# On the Nature of Unmotivated Components in Modern Chinese Characters

Tereza Slaměníková

*Abstract.* From an etymological perspective, the graphics of Chinese characters are in general supposed to encode at least semantic, but primarily both semantic and phonetic information concerning the recorded linguistic unit. This attribute of the Chinese writing system is often pointed out, even when referring to the composition of the graphemes used in modern Chinese signary. A careful look, however, at the individual characters suggests that, in view of the current meaning or sound of the characters, the relationship between the graphic and linguistic structure might be partly or entirely missing. This means, in other words, that apart from semantically and phonetically motivated components, unmotivated constituents can be identified in the composition of modern Chinese characters as well. Although the phenomenon of unmotivated constituents has been discussed in a number of grammatological studies, it is often viewed as a peripheral issue. This paper argues that these units deserve much more attention than they have so far received. Based on a new model of the classification system for Chinese characters, it demonstrates that there are two different types of unmotivated constituents to be distinguished, and thus it provides deeper insight into the characteristic features of the modern Chinese writing system.

## 1. Introduction

Chinese characters represent the oldest, uninterruptedly used, writing system in the world. Over the course of its development, the graphic form of the characters underwent radical changes influencing the basic characteristic features of the writing system. The corruption of the

Tereza Slaměníková   0000-0001-6929-7568
Department of Asian Studies, Palacký University in Olomouc
Czech Republic
tereza.slamenikova@upol.cz

graphics of the characters and the changes in the Chinese phonologi-
cal system as well as in the semantic content of the recorded linguistic
units contributed to the disruption of the originally motivated relation-
ship between the graphic and linguistic structures (cf. Schindelin in this
volume). The currently used system therefore represents a certain mix-
ture of graphemes with a different level of motivation. Despite this fact,
it is not unusual for researchers to employ the system created almost
2,000 years ago when categorizing the modern Chinese characters.

The so-called six-category classification was described in the old-
est known Chinese grammatological study 说文解字 *Shuō Wén Jiě Zì* [The
Meaning Explanation of Primary Characters and the Structure Analy-
sis of Secondary Characters]. Its author, the Han-dynasty scholar 许慎
Xǔ Shèn, conducted a thorough graphemic analysis of more than nine
thousand characters in order to explain the relationship between their
graphics and the recorded linguistic units. The definitions used also
placed each character into one of the categories referring to the con-
structional method employed by their creating. As the title of the book
indicates, there are two basic types of characters to be distinguished:
(1) those with a simple graphic form represented by 象形 *xiàngxíng* pic-
tograms (schematic depictions of the objects they represent) and 指事
*zhǐshì* symbols (characters expressing a certain idea through symbolic
strokes), and (2) those with a compound graphic form represented by 会
议 *huìyì* ideograms (combinations of two or more semantic components)
and 形声 *xíngshēng* phonograms (combinations of one semantic and one
phonetic components).[1] The last two of the six categories mentioned in
the Postface of *Shuo Wen Jie Zi*, i.e., 假借 *jiǎjiè* loanwords and 转注 *zhuǎnzhù*
variants, refer to the new usage of an existing character[2] and as such are
not identifiable in the definitions of characters.

Considering its significant role in the Chinese grammatological tra-
dition, it is not surprising that Xu Shen's legacy continues to shape the
field up until the present. One should not, however, overlook the fact
that the application of the traditional six categories to the modern Chi-
nese characters suffers from certain limitations. The main reason for this
is the fact that a comprehensive synchronic description of the graphic
form composition of the characters is impossible without involving a
new type of constructional units in the system, i.e., units that do not
provide a link to the pronunciation or to the meaning of the recorded
linguistic units. There are consequently two more basic types of char-
acters to be distinguished: (a) with partly, and (b) with entirely unmo-
tivated graphics.

---

1. Grammatologist are not in full agreement regarding the English equivalents of
the six categories. This paper adopts terms used by Uher (2013, pp. 297–303) in the
English translation of the Postface to *Shuo Wen Jie Zi*.

2. For details see Dong (1994, pp. 26–27).

The etymological explanation, by means of the six-category system, has represented an immutable classification paradigm for hundreds of years. Although researchers have often differed significantly in their views, not only on the theoretical concept of the categories as such, but especially on the classification of the individual characters, they never exceeded the established dogma.[3] An important shift in approach occurred in the first half of the twentieth century[4], when the palaeographer 唐兰 Táng Lán realized that the solution to the grammatological development crisis could be found in reducing the number of categories. His idea of adapting the traditional system inspired other grammatologists to propose their own modifications to the traditional model, particularly of importance being those developed by 陈梦家 Chén Mèngjiā, 裘锡圭 Qiú Xīguī or 王宁 Wáng Níng.[5]

According to another grammatologist 王凤阳 Wáng Fèngyáng (1989, p. 490), it was again Tang Lan who first used the term 记号字 *jìhào zì* unmotivated characters while referring to graphemes with corrupt graphics. It was Wang Fengyang himself, however, who in his extensive study highlighted the fact that the graphic composition of a significant number of Chinese characters no longer maintained the connection with their pronunciation or meaning. The already mentioned paleograph Qiu Xigui also paid significant attention to the issue of graphic form demotivation. His particular contribution lies in identifying the increase in the number of unmotivated characters and 半记号字 *bàn jìhào zì* partly unmotivated characters[6] as one of the three major changes in the devel-

---

3. For details see Qiu (1988, p. 103).

4. According to Su (2007, pp. 2–3), the increasing interest in the characteristic features of the currently used writing system at the beginning of the twentieth century had its roots in two main factors. It was first influenced by the efforts of young Chinese intellectuals to reform traditional Chinese society, since the modernization of the written Chinese language was one of their main requirements. The educational system radically changed as a result: the classical texts were replaced with texts written in colloquial language. The use of the written language was no longer the privilege of a limited amount of state officials and therefore the Chinese writing system had to face an increasing number of its users. Secondly, in light of the technological progress, a need arose to make the Chinese writing system accessible to new products, such as printing, typing machines, telegraphs, etc.

