

Research on Chinese word-formation

Lang Feng (✉ p25828@gmail.com)

Nation Pingtang University

Article

Keywords: Chinese characters, imagery, ancient culture, combination, cognition

Posted Date: October 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-83895/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Research on Chinese word-formation

An imagery word-formation theory based on the concept of empirical generation

Lang Feng¹

abstract

Pinyin text uses letters that are closely related to speech to record speech, but we don't know much about how ideographic text is represented. We analyzed the relationship between the glyph and the meaning of nine thousand Chinese characters and found that the combination of the glyph is actually to construct imagery, and this imagery is the psychological representation of the experience situation that forms the concept of its character, so the meaning and glyph of the characters are ideographically connected. We find out the imagery and system structure of these 9,000 Chinese characters, and summarize the five cognitive modes of their combination. These results reveal the ideographic mechanism of the dual-code hierarchical combination of the internal formation of Chinese characters and its imagery, demonstrate the application of cognitive linguistics and cognitive psychology in the study of ideographic characters, and put forward a guiding theory of Chinese word-formation.

Keywords: Chinese characters, imagery, ancient culture, combination, cognition.

¹ Nation Pingtung University. E-mail: p25828@gmail.com.

1 Introduction

According to the literature²³, the current writing systems in the world can be roughly divided into the alphabetic writing system(pinyin text) and ideographic writing system. The coding of the Pinyin writing system uses the correspondence between letters and pronunciation, that is, the "graphemes-phoneme correspondence rule" to combine letters to record speech, so that there is an obvious master-slave relationship between speech and text. However, there is no such direct correspondence between Chinese phonetics and glyphs, because Chinese characters do not directly express phonetics⁴, and glyphs are not coded according to phonetics, and there is no master-slave relationship between spoken language⁵ and Chinese characters like pinyin characters. So, can we say that Chinese characters are ideographic? So far, it cannot be said for certain lightly. Since Xu Shen(許慎)⁶, the Chinese character scholars have generally accepted and used the concept of pictophonetic characters, and pictophonetic characters accounted for more than 80% of Chinese characters. How can we say that the form of Chinese characters is ideographic under the current situation where sound symbols are in the majority? If so, how do they express it? Therefore, in the past two thousand years, the research status of Chinese characters has always been that phonograms and ideograms are unclear, and the glyph of Chinese characters is like water without a source and wood without a root.

Therefore, based on the firm belief that the glyph of Chinese characters is by no means arbitrary coding, we must re-examine Chinese characters scientifically. First of all, we limit the research object to regular script Chinese characters, that is, to research the synchronic plane, because the diachronic change of the glyphs involves the designer's cognition of their environment, otherwise the structure of the Chinese characters will not change⁷. We cannot use inter-generational cognition to explain before and after. The phenomenon can only be explained by the phenomenon itself. This is a scientific requirement. Therefore, the Chinese characters referred to below refer to regular script glyphs.

Secondly, we decided to change our view on the phonetic symbols of pictophonetic characters, that is, we did not think they were phonetic, because according to the statistics of scholars⁸, the motivation of phonetic symbols of pictophonetic characters to copy sounds is not strong. Therefore, although the phonological characters are similar or the same with their phonetic symbols, which may be because the phonetic symbols themselves are Chinese characters and thus lead to the phenomenon of rhyme in poetry, this is only an epiphenomenon, which is not borrowed to acquire the sound of phonetic symbols. We should not put the cart before the horse. Later, we found that the so-called phonetic symbols and their radicals are the same, they participate in the construction of

² Sampson, Geoffrey, 《Writing Systems: A Linguistic Introduction》 (Stanford: Stanford University Press, 1985).

³ Coulmas, Florian, 《The Writing Systems of the Word》 (Oxford: Blackwell Publisher, 1991).

⁴ Xie Guoping(), 《An introduction to linguistics 》 (Taipei: sanmin 2016), 16.

⁵ From the geographical area where Chinese characters are used, there are at least seven dialects in the domain.

⁶ Xu Shen(58-147) the compiler of the original Han dynasty dictionary 《ShuWen JieZi》 .

⁷ Chinese characters from oracle, bronze, large and small seals, to the current Di script and regular script characters.

⁸ Zhou Youguang (周有光), The pronunciation the sound of Chinese characters can be checked (漢字聲旁讀音便查) (Jilin: Jilin people's Press 1980). His statistical result is: 39% of the effective phonetic transcription rate of modern Chinese characters.

the imagery of Chinese characters, they are also part of the ideographic mechanism. In the same way, we have also changed our view on pure symbols of Chinese characters that are neither ideographic nor phonetic⁹, because later we also found the imagery of these pure symbols. In this way, the internal character of Chinese characters is single, that is, the internal characters of Chinese characters are all ideograms, and the nature of Chinese characters is completely ideographic.

