*

The complicated graphic form of the basic unit of the Chinese writing system and the amount of represented linguistic information has undoubtedly contributed to the fact that grammar is one of the key disciplines in Chinese linguistics.

\* This paper is based on my dissertation at Palacký University in Olomouc. Its publication was made possible with the support of the *Fond pro podporu vědecké činnosti* at the Faculty of Arts, Palacký University in Olomouc (FPVČ2017/16 Svět v sinogramech).

Almost two thousand years have passed since the Han-dynasty scholar Xu Shen 許慎 (?54–?125) finished the oldest known Chinese grammatological study *Shuo Wen Jie Zi* 說文解字 (The Structure Analysis of Primary Characters and Meaning Explanation of Secondary Characters, 121). In this extensive work, Xu Shen, among others, presented a six-category classification method for Chinese characters that until now represents a leading approach employed when analyzing Chinese characters. However, one cannot overlook the fact that Chinese script went through several major changes in the course of its development. On the one hand, these changes led to a simplification of its graphics, but on the other hand, it also led to the corruption of its graphic composition and the demotivation of the original semantically or phonetically motivated components. Due to this, using Xu Shen's categorization when analyzing the contemporary form of Chinese script seems to be rather inadequate. The diachronic point of view<sup>1</sup> simply cannot sufficiently reflect the specific features of modern Chinese characters.

Although in the 20<sup>th</sup> century several attempts were made to revise the traditional model of categorization,<sup>2</sup> all of them seem to be, in one way or another, insufficient in terms of an effective classification system for modern Chinese characters. The currently most recognized approach, called the new six categories, was introduced by Chinese linguist Su Peicheng 苏培成 (1935–).<sup>3</sup> It can be said

- 1 Chinese researchers have traditionally paid significant attention to the etymological development of the Chinese writing system. It was probably the depth of the diachronic perspective that paradoxically led to the situation where researchers have not felt the need to deal directly with the contemporary relationship between graphic and linguistic structures.
- 2 Especially to be mentioned are classifications introduced in the following publications: Tang Lan 唐兰, *Zhongguo Wenzixue* 中国文字学 [Chinese Grammatology] (Shanghai: Shanghai Guji Chubanshe, 2001), 59–98; Chen Mengjia 陈梦家, *Zhongguo Wenzixue* 中国文字学 [Chinese Grammatology], ed. by Yu Guolin 俞国林 (Beijing: Zhonghua Shuju, 2006), 24–94, 256–258, 354–357; Qiu Xigui 裘锡圭, *Wenzixue Gaiyao* 文字学概要 [Outline of Grammatology] (Beijing: Shangwu Yinshuguan, 1988), 97–204; Wang Ning 王宁, *Hanzixue Gaiyao* 汉字学概要 [Outline of Chinese Grammatology] (Beijing: Beijing Shifan Daxue Chubanshe, 2001), 63–82.
- 3 Su Peicheng 苏培成, *Xiandai Hanzixue Gangyao* 现代汉字学纲要 [Outline of Modern Chinese Grammatology] (Beijing: Beijing Daxue Chubanshe, 2001).

that its main innovation lies in two areas: (1) elimination of unnecessary categories of loangraphs and variants<sup>4</sup> and (2) involving the third type of constructional units, so-called unmotivated constituents. Only two of the traditional six categories can be found in the new system as well, i.e. ideograms and phonograms. The graphic form of characters belonging to these two categories has up to now been considered fully motivated. On the other hand, this is not the case for the four newly implemented categories. Two of them are partly motivated since they are combinations of an unmotivated constituent and a semantic or phonetic component. Finally, two categories of completely unmotivated characters can be identified. One of them is simple unmotivated characters indivisible into smaller graphic units and the other one are complex characters composed of two or more unmotivated constituents.

The system of the new six categories represents an important change in the approach to the classification of Chinese characters. It is an unquestionable sign of a shift towards synchronic description of the relationship between graphic units and their current function. However, the description is rather shallow, which is a fact that cannot be overlooked. The nature of two of the basic constructional units seems insufficiently explained. Su Peicheng does not sufficiently specify the parameters that must be present for a graphic unit to be considered a semantic component or an unmotivated constituent. In the former case, the rejection of the diachronic approach represents a basic principle of modern Chinese grammar; therefore, one cannot avoid the question whether it is still an etymological explanation that should be used as an evaluation tool. If so, another important question that arises is how the fact that the meanings of characters have evolved should be taken into consideration while analyzing the current semantic link between characters and their components.