5. Descriptions of these classification models can be found in Tang (2001, pp. 59–98), Chen (2006, pp. 24–94, 256–258 & 354), Qiu (1988, pp. 97–204) and N. Wang (2001, pp. 63–82).

6. The literal translation of the Chinese word 记号 *jìhao* is 'mark, sign'. In light of the fact that these English terms are overloaded with various meanings used in different contexts and thus might be rather misleading, the indirect expression 'unmotivated constituent,' which reflects the nature of these units in the Chinese writing system, was used in this paper. The derived terms of the unmotivated characters and partly unmotivated characters refer to graphemes whose graphics demonstrate no or only a partial relationship to the represented linguistic unit.

opment of the constructional composition of the characters as well as in putting these characters into the context of the general development of the Chinese writing system.

Although Qiu Xigui, as well as, for example, Wang Ning, discussed the issue of unmotivated and partly unmotivated characters in their studies, keeping them outside their basic classification scheme. The protagonists of modern Chinese grammatology, a new autonomous discipline that crystallized in the 1980s, undertook such an innovative step. 周有光 Zhōu Yǒuguāng, who is considered the founder of this newest branch of Chinese grammatology[7], claimed that its interest lies strictly in the currently used form of the Chinese writing system, the so-called modern Chinese characters. The question as to how the graphics of modern Chinese characters relate to their pronunciation or meaning naturally ranks among its main topics. In this context, the most recognized model of the new classification was introduced by by one of the leaders of modern Chinese grammatology 苏培成 Sū Péichéng.[8] The so-called new six categories have opened up an alternative synchronic perspective on the etymologically oriented traditional classification system.

Over the years, Su Peicheng introduced three different versions of the new classification. In the first one, published in the oldest edition of his book 现代汉字学 *Xiàndài Hánzìxué* [Modern Chinese Grammatology] (1994, pp. 72–80), he distinguished seven categories of characters. The first three categories include characters whose graphic form can still be viewed as fully motivated. Two of them were adopted from the traditional categorization, i.e., ideograms and phonograms. The third one contained characters whose graphics somewhat depicts their current meaning, that is the principle employed in Xu Shen's pictograms and symbols. The other four categories were newly implemented into the classification system, together with a new type component, i.e., an unmotivated constituent.[9] Two of these four categories are partly mo-

---

7. For details see Su (2007, p. 3).

8. Su Peicheng himself provides examples of two other classifications introduced by Qian (2001) and Hao (1994). Classifications similar to Su Peicheng's model, however, differing in the organization of the categories in terms of the subordination of several categories under a superior group with similar qualities can be found, for example, in studies presented by R. Yang (2008) or G. Gao (2002). Interesting models were also introduced by H. Yang and Zhu (1996) and Pan (2003).

9. The word 'constituent' in the English translation was chosen based on two factors. First, Su Peicheng distinguishes three types of 字符 *zìfú*, basic constructional units called components in this paper. Two of them are attributive compounds with the head 符 *fú*, i.e., 音符 *yīnfú* phonetic components and 意符 *yìfú* semantic components, and the third one 记号 *jìhào* was created with a different word formation principle. Second, the terminology used is somewhat inconsistent. Moreover, not all of the researchers use different terms when referring to the structural and constructional decomposition possibilities of the characters (for details see below). This is, for ex-

tivated, since they are combinations of an unmotivated constituent and a semantic or phonetic component. Finally, two categories with a completely unmotivated graphic form can be identified. One of them consists of simple unmotivated characters indivisible into smaller graphic units, while the other one involves complex unmotivated characters composed of two or more unmotivated constituents.

Unlike the first version, the two revised versions only employ six categories of characters.[10] In the second edition of *Modern Chinese Grammatology* (2001, pp. 93–101), the category of simple characters depicting the meaning was excluded. It is listed again in the third edition (2015, pp. 102–111) and the six-category arrangement is achieved this time through a fusion of the simple and complex unmotivated characters into one category. In this manner, Su Peicheng once again implemented all the first-version categories into the classification. What is rather unfortunate is the fact that he does not provide any explanation that could help one understand the reasons for the repeated modification of the classification system.

The system of the new six categories is an unquestionable shift towards a synchronic approach to the relationship between the graphic and linguistic structure. It is apparent that, over the course of time, Su Peicheng gave this issue serious consideration in order to achieve a more efficient classification system. Regrettably, however, one cannot fail to notice that the definitions of the basic constructional units are somewhat shallow. Su Peicheng is quite specific concerning the requirements for phonetic components, however, the parameters that need to be present for a graphic unit to be considered a semantic component or an unmotivated constituent are too general to provide the required information for evaluation. It should also be mentioned that while describing the classification system, Su Peicheng avoids stating how many characters belong to each category. One might therefore ask whether he actually verified the applicability of the proposed system through an in-depth analysis of a representative sample of modern Chinese signary.

As the title indicates, this paper primarily focuses on the issue of unmotivated constituents.[11] The problem concerning their definition is connected with the fact that modern grammatology emphasizes the im-

---

ample, the case of Wang Ning who uses one universal term 构件 *gòujiàn* and thus does not establish so clear a line between the structural or constructional approach. The word 'constituent' is supposed to prevent over-interpretation in terms implying one or the other approach.