Thirdly, since the invention of Chinese characters was after Chinese, our research focused on how Chinese characters are shaped according to their meanings, exploring why Chinese characters are so configured to characterize their meanings, and seeking the relationship between meanings and glyphs. The problem of Chinese character formation is the problem of word-formation. Only by clarifying the internal relationship between the glyph and meaning of a character can the problem of written word formation in the Chinese be solved. This is a major problem to be solved in linguistics about non-phonetic language.

Looking back on history, since the publication of 《Shuo Wen Jie Zi》 by Xu Shen of the Eastern Han Dynasty, a huge monument of Chinese characterology has been erected, and no one has been able to surpass this monument. However, this monument has been standing for nearly 2000 years, and his theory of six-books cannot meet the requirements of today's scientific theory. In the following Song Dynasty, although there were some bright spots in Wang AnShi's (1021-1086) " Zi Shuo " (王安石：字書) and his contemporaries Wang ShengMei's "You Wen Shuo"¹⁰(王聖美：佑文說), it is a pity that they only put forward ideas and did not carry out systematic sorting and empirical exploration of ideographic characters.

In the context of today's humanities, linguistics, cognitive linguistics, cognitive psychology, semiotics, linguistic semiotics, cultural semiotics, etc., have provided theoretical preparation and basis for us to study the relationship between the form and meaning of Chinese characters. Today, we see Chinese characters not only as a language phenomenon but also as a phenomenon of human social behavior, overall culture, and collective psychological cognition. Therefore, we study Chinese characters from the three levels of language, cognition, and culture and the relationship among them. Because as a linguistic level, words are only a kind of signifier symbol of material intermediary, and behind it is the concept produced by the mind's understanding of the world, and these cognitive contents are the reflection of the experience of cultural knowledge and natural common sense.

2 Results

We selected nine thousand Chinese characters, disassembled and analyzed them, and found that the form of Chinese characters is a hierarchical combination structure, behind

⁹ Chinese character study textbooks generally believe that there are three different types of characters in Chinese characters, namely, ideographic symbols, phonetic symbols, and pure symbols(that are neither ideographic nor phonetic). Qiu Xigui (裘錫圭), 《Summary of Chinese Characterology》(文字學概論), (Taipei: Wanjuanlou 萬卷樓 1995), 2. 15-17.

¹⁰ Huang Dehuan(黃德寬), Chen Bingxin(陳秉新), 《History of Chinese Philology》, (Taipei: linkingbooks 2008), 117-121.

which is the process of hierarchically constructing imagery, and this imagery is the psychological representation of the experience situation that forms the concept of its character. In other words, Chinese characters achieve the ideographic effect through certain empirical imagery related to the concept of their character. Therefore, exploring the dual-code structure of the hierarchical combination of Chinese characters and the cognitive model of their mutual combination is our greatest achievement.

2.1 Concepts and structure

According to Saussure's semiotics about the abstract relationship within words, that is, the opposition and coexistence relationship between signifier and signified, we have established three hierarchical structures of the upper, middle, and lower levels of Chinese characters. From the perspective of the glyph of the signifier, it is divided into three hierarchical levels: the upper, the middle, and the lower of the "glyph, component, and icon". From the perspective of the imagery of the signified, it is the upper, middle and lower of "imagery-glyph, imagery-component, and imagery-icon". The glyph corresponds to the imagery-glyph, the glyph is the signifier, and the imagery-glyph is the signified, which is the upper level or the level of the character, the glyph carries the imagery-glyph, the glyph is the carrier, and the imagery-glyph is the body. The imagery-glyph is the imagery of the formation of Chinese character, which is the core and purpose of the formation of the Chinese character. The glyph is the written form of a Chinese character. The icon corresponds to the imagery-icon, the icon is the signifier, the imagery-icon is the signified, which is the lower level or element level. The icon carries the imagery-icon, the icon is the carrier, and the imagery-icon is the body. Icons are some iconic symbols or pictographic characters. They are the smallest unit of loading imagery, that is, if the icons are divided into subdivisions, no imagery can be loaded. Similarly, the imagery-icons are some directly emerging concepts, which are some imagery that directly describe the shape, such as "人, 木, 日, 月" and so on. These few icons and their imagery-icons are combined hierarchically to construct the entire Chinese word system. The component corresponds to the imagery-component, the component is the signifier, the imagery-component is the signified, which is the middle level or the structure level, the component carries the imagery-component, the component is the carrier, and the imagery-component is the body, the component is the intermediate part except for the glyph and the icon. The imagery-component is the imagery of the component. A component is a combination of icons or other components. The component can be a combination of multiple levels until it is formed into a glyph. Similarly, an imagery-component is a combination of imagery-icons or other imagery-components. The imagery-component can be a combination of multiple levels until it is formed into an imagery-glyph. From icon to a glyph, there may be many combinations in the middle, and each combination is called a level or plane. And in each level or a combined plane, we call it a byte. At the level of imagery-component, its corresponding combined imagery-component is called imagery-byte. The dual-code structure relationship between the glyphs and the imagery is just like the relationship between the signifier and the signified in the semiotic theory. The signifier forms serve signified imagery. The imagery-glyph is the result of the continuous combination and construction of the signified imagery from the imagery-icon to the layered imagery-component. For example, the structure of the character "構" is shown in Figure 1.