As for unmotivated constituents, it is obvious that the aim of modern Chinese grammar is to set a clear boundary between the constructional and structural approaches to Chinese character graphics.<sup>5</sup> The first approach examines the

4 As for Xu Shen's six categories, this paper adopts English equivalents used by D. Uher: *xiangxing* 象形 (pictogram), *zhishi* 指示 (symbol), *huiyi* 会意 (ideogram), *xingsheng* 形声 (phonogram), *jiajie* 假借 (loangraph), *zhuanzhu* 转注 (variant). For details, see David Uher, *Hanská grammatologie* [Han Grammar] (Olomouc: Univerzita Palackého, 2013), 297–298.

5 Su Peicheng, *Xiandai Hanzixue Gangyao*, 64–65.

connection between the character's graphics on one side and the meaning or sound of the recorded linguistic unit on the other side. The structural approach is interested strictly in character graphics; it examines the number, typology and arrangement of the minimal graphic units, i.e. strokes, and basic graphic units, i.e. graphic elements. The unmotivated constituent is defined as a graphic unit carrying no useful information about the character's meaning or pronunciation, which is more problematic than it might seem because the definition does not specify how deep the decomposition is supposed to be carried out, and thus a blending with the structural decomposition methods is inevitable. Besides, one cannot fail to notice that while describing the classification system, Su Peicheng avoids stating how many characters belong to each category. Therefore, one doubts whether he has verified the applicability of the proposed system through a complex analysis of the modern Chinese signary.

Considering all the above-mentioned facts, an attempt was made to propose a new classification system that will more effectively reflect the current features of modern Chinese characters. Generally speaking, each character is interpreted as a meaningful syllable, usually a morpheme.<sup>6</sup> Thus, the conducted graphemic analysis examined how the current graphic composition of characters relates to the current sound or meaning of the recorded morpheme. Regarding the relatively high number of currently used characters, a representative sample of characters has been selected to perform the analysis. This paper presents a proposal for a new classification system obtained through analysis of the so-called 2,500 frequently used characters.

### I *Analyzed Signary*

One of the main interests of modern grammarology lays in the research on occurrence frequency in modern Chinese texts. Under the guidance of the State

6 For details on morphemes in Chinese, see David Sehnal, »Čínské znakové písmo. Jeho povaha a vývoj« [Chinese Writing System. Its Characteristics and Nature], *Zmýšlenek a představ Žluté země* [From the Thoughts and Imaginations of Yellow Earth] (Brno: Moravské zemské muzeum, 2002), 13–28

Language Work Committee, two normative lists of modern characters were assembled in the second half of the 1980s. The larger one, called *Xiandai Hanyu Tongyong Zi Biao* 现代汉语通用字表 ('Table of Commonly Used Modern Chinese Characters'), includes 7,000 items. These characters are supposed to represent the signary of modern Chinese.<sup>7</sup> Additionally, a list of half the size was compiled. The characters in the so-called *Xiandai Hanyu Changyong Zi Biao* 现代汉语常用表 ('Table of Frequent Chinese Characters') are divided into two parts: 2,500 *changyong zi* 常用字 ('frequently used characters') and 1,000 *ci changyong zi* 次常用字 ('less frequently used characters'). Statistically, high frequency characters cover a significant part of a common Chinese text. As for the 2,500 frequently used characters, the coverage ratio is 97.97%.<sup>8</sup>

In the view of the high coverage ratio, the 2,500 frequently used characters represent the core of the modern Chinese signary. Therefore, they stand in for a quantitatively representative sample for analysis. From a qualitative point of view, the signary had to be defined so that it would firstly not eliminate the characters of a particular constructional principle a priori, and secondly maintain (in the largest possible measure) the proportion of constructional principles. Therefore, the list of 2,500 frequently used characters compiled on the basis of a different classification criterion appears to be a sufficiently representative sample.

## 2 Analysis Procedure

The primary aim of the graphemic analysis was to decompose characters into smaller graphic units that possess a sound or meaning indicating function. Nevertheless, supplementary decomposition criteria need to be defined in order to process all the characters. Considering the possibility of a specific type of

7 In 2013, a new standardized list of 8,105 characters was issued. The research presented in this study was started before the *Tongyong Guifan Zi Biao* 通用规范字表 ('Table of General Standard Chinese Characters') was published, and thus is based on the older list from the 1980s. According to their frequency of use, characters are divided into three levels in the new list. There can be found 3,500 characters designated as frequently used. All of the 2,500 frequently used characters from the older list are included within them.

8 Su Peicheng, *Xiandai Hanzixue Gangyao*, 52.

decomposition, analyzed characters were divided into five groups. This section provides a description of the classification process. Moreover, criteria to evaluate the semantic and phonetic motivations had to be specified. This issue is going to be further discussed in the next section.

The key unit of the decomposition process was the component. However, rethinking the nature of modern Chinese characters, its definition was more precisely specified. First, to meet the requirement of a motivated graphic unit, which from a synchronic point of view possess the ability to provide a semantic or phonetic link to the morpheme represented by one character, the property of recurrence was added to the definition. This means that only graphic units that occur as a semantic or phonetic indicator in at least two characters of the modern Chinese signary<sup>9</sup> were considered as components. Thus, a set of 778 components was identified in the 2,500 frequently used characters.