10.  One cannot fail to notice that, through the exclusion of one category, Su Peicheng has reached an identical number of categories as can be found in Xu Shen's classification system.

11.  To describe in simple fashion the difficulty relating to semantic component, it is the rejection of the diachronic approach, one of the basic requirements of modern Chinese grammatology (Zhou 2004, pp. 306–316; Su 2001b, pp. 92–93; Su 2001a,

portance of a setting strict boundary between the constructional and the structural approach to the decomposition of Chinese characters. The first one explores the connection between the graphics of the characters and the meaning or sound of the recorded linguistic unit. The structural approach, in contrast, is strictly interested 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.[12] In view of this, the status of unmotivated constituents seems to be somewhat problematic since it is a constructional unit but carries no useful information concerning the character's meaning or pronunciation. Considering the graphemic analyses conducted by other researchers, the lack of a definition concerning the unmotivated constituent lies in the fact that it does not specify how deep the decomposition is supposed to be carried out and thus an uncontrolled blending with the structural decomposition methods is inevitable.[13]

This paper discusses the issue of unmotivated constituents on the basis of a new classification model that was proposed considering the above-mentioned limitations. The model was developed as an attempt to provide a system that will as effectively as possible reflect the current features of modern Chinese characters.[14] It should be emphasized that it has been primarily associated with searching for a solution to the issue of unmotivated constituents. To put it into a time context,

---

pp. 359–360; Su 2015, pp. 101–102), that raises the question as to whether it is still an etymological explanation that should be used as an evaluation device.

12. Different terms are used when referring to the basic units of these two approaches: the element is the basic unit of the structural approach, while the component is the basic unit of the constructional approach. The term (graphic) unit is used as a general term for part of the character without any further implications.

13. In this respect, 快速识字字典 *Kuàisù Shí Zì Zìdiǎn* [Chinese Characters: Quick and Easy] (H. Yang and Zhu, 1996) can be taken as a representative example. As the title indicates, it is a dictionary and as such provides a graphemic analysis of modern Chinese graphemes, not "merely" proposing a certain kind of theoretical model, such as the above mentioned classifications mostly do. Although it demonstrates a high level of systematicity and undoubtedly serves its pedagogical purpose well, in case of the unmotivated parts, the authors tend to decompose them into minimal possible graphical units, such as, for example, the character 带 *dài* 'belt' which is according to the authors composed of the semantic component 巾 'cloth' and two unmotivated constituents 卅⌒ (p. 47); or the character 令 *lìng* 'command' which is composed of three unmotivated constituents 人丶 and ▽ (p. 164). The question that arises is whether it is, in the case of the constructional approach, reasonable to decompose characters into such small parts.

14. The new model was introduced in my dissertation at Palacký University in Olomouc, which has been published, in a revised version, under the title *Čínské znakové písmo: synchronní model tradiční kategorizace* [The Chinese Writing System: A Synchronic Model of the Traditional Categorization] (2017). A brief description of the model in English can be found in Slaměníková (2017).

it was elaborated before the third version of Su Peicheng's categorization was published, i.e., at the moment when the simple and complex unmotivated characters each had its own separate category. The new model is based on the graphemic analysis of the so-called 2,500 常用字 *chángyòng zì* frequently used characters.[15] During the analysis procedure, it was discovered that due to the extent and diversity of the corruption of the original form, one cannot avoid employing both the constructional and structural approach while categorizing modern Chinese characters. Only a combination of both approaches enables the establishment of a comprehensive classification. Nevertheless, when considering the primary interest in examining the relationship between the graphic and linguistic representation, the constructional principle is considered the superior one, while methods of structural decomposition are applied as supplementary tools to make an adequate processing of all the characters possible. As will be demonstrated, an important difference from Su Peicheng's categorization lies in the fact that the two approaches do not blend together, but either one or the other is applied.

## 2.   Two Types of Unmotivated Constituents

The new model of categorization has a two-dimensional arrangement. It 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 entire character and its components in terms of semantic and phonetic motivation. It should be pointed that it was the unmotivated parts of the graphics of the characters that significantly determined its final arrangement. By means of the analysis, two types of unmotivated constituents were identified:

(a) those that are not motivated in a particular character, however, they are used as phonetic and/or semantic components in other characters within the modern Chinese signary[16]. For example, the graphic

---

15. As concerns the representativeness of the sample, it should be pointed out that even though the analyzed characters cover less than one third of the currently used signary, represented by the 现代汉语通用字表 *Xiàndài Hànyǔ Tōngyòng Zìbiǎo* [Table of the Commonly Used Modern Chinese Characters] with a total amount of 7,000 graphemes, one cannot overlook the fact that, in the view of the high occurrence rate in modern Chinese texts, they undoubtedly stand for the core of the modern Chinese signary. From a qualitative point of view, the list of characters compiled on the basis of another classification criterion firstly does not eliminate the characters of a particular principle a priori, and secondly maintains a certain proportion of possible constructional principles.

16. To be specific, the occurrence within characters listed in *The Table of the Commonly Used Modern Chinese Characters* was taken into consideration.

unit 巾 in the character 帮 *bāng* 'help' does not provide any useful link to the recorded morpheme, but is used as a semantic component in 帐 *zhàng* 'curtain,' 帽 *mào* 'hat' or 帆 *fān* 'sail'. A certain situation can be observed in the case of the graphic unit 巨 in the character 柜 *guì* 'cabinet' which is, however, used as a phonetic component in the characters 距 *jù* 'distance' or 炬 *jù* 'torch'

(b) those that appear neither as meaning nor as pronunciation indicators at all, such as for example the graphic unit 龶 used in the characters 責 *zé* 'duty,' 素 *sù* 'plain' or 青 *qīng* 'green'; or the graphic unit 皃 used in the character 貌 *mào* 'appearance'.