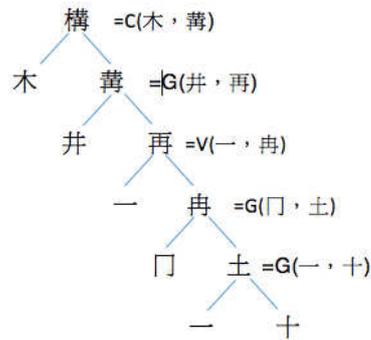


Figure1 the structure of the character “構”

The character "構" uses only five icons (十, 一, 冂, 井, 木) and its imagery-icon. After five layers of combination, it is five bytes (土, 冉, 再, 堯, 構), the character "構" is formed. Here, in addition to "構" and icons, the rest are components. The imagery-glyph of "構" is “Such as the wooden pole truss above the wellhead.” It uses the contextual imagery of digging a well to metaphorically represent the meaning of the character "構", and it also includes its multiple meanings. Of course, this imagery is very empirical, if you are familiar with the traditional folk digging craft. Each byte corresponds to its imagery-bytes, and the imagery-glyph is constructed by stacking layers of imagery-bytes according to a certain combination mode.

From the structure of Chinese characters, it can be seen that the entire Chinese character system is constructed by a small number of icons and their imagery-icons. Not only that, this way of constructing Chinese characters in modern times continues to be extended, and a large number of multi-character words combining Chinese characters and Chinese characters have emerged, that is, the establishment of the entire word system including multi-character words. Here, we can also see that the glyph, component, and icon of the character are dominant, that is, it can be observed. while the imagery-glyph, imagery-component, and imagery-icon of the character are recessive and exist in our brains. One of its tasks is to find them scientifically.

2.2 Combination mode

From the perspective of the form of characters, Chinese characters are composed of combinations, and their purpose is to construct the imagery-glyphs. This combination starts from the icon. Our research has found that the ancient Han people used different imagery to combine and construct new imagery, which is represented by the combination of icons or components to form new components or glyphs. For example, "土" is a combination of "一" and "十". Both "一" and "十" are one of the few icons. Why does the concept of "土" combine "一" and "十"? It turns out that the imagery of "一" is "the chaos of heaven and earth is like an egg and the PanGu(盤古) is born in it." This is the initial state that heaven and earth are inseparable. Here, "一" is below, and generally represents the earth. The imagery of "十" is "after PanGu's death, his body transformed into all things and formed a complete earth", so this two imagery jointly construct the imagery of "土": "a large number of PanGu's body transformed matter on the surface." Here, we are most concerned about the relationship between "一" and "十", and what is their combined rationale model. After a lot of research on the imagery-glyphs, we have summarized five combination modes, namely conceptual mode, metaphor mode, gestalt mode, verb group mode, and declension mode. The first three models belong to cognitive models, which are based on cognitive principles, while the latter two belong to syntactic principles and the principles of glyph transformation. These five combination modes are called the word-formation rules of Chinese characters, and the mechanism that constructs the imagery of Chinese characters through the

word-formation rules to represent the meaning of the characters is called the ideographic mechanism of Chinese characters. The theory of this kind of ideographic of Chinese characters is called the "imagery word-formation" hypothesis. We describe these five combination modes separately.

2.21 Conceptual mode

There are many theories about classification and concepts, and in modern times they have mainly developed into prototype theories. The key to the prototype is the characteristic attribute, which describes the characteristics of the prototype. When we explored the imagery-glyphs and the imagery-components, we found that the ancients had already mastered the essence of classification and concepts. That is to say, when two imagery are combined to construct new imagery, there are a large number of combination modes that are conceptual relationships such as [category + feature]. Therefore, we call this model a conceptual model. For example, the characters "桃, 柚, 梅, 棠, 梨, 楓" are generally called pictophonetic characters in Chinese character study textbooks. Among them, "兆, 由, 每, 尚, 利, 風" are called phonetic symbols, they are borrowed to distinguish the names of these fruit trees¹¹. However, we believe that this understanding is too far-fetched, because if we want to distinguish or characterize phonetics, why should we choose these characters instead of others? Because there are many homophones in Chinese characters. More importantly, this kind of cognition cannot represent the meaning of characters, although it only represents trees. Let's look at the character "桃". In our research, the imagery of "兆" is "burning oracle bones for divination, cracks like splashes of water." It is derived from divination and has the sign of "symptom". therefore, the imagery of "桃" is "Fruit trees that bring signs of spring", in which "木" is the category, and "兆" is the characteristic. This interpretation can be confirmed by a large number of literary works, such as the use of peach blossoms in our culture to symbolize the love between men and women, because peach blossoms herald the arrival of spring. Another example is the character "楓". "風" here is not a phonetic sign, but a characteristic of maple, because the color change of maple leaves cannot be separated from a large amount of autumn breeze. this common sense is in line with the modern principles of the botany of the maple family. As for the sound of "风" because it adopts the characteristics of the character "风", we should not put the cart before the horse.