Considering the main two-component construction of characters in *Shuo Wen Jie Zi*,<sup>10</sup> all characters were first examined in terms of the possibility of decomposition into two components. A new typology of components was elaborated through the decomposition process. In addition to the traditional semantic and phonetic components (noted as s-components and p-components in the proposed model), two more types of components were recognized. The first one was components that provided both a semantic and phonetic link to the morpheme represented by the character: these components were noted as s/p-components. The other one was graphic units identified within the set of 778 components; however, in the particular case of occurrence they do not possess any semantic or phonetic function. While the first three types of components are motivated, the semantic or phonetic function of the fourth type was neutralized, which was the reason to call them n-components.<sup>11</sup> Following the described

9 Including their occurrence on the higher constructional level, i.e. their occurrence as characters.

10 The dominant two-component composition can be assumed from the occurrence of the secondary characters (ideograms and phonograms) in *Shuo Wen Jie Zi*. For details, see footnote 29 below.

11 While providing character examples of categories described below, the four types of components are distinguished as follows: meaning can be found under an s-component; pronunciation can be

criteria, more than three fourths of the analyzed characters were classified as divisible into two components. In the proposed model, these characters can be found under Group C. Considering the different combinations of component types, seven categories were recognized.

After sorting out the two-component characters, the rest of the characters were examined. At first glance, they show a high diversity of composition complexity. Due to the fact that a certain amount of three or more component characters can also be found in *Shuo Wen Jie Zi*, they were first investigated in terms of the possibility of decomposition into more than two components. These kinds of characters (to be specific, characters composed of two motivated components and at least one other component of any kind) were put into Group D. However; this was the case for only a limited number of characters. Most of the remaining characters were still unclassified. At this point of the analysis, the decomposition method of the structural approach was employed in order to distinguish characters with a simple and a complex composition. The dictionary *Hanzi Xinxi Zidian* 汉字信息字典 [Dictionary of Chinese Character Information]<sup>12</sup> was used to determine divisibility or indivisibility into graphic elements.

Group B consists of characters that can be divided into two or more elements. Besides, further analysis has shown that some of them might also contain one graphic unit that has been recognized as a component; thus, four categories can be identified within this group. Usually, any of the given elements is an s-, p- or s/p-component at the same time; although their graphics may not necessarily overlap. On the other hand, group A contains characters indivisible into elements. The few exceptions that go against this principle are characters whose graphic form is indivisible in terms of constructional integrity, which is considered superior to the structural approach.

Last of all, a specific group of characters with the so-called zero meaning has to be mentioned. In fact, chronologically speaking, this was the first group of characters that was sorted out during the classification process. Since they themselves do not carry any meaning, they deviate from the typical arrangement

found under a p-component; both meaning and pronunciation can be found under an s/p-component; no additional information can be found under an n-component.

<sup>12</sup> Beijing: Kexue Chubanshe, 1988.

between graphic and linguistic structure. This was the reason for putting them in a separate group labeled Group E.

### 3 *Semantic and Phonetic Motivation*

In order to classify modern Chinese characters in terms of possible links between their composition and a represented linguistic unit, a complex graphemic analysis was carried out. To be able to conduct this, criteria had to be defined, which a graphic unit had to meet to be considered an effective semantic or phonetic component in relation to the current meaning or pronunciation of the character. Different studies were consulted during this process. As for phonetic motivation, a great variability of methods and results were observed. On the other hand, the issue of the synchronic relationship between character graphics and meaning is hardly discussed.

Generally speaking, there can be found two basic approaches to analyzing phonetic motivation. The first approach primarily focuses on phonetics; the relationship between the syllabic value of a component and all the characters it occurs with is examined in it and, as a result, the following types of phonetics are distinguished: a) ideal phonetics (all characters share the same pronunciation as the phonetic); b) several groups of phonetics with regular differences; and in a case of an indulgent approach c) several groups of irregular phonetics.<sup>13</sup> The second