In order to describe the two types of unmotivated constituents in detail, there is a need to pay attention to the motivated parts of the characters first. The synchronic point of view created a need to reconsider the definition of what kind of graphic units can be labeled as components. Apart from specifying the requirements on what is considered a semantically or phonetically motivated part of the character, one more parameter was added to the synchronic definition of the component, this being the recurrence. This means that only graphic units that occur as a semantic or phonetic indicator in at least two characters in the modern Chinese signary, represented by the 7,000 commonly used characters, were considered components.[17]

Various studies were consulted in order to define the criteria that a graphic unit had to meet to be considered an effective phonetic component in relation to the current pronunciation of the character or to be considered an effective semantic component in relation to its current meaning. As concerns the phonetic motivation, a great variability of methods, achieving noticeably different results, were observed. Simply speaking, the two basic approaches can be identified, when considering the graphic level which is being targeted. The first one focuses on phonetics that are classified according to the relationship between their syllabic value and the syllabic value of all the characters where they occur.[18] The second approach, in contrast, functions the other way around. It focuses on characters since it examines whether a character contains a component indicating its pronunciation. These characters are there-

---

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

18. The following basic types of phonetics are distinguished: (a) ideal phonetics whose pronunciation is identical with all the characters they occur in; (b) phonetics with regular differences whose pronunciation deviate in a systematic manner; and in case of an indulgent approach also (c) irregular phonetics with an unsystematic relationship to the character's pronunciation. This approach was employed, among others, by J. Gao, Fan, and Fei (1993); Zhang (1992); Guder-Manitius (1999); Schindelin (2007); Haralambous (2013).

fore sorted based on the level of phonetic component effectiveness.[19] Unfortunately, what is common to both approaches is that researchers significantly differ in terms of the required level of syllabic value adequacy between the character and its phonetics. While some of them recognize as phonetics only those components that share exactly the same pronunciation as the character or differ no more than in tone, the other considers a correspondence either in the initial syllable or the final acceptable. Considering the target of this paper, the second approach was adopted when analyzing the phonetic motivation. As for the required syllabic value adequacy, graphic units with a correspondence at least in the initial or final were considered phonetically motivated. The reason for adopting this broader viewpoint was the fact that even in Xu Shen's Postface the required level of adequacy is not specified.

When comparing the phonetic motivation, the nature of the synchronic connection between the character graphics and the meaning does not appear to be examined almost at all. Although a wide spectrum of different handbooks can be found analyzing the semantic relationship between single characters and their components[20], the general theoretical implications are rarely discussed. This was the reason why a complex semantic characterization of each component was developed to evaluate the semantic link between a component and the meaning of a particular character.[21] In order to achieve this, characters with the same component were gathered together and the meaning of each character[22] was compared with the component's meaning as described in the grammatological dictionaries.[23] It has been observed that the same type of connection often repeatedly occurs in characters with the same component. To provide an example, under characters with the component 钅 (金) 'metal,' there can be found those referring to the following four main semantic classes: types for metals (e.g., 铜 *tóng* 'copper,' 锡 *xī* 'tin,' 铅 *qiān* 'lead,'), different metal objects (e.g., 锤 *chuí* 'hammer,' 镰 *lián* 'sickle' 锁 *suǒ* 'lock,' 钉 *dīng* 'nail,' 锣 *luó* 'gong,' 锅 *guō* 'pot,' 链 *liàn* 'chain'), activi-

---

19. This approach was taken, among others, by Zhou (1980); Wen (1987); Defrancis (1984); H. Yang and Zhu (1996); Li and Kang (2002).

20. See e.g., Ye (2008); Huang and Ao (2009); H. Yu and Ch. (2010). Nevertheless, speaking of the semantic relatedness between characters and their constituents, an interesting approach of Haralambous (2013) has to be mentioned, who introduced an enhanced model for sinographic language processing. By exploring the semantic information stored in the so called subcharacters, he used three different WordNets.

21. This approach was inspired by the study of Shi (1992, pp. 76–92).

22. Specifically, the meanings mentioned in the two following dictionaries were considered: 现代汉语词典（汉英双语）*Xiàndài Hànyǔ Cídiǎn (Han-Ying Shuangyu)* [The Contemporary Chinese Dictionary. Chinese-English Edition] 2002; and 新华字典 *Xīnhuá Zìdiǎn* [Xinhua Dictionary] 2011.

23. Two dictionaries in particular were used: *Hanzi Xing Yi Fenxi Zidian* (Cao and Su, 1999) and *Kuaisu Shizi Zidian* (H. Yang and Zhu, 1996).

ties connected with the use of a metal object (e.g., 锻 *duàn* 'forge,' 销 *xiāo* 'melt,' 铸 *zhù* 'cast, found') and qualities of metal (e.g., 锐 *ruì* 'sharp').

Thus, based on the component distribution, a set of repeatedly used connections was identified representing the core of the component's semantic network. The semantic picture obtained in this manner was used to evaluate the motivation of the rest of the characters with this component. This method was applied in case of all characters containing one semantic component. When considering the complex nature of the semantic link in characters composed of two or more semantic components[24], the already mentioned grammatological dictionaries were used to determine their motivation. An emphasis, however, on a clear link with the current meaning of the characters was placed in the evaluation process.