If the propositional representation is used, the conceptual model should be:

Chinese character = C (category, feature)

Here C is a conceptual model, which means that there is a conceptual relationship between categories and features. Therefore, the propositional representation of "桃" can be written as:

桃 = C (木, 兆)

2.22 Metaphor mode

Early in ancient China, it was known to use metaphors in literary works, such as "The Book of Songs". But since George Lakoff & Mark Johnson put forward the "Conceptual Metaphor Theory", we suddenly raised our understanding of metaphor basis on cognition, so that we understand that metaphor is everywhere. The conceptual metaphor theory assumes that metaphors are a cognitive phenomenon, they appear in language, and they have a cognitive basis. The conceptual metaphor connects two conceptual domains: source domain and target domain. A conceptual domain is a collection of semantically related essence, characteristics, and

¹¹ Wang Ning's(王寧) "Lectures on Chinese Character Configuration" (Taipei: sanmin 2013), 41.

functions. The source domain usually consists of concrete concepts, such as cash, while the target domain involves abstract concepts, such as time. The conceptual metaphor theory assumes that we use the source domain to understand the target domain. For example, when we say "time is money", we use money as the source domain to explain time. The metaphorical relationship between the source domain and the target domain exists in the combination of two imagery in the Chinese characters to construct new imagery. We call this metaphorical model of [target domain + source domain] when composing imagery-glyphs or imagery-components. For example, the two characters "坡, 咚", in which the characters "皮, 冬" is used as the source domain to metaphor the two target domains of "土 and 口", that is, the ground (土) is inclined like skin and the sound (口) is like Winter hits the ice. Of course, the metaphor here is also used as a feature, combined with "土" and "口" in a conceptual model. This combination of multiple modes is called a compound mode.

If the propositional representation is used, this metaphorical model can be expressed as:

Chinese characters = M (target domain, source domain)

Here M is a metaphorical model, which means that there is a metaphorical relationship between the target domain and the source domain. Therefore, the propositional representation of the character "坡" can be written as:

坡 = CM (土, 皮)

The imagery-glyph is the imagery of Chinese character, and its essence is as a metaphor for the meaning of the Chinese character. In other words, familiar imagery is used to metaphor the abstract and universal meaning of the characters. For example, the character "權", the literal imagery-glyph (imagery) describes the scene of a bird's nest on a tree in the wild. Approaching the bird's nest will be attacked by biological instincts. Therefore, this metaphor can represent concepts such as power and measurement. Similarly, the imagery of the character "歡" is that the mother bird returns to the nest and the young birds each open their mouths to greet each other, which is a metaphor for the mood at the moment. A bird's nest imagery of "窠" can be combined into the imagery of characters such as "灌, 權, 歡, 罐, 觀, 驩, 勸, 顴, 鵲, 矚, 謹, 鑰, 權, 權, 燿", etc. Its magic and efficiency are evident.

2.23 Gestalt mode

Gestalt theory is also called the Gestalt theory. It developed from Germany as early as the 20th century. The famous argument of Gestalt psychology is that the whole is greater than the sum of its parts. That is, the best way to understand a psychological phenomenon is to see it as an organized and structured whole. According to this view, much linguistic information is incomplete, and they need to be made up of human gestalt ability (Kurt Koffka 1935). When humans observe and perceive the objective world, they always involuntarily allow their subjective experience to participate in the understanding, linking the characteristics of seemingly unrelated things, to achieve an overall grasp of the objective. The various imagery (mind imagery) that we store in our brains cannot arise out of thin air. Imagination without imagery is like wood without roots, water without sources. And we put forward the theory of "imagery word-formation", also using the Gestalt principle, from the three levels of language, cognition, and culture. In other words, we start from the human cognitive instinct and cognitive rules, regard the concept formation of the meaning of the character and the structural imagery representation of Chinese characters as the result of the interaction between humans and the external world, and add the cultural background to integrate them into a whole. Gestalt theory is one of the rules of Chinese word-formation.

If the integrity of this kind of gestalt is limited to the composition of the imagery-glyph, a considerable number of Chinese characters are combined using the principle of gestalt cognition. That is, two or more imagery-components together represent the overall imagery of something. Its basic model is [feature + feature], therefore, we call this model the Gestalt model.