13 Among others, this approach was taken in the following studies: Gao Jiaying 高家莺, Fan Keyu 范可育 and Fei Jinchang 费锦昌, *Xiandai Hanzixue* 现代汉字学 [Modern Grammatology] (Beijing: Gaodeng Jiaoyu Chubanshe, 1993); Zhang Jingxian 张静贤, *Xiandai Hanzi Jiaocheng* 现代汉字教程 [Tutorial of Modern Chinese Characters] (Beijing: Xiandai Chubanshe, 1992); Andreas Guder-Manitius, *Sinographemdidaktik. Aspekte einer systematischen Vermittlung der chinesischen Schrift im Unterricht Chinesisch als Fremdsprache. Mit einer Komponentenanalyse der häufigsten Schriftzeichen.* [Sinographemdidaktik. Aspects of a Systematic Explanation of Chinese Writing in Teaching Chinese as a Foreign Language. Including a Component Analysis of the Most Common Characters] (Heidelberg: Julius Gross Verlag, 1999); Cornelia Schindelin, *Zur Phonetizität chinesischer Schriftzeichen in der Didaktik des Chinesischen als Fremdsprache. Eine synchrone*

approach investigates whether a character contains a component that indicates its pronunciation and if so, it distinguishes characters according to the level of phonetic effectiveness.<sup>14</sup> Considering the target of this study, the second approach was adopted while analyzing the phonetic motivation. Besides, it has to be pointed out that the mentioned studies significantly differ in terms of the required level of syllabic value adequacy. While some of them recognize as phonetics only those components whose pronunciation is exactly the same as that of the character or differs no more than in tone, the others find correspondence either only in initial or in final, acceptable as well. The present study accepts the second approach, and thus regards as p-components those graphic units whose syllabic value corresponds with the pronunciation of the character at least in initial or in final.

As for the semantic motivation, a semantic characterization of each component was developed to evaluate the semantic link between a component and the character's meaning. In order to achieve this, characters with a common component were gathered together and the meaning of each character<sup>15</sup> was

*Phonetizitätsanalyse von 6.535 gebräuchlichen Schriftzeichen* [On the Phonicity of Chinese Characters in Teaching Chinese as a Foreign Language. A Synchronous Phonetic Analysis of 6,535 Common Characters] (Munich: Iudicium, 2007).

- 14 Among others, this approach was taken in the following studies: Zhou Youguang 周有光, *Hanzi Shengpang Duiyin Bian Cha* 汉字声旁读音便查 [Guide to Phonetics Pronunciation in Chinese Characters] (Changchun: Jilin Renmin Chubanshe, 1980); Wen Wu 文武, »Guanyu Hanzi Pingjia de Ji Ge Jiben Wenti« 关于汉字评价的几个基本问题 [Several Basic Issues on the Evaluation of Chinese Characters], *Yuwen Jianshe* 语文建设 [Language Planning] 2 (1987), 7–12; John DeFrancis, *The Chinese Language. Fact and Fantasy* (Honolulu: University of Hawaii Press, 1984); Yang Hongqing 杨洪清 and Zhu Xinlan 朱新兰, *Kuaisu Shi Zi Zidian* 快速识字典 [Chinese Characters: Quick and Easy] (Nanjing: Jiangsu Guji Chubanshe, 1996); Li Yan 李燕 and Kang Jiashen 康加深, »Xiandai Hanyu Xingsheng Zi Shengfu Yanjiu« 现代汉语形声字声符研究 [A Study of Phonograms in Modern Chinese], in *Xiandai Hanzixue Cankao Ziliao* 现代汉字学参考资料 [A Collection of Papers on Modern Chinese Grammatology], ed. by Su Peicheng 苏培成 (Beijing: Beijing University Press, 2002), 141–154.
- 15 To be specific, meanings mentioned in the two following dictionaries were considered: *Xiandai Hanyu Cidian (Han-Ying Shuangyu)* 现代汉语词典 (汉英双语) [The Contemporary Chinese Dictionary. Chinese-English Edition], ed. by Zhongguo Shehui Kexueyuan Yuyan Yanjiusuo 中国

compared with the component's meaning as described in grammatological dictionaries.<sup>16</sup> The observation showed that the same type of connection often repeatedly occurs in characters with the same component. For example, regarding the component 水 'water', there can be found characters relating to different water sources, types of liquids, qualities of water or activities connected with water. Thus, based on the component distribution, a set of repeatedly used semantic connections that can be considered a core of the component's semantic space was identified.<sup>17</sup> The semantic picture in this way obtained was used as a means of evaluating the motivations of the rest of the characters with this component. Additionally, considering the fact that the motivation in the case of ideograms may be of a more complex nature, and thus not necessarily at the first glance derivable from the basic meaning of components, the above mentioned grammatological dictionaries were used to determine semantic motivations in characters that seemed to be composed of two or more semantic components.

#### 4 *A New Model of Categorization*

The proposed model has a two-dimensional arrangement. It includes five groups subdivided into 20 categories. The group status indicates the decomposition specification, and the category status the nature of the relationship between the graphic and linguistic representation. After a brief description, some examples are given for each category. To locate the position of components in the character

社会科学院语研究所 [Institute of Linguistics, Chinese Academy of Social Sciences] (Beijing: Waiyu Jiaoxue yu Yanjiu Chubanshe, 2002); and *Xinhua Zidian* 新华字典 [Xinhua Dictionary] (Beijing: Shangwu Yinshuguan, 2011).