Following the described procedure, a set of 613 phonetically used graphic units and a set of 270 semantically used graphic units were identified. Since some of them may possess both a semantic and phonetic function, the final list of all components includes 778 items. When considering their function in particular characters, four types of components can be distinguished: (a) phonetic components, abbreviated as the p-component; (b) semantic components, abbreviated as s-components; (c) components that provide both a semantic and phonetic link to the morpheme represented by the character, abbreviated as s/p-components[25]; and finally (d) graphic units listed in the set of 778 components that in a particular case of occurrence do not possess any semantic or phonetic function, i.e., it can be argued that this function was neutralized, this being the reason for calling them n-components.[26]

---

24. The complexity of the semantic link in these characters lies in the fact that it may not be derivable from the separate meanings of the individual components. This is because it is encoded as a combination of two semantic entities and thus associated with a more complex association process. For details, see Slaměníková (2013).

25. Components with both a semantic and phonetic function can be found in *Shuo Wen Jie Zi* (Xu, 1963) even though none of the definitions of the categories mentions how the characters containing this component should be evaluated. It is therefore not surprising that there is no consensus regarding their classification: some researchers consider them ideograms, others phonograms (Dong, 1994, p. 21). A previous analysis of ideograms has shown that in the case of 68% commonly used characters, which are according to *Hanzi Xing Yi Fenxi Zidian* composed of two semantic components, one of these is classified as a phonetic component as well (Slaměníková, 2013). Thus, considering the fact that the combination of one semantic component and one component with both a semantic and phonetic function is more common than the combination of two semantic components, it is my opinion that components with both functions should be considered as a specific type of components.

26. While providing examples of the characters of the categories described below, the four types of components are distinguished as follows: meaning can be found behind an s-component; pronunciation can be found behind a p-component; both mean-

N-components represent the above-mentioned first type of unmotivated constituents. Analysis has shown, however, that the principle of in/divisibility into components itself is not enough to create an effective classification system. A significant number of characters can be found with complex graphics, apparently composed of more than one graphic unit, yet, one of them or all of them, do not match the criteria to be considered a component. Another aspect had to therefore be implemented into the model in order to ensure that characters with an obvious different level of composition complexity will be divided into separate groups. This was the moment when the unmotivated units belonging to the second type were taken into consideration.

## 3.   Unmotivated Constituents in the New Model

Based on the statistics provided by the Qing-dynasty scholar Wang Yun (G. Yu, 1995, pp. 51–58), 264 (2.8%) pictograms can be found, 129 (1.4%) symbols, 1,254 (13.4%) ideograms and 7,697 (82.3%) phonograms in *Shuo Wen Jie Zi*. It is apparent that most of the characters listed in Xu Shen's work were created as compositions of two components.[27] In view of this fact, all the characters were first examined in terms of the possibility of the decomposition into two components. Following the above described requirements on components, it has been determined that despite the graphic form corruption and other development changes, two-component arrangement still represented the dominant construction principle. The difference, however, is the wider range of possible combinations resulting from the four-type classification system of components. Altogether, seven different combinations were identified, each of them representing one category in the proposed classification model. The letter C refers to the fact that the two-component characters represent the third group in terms of the graphic form complexity (i.e., group C). The numbers in abbreviations indicate the absolute frequency of occurrence within the analyzed signary. To locate the position of the components in the character 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.

---

ing and pronunciation can be found behind an s/p-component; no additional information can be found behind an n-component.

27. Based on Xu Shen's definitions, ideograms represent the only category of characters that can be composed of more than two components. The occurrence, however, of three or more-component ideograms is quite small. Specifically, an analysis of ideograms within the *Table of the Commonly Used Modern Chinese Characters* has shown that less than 5% of the 1,241 identified ideograms are composed of three or more components (ibid.).

*Category C1*: n-component + n-component (142 characters)
    猜 *cāi* 'guess': L 犭 (犬) R 青
    遗 *yí* 'inherit': I 贵 O 辶
*Category C2*: s-component + n-component (336 characters)
    鹊 *què* 'magpie': L 昔 R 鸟 'bird'
    海 *hǎi* 'sea': L 氵 (水) 'water' R 每
*Category C3*: p-component + n-component (166 characters)
    辅 *fǔ* 'assist': L 车 R 甫 *fǔ*
    常 *cháng* 'often': U 尚 *shàng* D 巾
*Category C4*: s/p-component + n-component (27 characters)
    皇 *huáng* 'emperor': U 白 D 王 *wáng* 'king'
    银 *yín* 'silver': L 钅 (金) *jīn* 'metal' R 艮
*Category C5*: s-component + s-component (188 characters)
    库 *kù* 'warehouse'; O 广 'shed' I 车 'vehicle' (Cao and Su, 1999, p. 662)
    岩 *yán* 'rock, cliff'; U 山 'mountain' D 石 'stone' (ibid., p. 295)
*Category C6*: s-component + p-component (939 characters)
    筐 *kuāng* 'basket': U 竹 'bamboo' D 匡 *kuāng*
    裙 *qún* 'skirt': L 衤 (衣) 'clothing' R 君 *jūn*
*Category C7*: s-component + s/p-component (139 characters)
    箩 *luó* 'basket': U 竹 'bamboo' D 罗 *luó* 'net'—basket woven from grass
      (ibid., p. 610)
    泡 *pào* 'bubble, blister': L 氵 (水) 'water' R 包 *bāo* 'bag'—bag full of water
      (ibid., p. 348)

The total value of 1,937 characters indicates that the two-component characters cover nearly fourth-fifths of the 2,500 frequently used characters. The most productive combinational principle is the connection between the p-component and the s-component. It can therefore be stated that the dominant constructional pattern of the minor script observed in Xu Shen's *Shuo Wen Jie Zi* is still being preserved in the writing system of modern Chinese. A significant decrease in occurrence cannot be overlooked, however, since the current percentage value is less than 40%. The second most productive category is represented by characters that combine one s-component and one n-component. This implies the significance of the unmotivated constituents in modern Chinese signary.