For example, the character "火" is a combination of "人" and "灬". Its imagery-glyph is "light and flames that make people sweat profusely." Here, "人" (people) and "灬" (sweat) are both features and features points to the unstructured shape alternative things, namely the concept of fire. Another example is "囿", which is a non-character component, and its imagery-component is a square upper and lower symmetrical container (囿) containing millet and other ancient millets. It is composed of "一, 口, 田", and they are all characteristics of "囿": "一" signifies a symmetrical lid, "口" signifies the opening of the container, and "田" signifies the food in the container. Another example is the character "爲", which is a combination of "爪, 尸, 尸, 勹, 灬", "爪" represents the claw, "尸" represents the body, "勹" represents the handle of a hand-held object, and "灬" represents the dynamics. They are combined into character, which together represents the imagery of the two monkeys catching lice and each other. "爪, 尸, 尸, 勹, 灬" are the constructive features.

If you use the propositional representation, this Gestalt model can be represented as:

$$\text{Chinese character} = G(\text{feature}, \text{feature})$$

Here G is Gestalt mode, which means that there is a Gestalt relationship between features. Therefore, the propositional representations of the two characters "火" and "爲" can be written as:

$$\begin{aligned} \text{火} &= G(\text{人}, \text{灬}) \\ \text{爲} &= G(\text{爪}, \text{尸}, \text{尸}, \text{勹}, \text{灬}) \end{aligned}$$

The difference between the conceptual model and the Gestalt model is that the subject of the conceptual model is the same as the category component, while the subject of the Gestalt model is constructed by constructive features, and they have no category relationship with the subject. For example, the concept of "身" has no categorical relationship with its characteristic components "自" and "才". while "楓" is similar to its component "木".

2.24 Verb group mode

Verb group (verb phrase, VP) is a concept in linguistic syntax. The reason why we are interested in verb groups is that imagery-components combinations of many Chinese characters are similar to the verb group structure, which means that the group of phrases surrounding the verb is combined, including verb-object phrases, verb-complement phrases, or verb-auxiliary phrases. A verb-object phrase refers to the combination of a verb and an object, and a verb complement or a verb auxiliary phrase refers to a combination of a verb and a complement or auxiliary word. For example, "打" is a combination of "扌" and "丁". The imagery-glyph of "打" is "Hold the tool to hit the dowel." Because "扌" and "丁" are the relationship between verb and object phrase, we call the imagery-glyph or imagery-component that has the verb group relationship as the verb group mode. Similarly, the character "有" is a combination of "ナ" and "月". Its imagery-glyph is to hold beast meat, which is also a relationship between verb and object. Another example is "刳", which is composed of "牛" and "刃". The imagery-glyph means that a knife can move easily between the bones and flesh of the beef. "牛" and "刃" have a dynamic complement relationship.

If you use the propositional representation, this verb pattern can be represented as:

Chinese characters = V (verb, object)

The V here is the Verb pattern, which means that the verb and the object are the verb-object relationship. Therefore, the propositional representation of "打" can be written as:

打=V(扌, 丁)

2.25 Declension mode

The phenomenon of declension of Chinese characters refers to the non-combination of the form of Chinese characters, that is, the form of Chinese characters directly changes to change the original imagery. The first four combination modes are composed of two or more components and their imagery-components to construct new imagery. These imagery-components are combined with the components from the imagery-icon layer by layer until the imagery-glyph is completed. The combination of Chinese characters refers to the combination of more than one configuration in certain plane space. However, the declension mode is to change the shape of a single glyph or component to achieve a change in the imagery-glyph or imagery-component. For example, the character "才" is formed by missing the right hand of "木", which signifies that the tree becomes usable wood after removing branches and finishing. Another example is the character "夂", which is formed by the character "夂" missing one vertical line on the left. because the "夂" means that the adult is right in the inside, while the gaped "夂" means incompleteness and gaps. They all use physical changes to achieve the purpose of changing the original imagery. Therefore, we call this phenomenon the inflection of the Chinese character configuration. The imagery-component or imagery-glyph generated in this way is called the declension mode.

If you use the propositional representation, this inflection mode can be represented as:

Chinese character = D (component, deformation)

Here D is the declension mode, which means between the component and the deformation is the declension relationship. Therefore, the propositional representation of "才" can be written as:

才 = D (木, -)

Also, because there are seven ways of declension, the declension symbol must be added before the declension component in the propositional representation. According to our empirical research, the number of declension Chinese characters is not many, less than a hundred, but they still play an interpretive role in the production of imagery-glyph.

The above five combination modes occur in each combination plane, including their compound modes. The configuration of Chinese characters first develops from the combination of icons, then develops to the combination of components, and finally ends in the glyphs of the character. The creation of each new character and its imagery is constructed by combining existing shapes and imagery according to the combination mode. In this way, new words continue to expand.