16 Yang Hongqing and Zhu Xinlan, *Kuaisu Shi Zi Zidian; Hanzi Xing Yi Fenxi Zidian* 汉字形义分析字典 [Analytic Dictionary of Chinese Character Graphics and Meanings], ed. by Cao Xianzhao 曹先擢 and Su Peicheng 苏培成 (Beijing: Beijing Daxue Chubanshe, 1999).

17 This approach was inspired by the study of Shi Zhengyu 施正宇 »Xiandai Xingsheng Zi Xinfu Gongneng« 现代形声字形符功能分析 [An Analysis of Determinatives Function in Modern Chinese Phonograms], *Yuyan Wenzhi Yingyong* 语言文字应用 [Applied Linguistics] 4 (1992), 76–92.

graphics, the following abbreviations are used: L for left, R for right, U for up, D for down, I for inside and O for outside.

#### 4.1 *Group A*

The attribute connecting all the characters in this group is “indivisibility”. Over 96% of these characters are composed of only one graphic element. The remaining 4% can be divided into more than one element; however, their indivisibility is ensured by constructional coherence. As described above, constructional motivation was considered a superior classification criterion in problematic cases.

Group A contains 257 characters altogether. The low number of characters in this group can be considered inversely proportional to their importance. Approximately 80% of group A characters are also used in the Chinese writing system as components. As for their graphic structure, it was observed that, on average, they consist of 4.7 strokes. Given the predominant two-component composition of Chinese characters, this number corresponds to the fact that most of the characters in modern Chinese are comprised of nine strokes.<sup>18</sup>

It can be assumed that at least the initial graphic form of these characters reflected the meaning of the recorded linguistic unit. Considering the development features of the Chinese writing system, it is of no surprise that understanding this connection in relation to the current graphic form requires, in most cases, a more or less extensive etymological explanation. In principle, two basic kinds of disruption of the originally motivated relationship could be identified: the first one is the increasing corruption of the graphic form that culminated during the shift to the clerical script; the second one is the borrowing of a grapheme in order to record another word, mostly of an abstract meaning.

The characters of group A are further subdivided into the following three categories.

##### *Category A1*

This category includes 215 one-element characters whose current graphics do not depict the current meaning of the character. Most of them originated as

18 Yang Runlu 杨润陆, *Xiandai Hanzixue* 现代汉字学 [*Modern Grammatology*] (Beijing: Beijing Shifan Daxue Chubanshe, 2008), 132.

pictograms or symbols. Thus, the following semantic groups can, for example, be identified in this category: characters for different animals (such as *xiàng* 象 ‘elephant’, *shǔ* 鼠 ‘mouse, rat’, *niǎo* 鸟 ‘bird’, *mǎ* 马 ‘horse’), body parts (such as *shǒu* 手 ‘hand’, *shǒu* 首 ‘head’, *ěr* 耳 ‘ear’, *yá* 牙 ‘tooth’, *máo* 毛 ‘hair, feather’), parts of an inanimate nature (such as *huǒ* 火 ‘fire’, *shuǐ* 水 ‘water’, *shí* 石 ‘stone’, *yǔ* 雨 ‘rain’) or artefacts (such as *zhōu* 舟 ‘boat’, *chē* 车 ‘vehicle’, *yī* 衣 ‘clothing’, *mén* 门 ‘door’, *shū* 书 ‘book’, *máo* 矛 ‘spear’, *sháo* 勺 ‘spoon’).

### *Category A2*

The graphic form of these characters reflects the meaning; either pictographically or by means of symbolic strokes. With regard to the total of only 30 items, it can be said that these graphemes are no more than rarities in the writing system of modern Chinese. Considering the current stage of simplification, one could ask if these characters should still be mentioned as a separate category. To answer this question one should consider the fact that the successful identification of the depicted object was the basic requirement assigned to pictograms, not the exact graphic representation of the actual shape. In this sense, even within the modern characters such ones can be found whose graphic form schematically, but still distinctively reflects the recorded meaning.

*tián* 田 ‘field’: Earth’s surface divided by water canals into small fields

*chuān* 川 ‘river’: three strokes, one next to each other symbolize the river current

*rén* 人 ‘person’: a person with one foot forward

*kǒu* 口 ‘mouth’: opened mouth

*yī* 一 ‘one’: one horizontal stroke; *èr* 二 ‘two’: two horizontal strokes; *sān* 三 ‘three’: three horizontal strokes

### *Category A3*

In traditional categorization, these kinds of characters would be considered symbols. Compared with these, the characters of category A2 are limited to those graphemes in whose graphics another independently existing character with exactly one more or one less distinctive stroke could be recognized. Although it was not the original intention, the graphemic analysis has shown that only in these cases can the initial character actually be recognized, and therefore provide

significant information about the meaning or pronunciation of the derived character.