After sorting out the two-component characters, the rest of the characters were examined. Two more specific groups of characters were separated: group D including three or more-component characters[28] and group E including the so-called characters with zero meaning[29]. Neither

---

28. For example the character 狱 *yù* 'lawsuit' which is composed of 讠 (言) 'words' between two 犭 (犬) 'dogs' expressing the meaning that two dogs are fighting each other (Cao and Su, 1999, p. 652).

29. Characters with zero meaning represent a specific group of graphemes that deviate from the general arrangement between graphic and linguistic units in Chinese

of the characters cover more than 1.5% of the analyzed signary. They are therefore only briefly mentioned in this paper. The remaining one fifth of the characters are going to be further discussed in the paper.

At first glance, these characters display a high diversity of composition complexity. Although the two-component decomposition based on the above described principles of motivation, is impossible, a number of them can be apparently divided into smaller graphic units that repeatedly occur in the Chinese writing system. To distinguish these characters from those with a single indivisible graphic form, the decomposition method of the structural approach was applied. Specifically, the dictionary 汉字信息字典 *Hànzì Xìnxī Zìdiǎn* [Dictionary of Chinese Character Information][30] was used to determine the divisibility into graphic elements. It has been determined that almost exactly one half of these characters can be divided in two or more elements and one half cannot. The first mentioned are included in group B and the other in group A.

Within the analyzed sample, a total of 256 group B characters were identified. Despite the fact that the graphic units composing these characters do not meet the requirements that would enable the characters to be classified as a group C member, about one third of the group B characters contains one s-component, p-component or s/p-component. Four different categories can therefore be identified within the group B characters. The largest category is composed, however, of characters whose specification is only divisible into two or more elements. Considering the fact that the category status is supposed to reflect the nature of the relationship between the graphic and linguistic form, there was no need to separate the category of characters containing the n-component.

*Category B1*: divisible into two or more elements (158 characters)
　　能 *néng* 'can, be able': composed of elements 厶月匕匕
　　建 *jiàn* 'construct': composed of elements 廴聿
*Category B2*: divisible into two or more elements + contains one s-component (71 characters)
　　句 *jù* 'sentence': composed of elements 勹口, the second one functions as an s-component 口 'mouth'

---

since they, unlike other characters, do not carry any meaning. In order to do so, they need to be combined with another character and become part of a two- or more-syllable morpheme. They themselves are thus linked with the language only on the phonetic level. This is, for example, the case with the characters 菠 *bō* or 蜻 *qīng*. The first is used as part of two-syllable morphemes 菠萝 *bōluó* 'pineapple' and 菠菜 *bōcài* 'spinach'; the second one is used along with another character with a zero meaning as part of the two-syllable morpheme 蜻蜓 *qīngtíng* 'dragonfly'.

30. This dictionary was chosen since it represents the source from which modern Chinese grammatologist often quote statistical data about the structural composition of modern Chinese characters (e.g., Su 2001a, pp. 331−332, 350−352, 428; Su 2015, pp. 97−98; Ma 2013, pp. 85, 113−114, 220−221; R. Yang 2008, pp. 133−134).

骨 *gǔ* 'bone': composed of elements 冎月, the second one functions as an s-component 月 (肉) 'flesh'

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

齿 *chǐ* 'tooth': composed of elements 止人凵, the first one functions as a p-component 止 *zhǐ*

聚 *jù* 'gather, get together': composed of elements 耳又乑, the combination of the first and second element functions as a p-component 取 *qǔ*

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

眉 *méi* 'eyebrow': composed of elements 尸目, the second one functions as an s/p-component 目 *mù* 'eye'

贵 *guì* 'expensive': composed of elements 贝虫, the first one functions as s/p-component 贝 *bèi* 'shell'

The attribute connecting the A group characters is "indivisibility". This is principally ensured through the one-element graphic structure. Although most of these characters originated as pictograms or symbols, understanding the current connection between their graphic form and the meaning of the recorded morpheme usually requires a more or less extensive etymological explanation. Only a small number of characters can be found about whose graphics it can be said that they distinctively reflect the recorded meaning (labeled as Category A2). These characters are characterized by constructional indivisibility which is superior to the structural decomposition possibilities. In addition, one more specific category can be identified: graphemes where another independently existing character with exactly one more or less distinctive stroke can be recognized (labeled as Category A3). In traditional classification, characters composed on this principle would be considered symbols, however, compared with these it is important to highlight the one-stroke difference between the initial and derived character. 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 as such provide significant information in relation to the meaning or pronunciation of the derived character.

*Category A1* (215): one unmotivated element

马 *mǎ* 'horse'

石 *shí* 'stone'

*Category A2* (30): pictographic or symbolic reflection of the meaning

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

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

*Category A3* (12): existing character ± one distinguishing stroke

本 *běn* 'root': horizontal stroke added to the lower part of the character 木 *mù* 'tree' symbolizes the meaning 'root'.