However, the internal space of the Chinese character box is limited, new characters are constantly increasing, and the internal structure of some Chinese characters is unbearably swollen, which is inconvenient for writing and memory. Therefore, people think of the further expansion of the combination of Chinese characters, that is, the further expansion of word-formation from the combination of characters and characters. In this way, the number of characters used and the number of characters created can be controlled within a certain range, and the formation of characters and words is in the same line, and the combination machine is

the same. The concept of Chinese characters and modern words are integrated and unified. That is, a Chinese character can be a single-character word or a morpheme of a multi-character word, or a word can be composed of one character or multiple characters. However, the combination of characters in a multi-character word is not a combination of imagery-glyphs, but a combination of the meaning of Chinese characters, but their combination relationship still follows the aforementioned four combination modes (without declension mode). For example, "車頭, 車頂, 車窗, 車門, 車椅, 車燈, 車尾, 車體, 車胎" are still conceptual models. "火箭" is a gestalt and metaphor model. "革命" is a verb mode. "意象" is the Gestalt model. In short, multi-character words can greatly reduce the amount of character creation and literacy, and it is still in line with the effect of using simplicity to control the complexity so that we can predict the meaning of the word from the meaning of the character. For continuous word(連綿詞), such as 葡萄、蜻蜓、蝌蚪、蚯蚓, etc., although a single character has no meaning of the character, they do have imagery-glyph because they have a configuration and must have an imagery-glyph or imagery-component. The gestalt integration of characters and words can be characterized by the term "字詞".

2.3 Verification

Today, it is gratifying that the nine thousand commonly used regular script Chinese characters we have selected have all been cracked to obtain the imagery-glyphs, including their components and their corresponding imagery-components, as well as the icons and its imagery-icons of the entire system. There are a total of 87 icons and their corresponding imagery-icons, there are 246 non-character components, and the component combination has a maximum of ten levels. The level distribution of nine thousand characters is shown in Table1. Each combination confirms its combination pattern, for the statistical distribution of their various combination modes, including multiple combinations, see Table2. The character icons of the whole system are combined into a whole word system like DNA ladders, which shows that this system is very dense and efficient. Moreover, Chinese characters at any level are the result of the combination of lower-level Chinese characters or components, that is, all Chinese characters are the result of the combination of icons or imagery-icons. This provides the convenience of class and module association for the learning Chinese characters and also provides a standard for the disassembly of Chinese characters components. The comparison between the regular script Chinese character system and other inter-generational Chinese character systems is shown in Table3.

Table1 Statistical table of hierarchical distribution of nine thousand Chinese characters.

Icon 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10
87	297	1244	2606	2589	1339	671	103	108	33

It can be seen from Table1 that most Chinese characters appear in the fourth and fifth levels, within the sixth floor accounted for 90%, that is, eight thousand Chinese characters.

Table2 For the statistical distribution of various combination of nine thousand Chinese characters, including multiple combinations.

C	D	G	V	M
5578	90	2247	1446	4040

Table¹²3 The comparison between the regular script system and other previous Chinese characters system.

text type	total taken	icon	average icon
Oracle released	1380	412	3.3
QinJian sampling	1773	361	4.9
《ShuoWen》	10422	414	25.05
Regular script	9032	87	103.82

It can be seen from Table3 that the regular script system is more systematic than the 《ShuoWen》 system.

3 Discuss

Let us imagine that at the beginning of CangJie's¹³ creation of Chinese characters, there were no characters but language. Language evolved naturally, while writing is the invention of human culture. In other words, there was already spoken conversation when writing was invented, and the meaning can be expressed through spoken language. As far as CangJie is concerned, if he wants to invent written characters, the first thing he faces is what he must base on to encode the configuration so that these characters can record the sound and meaning of the language. However, speech is an invisible sense of hearing, and it is a fleeting phenomenon in person. Therefore, speech is limited to a specific time and space. The meaning is the relationship between us and the world, and this relationship is the result of human cognition and understanding of the world. Since this world is open to human beings, we complete the construction of meaning through the perception process of recognition, organization, and understanding from the sensations received from environmental stimuli. Knowledge is the key to perception¹⁴. Perception is a psychological phenomenon of perceptual channels such as vision, hearing, touch, taste, smell, etc. However, cognitive psychologists generally value visual perception because visual perception is the most widely recognized and most widely studied perceptual channel¹⁵. However, we do not perceive the external world with the appearance of the eyes, but the brain tries its best to understand the stimuli that enter the eyes and fall on the retina¹⁶. Therefore, compared to hearing, most people are more likely to raise the imagery of things in their minds to represent meanings¹⁷, including the intuitive depiction of objects. According to experimental evidence, the brain processes logographic text (image) and Pinyin text (sound) in different places¹⁸.