The category A<sub>3</sub> is even less productive than the already described category A<sub>2</sub>. In most of the cases the added strokes somehow specify the meaning carried by the initial character. For example, the stroke added to lower part of the character *mù* 木 ‘tree’ symbolizes the meaning *běn* 本 ‘root’. The only case of stroke reduction can be observed in the graphic form of the character *wū* 烏 ‘crow’, i.e. a black bird whose black eye cannot be seen<sup>19</sup> that was derived from the character *nǐǎo* 鸟 ‘bird’.

#### 4.2 Group B

Group B includes the characters that are not indivisible like group A characters, nor completely divisible into components like group C and D characters. Considering their graphic structure, they can be divided into two or more graphic elements; however, these graphic units do not meet the requirements that would enable the characters to be classified as a group C or D member. Despite this fact, about one third of group B characters contain one s-component, p-component or s/p-component;<sup>20</sup> therefore, the group B characters can be divided into four categories. Nevertheless, the largest category is composed of characters whose specification is just divisibility into two or more elements. Within the analyzed signary, a total of 256 group B characters were identified.

*Category B1*: divisible into two or more elements (158 characters)

*gòng* 共 ‘common, general’

*xià* 夏 ‘summer’

*Category B2*: divisible into two or more elements + contains one s-component (71 characters)

*jǐ* 脊 ‘spine, backbone’: contains component 肉 ‘flesh’

*xuě* 雪 ‘snow’: contains component 雨 ‘rain’

19 *Hanzi Xing Yi Fenxi Zidian*, 553.

20 In most of the cases, the element of the first level decomposition and the motivated component represent the same graphic unit; however, a few exceptions can be identified. For example, the character *lǚ* 旅 ‘travel’ can be divided into two first level elements, one on the left side and the other on the right, but it contains an s-component 旃 ‘flag’ that encloses its other part from the left side.

*Category B3*: divisible into two or more elements + contains one p-component (22 characters)

*cóng* 丛 ‘crowd together’: contains p-component *cóng* 从

乔 *qiáo* ‘tall’: contains p-component *yāo* 夭

*Category B4*: divisible into two or more elements + contains one s/p-component (5 characters)

*méi* 眉 ‘eyebrow’: contains s/p-component *mù* 目 ‘eye’

#### 4.3 Group C

Group C characters can be divided into exactly two components. The total value of 1,937 indicates that these characters cover nearly four-fifths of the 2,500 frequently used characters. Since different types of components can be combined, seven different categories can be identified. The most productive combinational principle is the connection of a p-component and an s-component.

*Category C1*: n-component + n-component (142 characters)

*xiàn* 献 ‘donate’: L 南 R 犬

*zhè* 这 ‘this’: I 文 O 讠

*Category C2*: s-component + n-component (336 characters)

*miào* 庙 ‘temple’: I 由 O 广 ‘shelter’

*zhí* 执 ‘hold’: L 手 ‘hand’ R 丸

*Category C3*: p-component + n-component (166 characters)

*yàn* 验 ‘examine’: L 马 R *qiān* 钅

*tǎng* 倘 ‘if, supposing’: L 人 R *shàng* 尚

*Category C4*: s/p-component + n-component (27 characters)

*huáng* 皇 ‘emperor’: U 白 D *wáng* 王 ‘king’

*xiá* 狭 ‘narrow’: L 犬 R *jiā* 夹 ‘press from both sides’

*Category C5*: s-component + s-component (188 characters)

*zāi* 灾 ‘catastrophe’: U 宀 ‘roof’ D 火 ‘fire’—fire under roof<sup>21</sup>

*míng* 鸣 ‘cry of birds, animals, insect’: L 口 ‘mouth’ R 鸟 ‘bird’—bird with an open beak<sup>22</sup>

21 *Hanzi Xing Yi Fenxi Zidian*, 662.

22 *Ibid.*, 370.

*Category C6*: s-component + p-component (939 characters)

*táo* 桃 ‘peach’: L 木 ‘tree’ R *zhào* 兆

*chuí* 锤 ‘hammer’: L 金 ‘metal’ R *chuí* 垂

*Category C7*: s-component + s/p-component (139 characters)

*zuò* 座 ‘seat’: I *zuò* 坐 ‘sit’ O 广 ‘shelter’<sup>23</sup>

*jiǔ* 酒 ‘alcoholic drink’: L 水 ‘water’ R *yǒu* 酉 ‘wine vessel’ —fluid in a wine vessel<sup>24</sup>

#### 4.4 Group D

These characters consist of three or more components of which at least two are motivated. The total of 33 characters suggest that this type of construction is quite rare. Within the traditional categorization, there is only one category whose definition explicitly mentions the possibility of more than two-component construction, i.e. the category of ideograms. Within group D, the combination of three or more s-components is the most productive one, but is not the only one. Based on the analysis, three more types of combinations were identified. Therefore, group D is subdivided into four categories. Considering the limited productivity, only one example of each category is given.