灭 *miè* 'extinguish': a horizontal stroke added on top symbolizes the object that covered 火 *huǒ* 'fire'

## 4.   Conclusion

The new model of classification used in this paper was primarily designed as an attempt to find a systematic solution for the evaluation of unmotivated units occurring in the graphics of modern Chinese characters. This was achieved through the adoption of a two-dimensional arrangement that enables an evaluation of the decomposition possibilities of the characters and the relationship between the graphic and linguistic structure separately. The limitations of Su Peicheng's new six categories were resolved, however, through a better specification of the principles applied by an evaluation of the semantic and phonetic motivation of the graphics of the characters from a synchronic perspective. In addition, the new model argues that only the implementation of both the constructional and structural decomposition methods can establish a good set of criteria for an effective classification system. It is important, however, to notice that these approaches do not blend together, but either one or the other is applied at a certain stage of decomposition process.

Thanks to the precisely defined parameters, two types of unmotivated constituents were identified, considering the potential of being used as a semantically or phonetically motivated graphic unit in other characters of modern Chinese signary. There are consequently different types of characters with entirely, and with partly unmotivated, graphics which can be distinguished. As can be observed below, the first mentioned can be divided into three types and the second one into two types.

Types of unmotivated characters:

(a)  indivisible represented by category A1;
(b)  divisible into elements represented by category B1;
(c)  divisible into two n-components represented by category C1.

Types of partly unmotivated characters:

(a)  divisible into elements represented by categories B2, B3 and B4—an s-component can be identified in the composition of the B2 characters, a p-component in the composition of the B3 characters and an s/p-component in the composition of the B4;
(b)  divisible into two-components represented by categories C2, C3 and C4—together with one n-component, the C2 characters are composed of one s-component, C3 characters of one p-component, and C4 characters of one s/p-component.

As mentioned above, the model was proposed before the third revised edition of *Modern Chinese Grammatology* was published. It should be admitted that, compared to Su Peicheng's classification, the proposed model has a rather complicated arrangement. It is, however, a two-dimensional layout that provides deeper insight into the nature of the graphemes composition and thus provides a better understanding of the characteristic features of the modern Chinese writing system. In closing, it is important to note that the proposed model, in its current form, represents a prototype that aims to provide the foundation for examining a larger data sample. Due to the fact that it is based on an analysis of only one part of modern Chinese signary, it has been developed with the intention of outlining the most complex spectrum of possibilities. Although the general pattern appears to sufficiently reflect the characteristics of the currently used writing system, certain modifications can be expected, especially considering the fact that groups D and E contain only a limited number of characters.

## References

Cao, X. [曹先擢] and P. Su [苏培成], eds. (1999). 汉字形义分析字典 *[Analytic Dictionary of Chinese Character Graphics and Meanings]*. Beijing: 北京大学出版社 [Beijing Daxue Chubanshe].

Chen, M. [陈梦家] (2006). 中国文字学 *[Chinese Grammatology]*. Beijing: 中华书局 [Zhonghua Shuju].

Chinese Academy of Social Sciences, Institute of Linguistics, Dictionary Department [中国社会科学院语言研究所词典编辑室] (2002). 汉英双语现代汉语词典 *[The Contemporary Chinese Dictionary: Chinese-English Edition]*. Beijing: 外语教学与研究出版社 [Waiyu Jiaoxue yu Yanjiu Chubanshe].

Defrancis, J. (1984). *The Chinese Language. Fact and Fantasy*. Honolulu, HI: University of Hawaii Press.

Dong, X. [董希谦] (1994). 《说文解字》一夕谈 *[One Evening Talking about "Shuo Wen Jie Zi"]*. Zhengzhou: 河南人民出版社 [Henan Renmin Chubanshe].

Gao, G. [高更生] (2002). 现行汉字规范问题 *[The Issue of Modern Chinese Characters Standardization]*. Beijing: 商务印书馆 [Shangwu Yinshuguan].

Gao, J. [高家莺], K. Fan [范可育], and J. Fei [费锦昌] (1993). 现代汉字学 *[Modern Grammatology]*. Beijing: 高等教育出版社 [Gaodeng Jiaoyu Chubanshe].

Guder-Manitius, A. (1999). *Sinographemdidaktik. Aspekte einer systematischen Vermittlung der chinesischen Schrift im Unterricht Chinesisch als Fremdsprache. Mit einer Komponentenanalyse der häufigsten Schriftzeichen*. Heidelberg: Julius Gross Verlag.

Hao, E. [郝恩美] (1994). "现代汉字教学法探讨 [Investigation of Teaching Strategies for Chinese Characters]". In: 语言文字应用 *[Applied Linguistics]* 1994, pp. 83–87.

Haralambous, Y. (2013). "New Perspectives in Sinographic Language Processing through the Use of Character Structure". In: *Computational Linguistics and Intelligent Text Processing, CICLing 2013*. Ed. by A. Gelbukh. Vol. 7816. Berlin, Heidelberg: Springer, pp. 201–217.

Huang, W. [黄伟嘉] and Q. Ao [敖群] (2009). 汉字部首例解 *[Chinese Radicals in Examples]*. Beijing: 商务印书馆 [Shangwu Yinshuguan].

Li, Y. [李燕] and J. Kang [康加深] (2002). "现代汉语形声字声符研究 [A Study of Phonograms in Modern Chinese]". In: 现代汉字学参考资料 *[A Collection of Papers on Modern Chinese Grammatology]*. Ed. by P. Su [苏培成]. Beijing: 北京大学出版社 [Beijing Daxue Chubanshe], pp. 141–154.

Ma, X. [马显彬] (2013). 现代汉字学 *[Modern Grammatology]*. Guangzhou: 济南大学出版社 [Jinan Daxue Chubanshe].

Pan, J. [潘钧] (2003). 现代汉字问题研究 *[Analysis of Selected Issues of Modern Chinese Characters]*. Kunming: 云南大学出版社 [Yunnan Daxue Chubanshe].