In history, no matter the east or west, human beings used pictures to represent an interest in the early days of writing. The existing evidence strongly points out that the text is not meant to express spoken language at the beginning, it comes from an interest in pictures and graphical representations¹⁹. These pictures have communicative intent. This writing system is called logographic text. However, with the complexity and abstraction of the concept of things, it

¹² Table 3 Except for the data of regular script Chinese characters, the rest of the data comes from Wang Ning's(王寧) "Lectures on Chinese Character Configuration" (Taipei: sanmin 2013), 18.

¹³ Legend has it that CangJie invented Chinese characters. Here the name CangJie is borrowed to represent those ancient people who designed Chinese characters.

¹⁴ Robert J. Sternberg, Karin Sternberg, translated by Li Hongy(李宏鑑), 《Cognitive Psychology》 (Taipei : Cengage Learning 2017), 82.

¹⁵ Same as above, 82.

¹⁶ Same as above, 83.

¹⁷ Same as above, 296.

¹⁸ George A. Miller, Translation by Hong Lan(洪蘭) 《The Science of Word 》 (Taipei: Yuan-Liou Publishing, 2002), 79.

¹⁹ Same as above, 81.

becomes more difficult to express in this way of images. In order to solve this problem, the ancients had to invent symbols that represent names and sounds, so a sequence of pictorial representations of the language had to be relegated to the representation of sequential syllables. This writing system is called syllabic and alphabetic text(Pinyin). Gelb²⁰(I. J. Gelb) wrote: Once this system was invented, the principle of phonetics spread quickly, and everything, no matter how abstract, can now be expressed in a text system. However, why didn't Chinese characters turn to phonetic coding? In the early stage of Chinese characters, a few iconic signs were also established, that is, after pictographic characters, they continued to move stubbornly toward the path of ideographic. How did it do that? How does it overcome the difficulty of an image representation of complex abstract things?

To explore this problem, let us first review how the concept or meaning of humans is formed, which involves human cognition. We know that cognitive linguistics regards the formation of the human body's phenomenal experience on the human conceptual system and the form of reasoning as the core position²¹. Husserl²² (Edmund Gustav Albrecht Husserl) said that primitive movements are the ancestor of all cognition. Cognition is the result of highly complex interactions among the body, brain, and situation. Many psychological experiments have established that language understanding is closely related to physical behavior^{23,24}. The famous neuroscientist Nobel laureate Edelman said: The idea of thinking of meaning as an abstract symbol is one of the biggest fallacies in the history of science. This experience-based cognitive theory gives the concept of meaning a grounded foundation. At the same time, the evolution of the human brain makes it possible to do high-level simulations of various actions of the human body, as well as the relative position of the human body in space, society, and causality. One of the final results of this continuous evolution of cognitive stimulation is the formation of human consciousness, and language plays an important role in supporting this continuous evolution of simulation²⁵. Wilson (Wilson, M. 2002) once concluded: Cognition is produced by action, and memory can be developed to transform the actual interaction experience of the external world into a form that can be operated in the brain. These off-site situational cognitive activities are based on concepts related to the body's perception of movement, many pieces of evidences show that we often perform off-site and non-real-time simulations of external situations. For example, "mental imagery" is a psychological simulation of external events Operation. Embodied thinking maybe the essence of human thinking²⁶. From the perspective of cognitive psychology, cognition is generated by the human nervous system and various perceptual senses, and then enters the attention of consciousness, and encodes information and transfers it to memory, until it is represented and organized in the heart to form knowledge, It is the journey from experience to knowledge. Therefore, the formation of a concept or meaning originates from the cognitive process of senses, perception, and experience.

In the formation of meanings and concepts, there are two issues that arouse our attention: one is how to represent these meanings or concepts in our hearts? The other is how these meanings or concepts extend from direct experience to complex abstraction to the entire conceptual system? Regarding the first question, cognitive psychologists generally agree that there are three encoding

²⁰ I. J. Gelb , 《A Study of Writing》 (Chicago: University of Chicago Press, 1963).

²¹ Huang Xuanfan(黃宣範): "Embodied cognition"(棲於身的體現認知), 《Language and Cognition》 (Taipei : Nation Taiwan University Publishing Center 2009,08), 344.

²² Husserl, E. 《Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy》 (Boston: Kluwer Academic 1980).

²³ Glenberg, A. M. & Roberson, D. A. 《Symbol grounding and meaning》 (Journal of Memory Language 43, 2000), 379-401.

²⁴ Kaschak, M. P. & Glenberg, A. M. 《Constructing comprehension》 (Journal of Memory and Language 43, 2000), 508 - 529.

²⁵ Huang Xuanfan(黃宣範): "Embodied cognition"(棲於身的體現認知), 《Language and Cognition》 (Taipei : Nation Taiwan University Publishing Center 2009,08), 364.

²⁶ Same as above, 357.

forms of representing knowledge²⁷, that is, imagery, language(or other symbols), and propositions. Because in the CangJie era, it was the era when writing symbols were created. Therefore, the coding form of representing knowledge is mainly imagery and propositions, and the propositional form is the product of the very developed logic of future generations.