*Category D1*: n-component + s-component + p-component (1 character)

*qú* 渠 ‘canal’: U 水 ‘water’ + *jù* 巨 D 木

*Category D2*: three or more s-components (20 characters)

*sēn* 森 ‘forest’: three 木 ‘trees’<sup>25</sup>

*Category D3*: two s-components and one p-component (5 characters)

*shì* 饰 ‘decorations’: L *shí* 食 R ‘person’ 人 + 巾 ‘scarf’<sup>26</sup>

*Category D4*: two or more s-components and one s/p-component (7 characters)

*bì* 碧 ‘bluish green’: L 玉 ‘jade’ + *bái* 白 ‘white’ D 石 ‘stone, clear and jade green stone’.<sup>27</sup>

23 *Hanzi Xing Yi Fenxi Zidian*, 723.

24 *Ibid.*, 272.

25 *Ibid.*, 463.

26 *Ibid.*, 484.

27 *Ibid.*, 25.

4.5 *Group E*

One character is generally used to record one monosyllabic morpheme. However, there is a specific group of characters that deviate from this arrangement. The characters with zero meaning, as the name indicates, are the characters that in themselves do not carry any meaning.<sup>28</sup> To do so, they need to combine with another character, that is why they became a part of a two- or sometimes even more-syllable morphemes. This means that they are linked with the language only on the phonological level. Because of the missing semantic value, these characters were automatically excluded from the semantic analysis. Hence, only the potential connection between the pronunciation of a character and some of its parts was examined. Based on this criterion, characters with zero meaning were divided into two categories.

*Category E1*: n-component + n-component (1 character)

*bō* 玻: L 王 R 皮 (→ *bōli* 玻璃 ‘glass’)

*Category E2*: p-component + n-component (16 characters)

*qīng* 蜻: L 虫 *qīng* R (→ *qīngtíng* 蜻蜓 ‘dragonfly’)

The occurrence of these characters in the analyzed signary was quite small; to be specific, only 17 characters were identified. Nevertheless, data mentioned in other studies suggests that a significantly higher concentration of characters with zero meaning can be expected within a group of less frequent characters.<sup>29</sup> Besides, it cannot go unnoticed that even though these characters do not carry their own meaning, at least within the analyzed corpus, an obvious semantic connection to a higher linguistic level, i.e. a connection between one of the components and the meaning of the two-syllable morpheme, was observed. Considering all of this, there is no doubt that these characters require more in-

28 For details on zero meaning characters, see Ma Xianbin 马显彬, *Xiandai Hanzixue* 现代汉字学 [Modern Grammatology] (Guangzhou: Jinan Daxue Chubanshe, 2013), 113.

29 For example, Švarný et. al. state that two syllable bases belong to the layer of less frequent units in language. See Oldřich Švarný et al., *Hovorová čínština v příkladech III*. [Colloquial Chinese in Sentence Examples] (Olomouc: Vydavatelství Univerzity Palackého, 1998), 87. As for the characters with zero meaning, Ma Xianbin states that they cover 14% of the 7,000 commonly used characters. See Ma Xianbin, *Xiandai Hanzixue*, 113.

depth analysis that should be based on a larger sample of characters. That was the reason to assign them a special category, rather than merge them into group C.

### 5 Summary of Results

The following table summarizes the proposed model. The amount of characters belonging to the group or category can be found in parentheses; for every group, the relative frequency as a percentage is noted as well.

Group A indivisible (257, i.e. 10.3 %)	<i>Category A1</i> (215) one unmotivated element
	<i>Category A2</i> (30) pictographic or symbolic reflection of the meaning
	<i>Category A3</i> (12) existing character ± one distinguishing stroke
Group B divisible into graphic elements (256, i.e. 10.2 %)	<i>Category B1</i> (158) without any motivated component
	<i>Category B2</i> (71) with one s-component
	<i>Category B3</i> (22) with one p-component
	<i>Category B4</i> (5) with one s/p-component
Group C divisible into exactly two components (1,937, i.e. 77.5 %)	<i>Category C1</i> (142) n-component + n-component
	<i>Category C2</i> (336) s-component + n-component
	<i>Category C3</i> (166) p-component + n-component
	<i>Category C4</i> (27) s/p-component + n-component
	<i>Category C5</i> (188) s-component + s-component
	<i>Category C6</i> (939)

	s-component + p-component
	<i>Category C7</i> (139)
	s-component + s/p-component
Group D divisible into two motivated components and at least one other component of any kind (33, i.e. 1.3 %)	<i>Category D1</i> (1)
	s-component + p-component + n-component
	<i>Category D2</i> (20)
	three or more s-components
	<i>Category D3</i> (5)
	two s-components + one p-component
	<i>Category D4</i> (7)
	two or more s-components + s/p-component
Group E characters with zero meaning (17, i.e. 0.7 %)	<i>Category E1</i> (1)
	n-component + n-component
	<i>Category E2</i> (16)
	p-component + n-component