Qian, N. [钱乃荣] (2001). 现代汉语 *[Modern Chinese]*. Nanjing: 江苏教育出版社 [Jiangsu Jiaoyu Chubanshe].

Qiu, X. [裘锡圭] (1988). 文字学概要 *[Outline of Grammatology]*. Beijing: 商务印书馆 [Shangwu Yinshuguan].

Schindelin, Cornelia (in this volume). "The Li-Variation (隶变/隸變) *lìbiàn*. When the Ancient Chinese Writing Changed to Modern Chinese Script".

———— (2007). *Zur Phonetizität chinesischer Schriftzeichen in der Didaktik des Chinesischen als Fremdsprache. Eine synchrone Phonetizitätsanalyse von 6.535 gebräuchlichen Schriftzeichen*. Munich: iudicium Verlag.

Shi, Z. [施正宇] (1992). "现代形声字形符表义功能分析 [An Analysis of Determinatives Semantic Function in Modern Chinese Phonograms]". In: 语言文字应用 *[Applied Linguistics]* 1992, pp. 76–92.

Slaměníková, Tereza (2013). *Ideogramy v moderní čínštině [Ideograms in Modern Chinese]*. Olomouc: Univerzita Palackého.

———— (2017a). *Čínské znakové písmo: Synchronní model tradiční kategorizace [Chinese Writing System: A Synchronic Model of the Traditional Categorization]*. Olomouc: Univerzita Palackého.

———— (2017b). "Proposal for a New Classification System for Modern Chinese Characters". In: *Studia Orientalia Slovaca* 16, pp. 87–106.

Su, P. [苏培成] (1994). 现代汉字学纲要 *[Outline of Modern Chinese Grammatology]*. Beijing: 北京大学出版社 [Beijing Daxue Chubanshe].

———— (2001a). 二十世纪的现代汉字研究 *[20th Century Research in Modern Chinese Characters]*. Taiyuan: 书海出版社 [Shuhai Chubanshe].

———— (2001b). 现代汉字学纲要 *[Outline of Modern Chinese Grammatology]*. Beijing: 北京大学出版社 [Beijing Daxue Chubanshe].

———— (2007). "现代汉字学的学科建设 [Founding of the Discipline of Modern Chinese Grammatology]". In: 语言文字应用 *[Applied Linguistics]* 2007, pp. 2–11.

Su, P. [苏培成] (2015). 现代汉字学纲要 *[Outline of Modern Chinese Grammatology]*. Beijing: 北京大学出版社 [Beijing Daxue Chubanshe].

Tang, L. [唐兰] (2001). 中国文字学 *[Chinese Grammatology]*. Shanghai: 上海教育出版社 [Shanghai Guji Chubanshe].

Uher, D. (2013). *Hanská grammatologie [Han Grammatology]*. Olomouc: Univerzita Palackého.

Wang, F. [王凤阳] (1989). 汉字学 *[Chinese Grammatology]*. Changchun: 吉林文史出版社 [Jilin Wenshi Chubanshe].

Wang, N. [王宁] (2001). 汉字学概要 *[Outline of Chinese Grammatology]*. Beijing: 北京师范大学出版社 [Beijing Shifan Daxue Chubanshe].

Wen, W. [文武] (1987). "关于汉字评价的几个基本问题 [Several Basic Issues on the Evaluation of Chinese Characters]". In: 语文建设 *[Language Planning]* 1987, pp. 7–12.

Xu, S. [许慎] (1963). 说文解字 *[The Meaning Explanation of Primary Characters and Structure Analysis of Secondary Characters]*. Beijing: 中华书局 [Zhonghua Shuju].

Yang, H. [杨洪清] and X. Zhu [朱新兰] (1996). 快速识字字典 *[Chinese Characters: Quick and Easy]*. Nanjing: 江苏古籍出版社 [Jiangsu Guji Chubanshe].

Yang, R. [杨润陆] (2008). 现代汉字学 *[Modern Grammatology]*. Beijing: 北京师范大学出版社 [Beijing Shifan Daxue Chubanshe].

Ye, Ch. [叶昌元] (2008). 字理 —— 汉字部件通解 *[Chinese Characters Motivation: General Explanation of Characters Components]*. Beijing: 东方出版社 [Dongfang Chubanshe].

Yu, G. [余国庆] (1995). 说文学导论 *[Introduction to The Meaning Explanation of Primary Characters and Structure Analysis of Secondary Characters]*. Hefei: 安徽教育出版社 [Anhui Jiaoyu Chubanshe].

Yu, H. [余辉] and Liu Ch. [刘诚] (2010). 说字理解汉字 *[Explanation of Chinese Characters by Principles of Motivation]*. Beijing: 气象出版社 [Qixiang Chubanshe].

Zhang, J. [张静贤] (1992). 现代汉字教程 *[Tutorial of Modern Chinese Characters]*. Beijing: 现代出版社 [Xiandai Chubanshe].

Zhou, Y. [周有光] (1980). 汉字声旁读音便查 *[Guide to Phonetics Pronunciation in Chinese Characters]*. Changchun: 吉林人民出版社 [Jilin Renmin Chubanshe].

———— (2004). "现代汉字学发凡 [Introduction to Modern Chinese Grammatology]". In: 周有光语言学论文集 *[Collection of Zhou Youguang's Linguistic Articles]*. Beijing: 商务印书馆 [Shangwu Yinshuguan].

新华字典 *[Xinhua Dictionary]* (2011). 11th ed. Beijing: 商务印书馆 [Shangwu Yinshuguan].