Table 1  
*The Proposed Model of Categorization.*

As per the table above, group C was identified as the most productive composition type in the analyzed signary of the modern Chinese characters. The total amount of 77.5 % characters proves that the graphics of most of the characters contain exactly two components. Moreover, the most productive category of group C is actually the most productive category overall: combinations of one s-component and one p-component (category C6) cover 48.5% of group C characters and 37.6% of analyzed signary. Therefore, it can still be stated that the dominant constructional pattern of minor script characters observed in Xu Shen's *Shuo Wen Jie Zi* is still being preserved in the writing system of modern Chinese. However, a significant decrease in occurrence cannot be overlooked: in comparison to their relative frequency in *Shuo Wen Jie Zi*,<sup>30</sup> the current percentage

30 According to the statistics provided by Qing-dynasty scholar Wang Yun 王筠 (1784–1854), there can be found 264 (2.8%) pictograms, 129 (1.4%) symbols, 1,254 (13.4%) ideograms and 7,697 (82.3%)

value is less than one half. The second most productive category too can be found within group C. Characters of category C2 that combine one s-component and n-component cover 13.4% of the analyzed signary. This implies that the neutralization of motivation represents an important development tendency.

Although with a difference of only one item, the characters indivisible into smaller graphic units represent the second most productive group. To be specific, group A characters cover 10.3% of the analyzed signary. Within Xu Shen's classification system, pictograms and symbols are the characters that are characterized by indivisibility into smaller constructional units. The identified number of group A characters is six times higher than the sum of pictograms and symbols in *Sbuo Wen Jie Zi*. Therefore, the fusion of original components into one graphic unit needs to be considered as one of the important types of graphic form corruption. On the other hand, the almost identical coverage rate of group B characters indicates that another type of corruption is rather productive as well.

The new model of categorization reflects the characteristics of the modern Chinese writing system. Among other things, conclusions about the current amount of motivation can be drawn. With regard to this, three different types of characters can be distinguished: fully motivated characters whose entire graphics provide a link to the meaning or pronunciation; partially motivated characters whose one part does not provide any phonetic or semantic link; and finally, the fully unmotivated characters whose current graphics show no relation to the recorded morpheme.

As can be seen in the table, characters with fully motivated and partly motivated composition are spread across 8 different categories each. To be specific, the categories of fully motivated characters are as follows: A2, A3, C5, C6, C7, D2, D3 and D4. Altogether they cover 53.6% of the analyzed 2,500 frequently used Chinese characters. The characters with partly motivated composition can be found in the following categories: B2, B3, B4, C2, C3, C4, D1 and E2. By summing up their occurrences, a coverage rate of 25.8% was obtained. In other words, these numbers mean that despite the extensive corruption of character

phonograms in *Sbuo Wen Jie Zi*. See Yu Guoqing 余国庆, *Sbuowenxue Daolun* 说文学导论 [Introduction to research on The Structure Analysis of Primary Characters and Meaning Explanation of Secondary Characters] (Hefei: Anhui Jiaoyu Chubanshe, 1995), 51–58.

graphics and language changes, the composition of more than three fourths of 2,500 frequently used characters can still be considered at least partially motivated. Quantitatively speaking, it can be observed that the semantic motivation is more productive than the phonetic one. While the analyzed signary contains 69.5% characters with at least one s-component; the characters with p-component cover 53.1%.

### *Conclusion*

This analysis has shown that due to the corruption of the original graphic form, one cannot avoid employing both the constructional and structural approach while categorizing modern Chinese characters. Only a combination of both approaches enables us to establish a comprehensive classification. Nevertheless, it is important to notice that the constructional approach is considered the superior one, while structural decomposition is limited to the cases where a character cannot be divided into components following the defined criteria. The important difference from Su Peicheng's categorization is that the two approaches are not blending together, but either one or the other is being applied at a certain stage of the classification process. Despite attempting not to establish categories with few characters, several of them can be found in the model. The reason for this is that even though the analyzed characters are supposed to cover over 97% of a common Chinese text, the analysis involved only about one third of the modern Chinese signary. As such, the presented model has to be understood as a prototype that in the first place aims to outline a most complex spectrum of possibilities and to provide a basis for examining a larger data sample. Until an analysis of the complete modern Chinese signary is conducted, the definite version of the new categorization model cannot be presented. Even though the general outline of the model appears to sufficiently reflect the characteristics of the currently used writing system, some modification can be expected.

*Palacký University in Olomouc, Department of Asian Studies*