What needs to be emphasized is that regarding the external representation of imagery, such as images and photos in books, and the internal mental representation that we care about, that is, mental imagery, Shepard (1971) and other researchers put forward the hypothesis of functional equivalence²⁸. It is argued that the representational form of imagery and the perceptual experience triggered by visual stimuli are functionally equivalent, even if the two(images and imagery) are not truly equal. Imagery is a specialized concept of cognitive psychology, it has been widely used in the fields of literature and art, it is a phenomenon in consciousness, it is the external scene reappeared in the brain, it is different from objective existence, it is the imaginary experience after being observed subjectively, it is the unity of subjective and objective.

Regarding the second question, that is, how meaning or concept extends from direct experience to complex abstraction to the entire conceptual system, the "conceptual metaphor theory" of cognitive linguistics has put forward a good explanation. George Lakoff & Mark Johnson said in "Metaphors we live by": We find that the system of daily concepts is essentially metaphorical²⁹. They analyzed embodied metaphors, from directly emerging concepts to metaphorically emerging concepts. They emphasized that the main function of metaphor is to understand, to conceive something from other things. This metaphor of familiar things with more abstract and complex things is precisely the main way in which meanings or concepts continue to deepen, expand, and accumulate into a system with cognition, in a word, it is achieved from direct emergent concepts to metaphor emergent concepts.

At this point, we return to the question of how CangJie encodes characters. In front of CangJie, he wanted to encode a known meaning or concept. He wanted to ask: Where did the concept or meaning to be encoded come from? He wants to go back to the source of the formation of meaning or concept, that is, back to the source of experience established by the meaning or concept, that is, the scene of the empirical situation where an event occurred, and the imagery of the situation will appear in his mind. So he must code according to this imagery because this imagery can represent this concept or meaning. This is the main way he encodes the meaning or concept (knowledge). For example, if he wants to represent the concept of "tree", he can draw the simple shape of a tree (木). If he wants to represent the concept of "person", he draws the shape of a person walking sideways (人), and so on. These are the so-called direct emergence concepts. Therefore, to understand the configurational rationale of Chinese characters, the key is to find out the configurational imagery of the character. This configurational imagery is connected with the meaning or concept of the character above and at the same time connected with the context of experience below. It is essentially metaphorical or analogous. Therefore, the way of the ideographic meaning of Chinese characters is mainly to complete the coding configuration through the mental representation of environmental cognition, that is, the imagery.

However, simple things can be traced directly, and such things are very few after all. How to represent more abstract and complex concepts? For example, how to represent the concept of "more"? Although "more" does not mean specific things, it is still extracted from specific things. As a result, CangJie remembered common imagery of concrete things that could represent the

²⁷ Robert J. Sternberg, Karin Sternberg, translated by Li Hongyi(李宏鑑), 《Cognitive Psychology》(Taipei: Cengage Learning 2017), 298-302.

²⁸ Same as above, 309.

²⁹ George Lakoff & Mark Johnson, translated by Zhou Shizhen(周世箴), 《Metaphors we live by》(Taipei: linkingbooks 2006), 9.

concept of "many": everything in the world is dark at night, and only the sky is full of stars. Therefore, the imagery of the night sky full of stars can represent the concept of metaphor "more". As a result, the concept of "moon"(月) emerges directly. Its deformed "Evening" (夕) means that the "月" is a little less to represent the light moon, which is the twilight period, and the overlapping combination of two "evenings" (夕+夕) can represent aggravation. The light moon is late at night. Thus, the character "多" was born. It can be seen from this that the configuration of Chinese characters is divided into two stages: the first stage is that a small number of directly ideographic characters appear first, that is, iconic characters. They are the result of directly tracing the shape of specific objects, and they are direct emergence concept. However, after all, there are not many iconic signs that can be directly described, but these iconic signs can be used as the features of the imagery, and the combination of several such imagery features, especially the combination that complies with the cognitive principle, can represent more complex and abstract imagery. In other words, complexity can be represented by simple combinations, and complex conceptual imagery can be represented by simple imagery combinations based on cognitive principles. As a result, the combination of these few iconic signs produces imagery features to represent the imagery of Chinese characters, so that Chinese characters can continue to extend and expand. This is the second stage of Chinese character formation. The emergence of metaphorical concepts in the second stage, they all use iconic signs to generate imagery features, and under the guidance of the cognitive model, the relevant imagery features are combined to represent the imagery of Chinese characters. It is manifested in the shape of the character, that is, a combination of hierarchical components with a few icons, and finally combined into the glyphs of Chinese characters. As for the combination of character and character into multi-character words, we call it the third stage. Therefore, this combination greatly improves the efficiency of the use of symbols and firmly controls the use of Chinese characters. The basic principle of constructing a Chinese word system can refer to Figure 2.

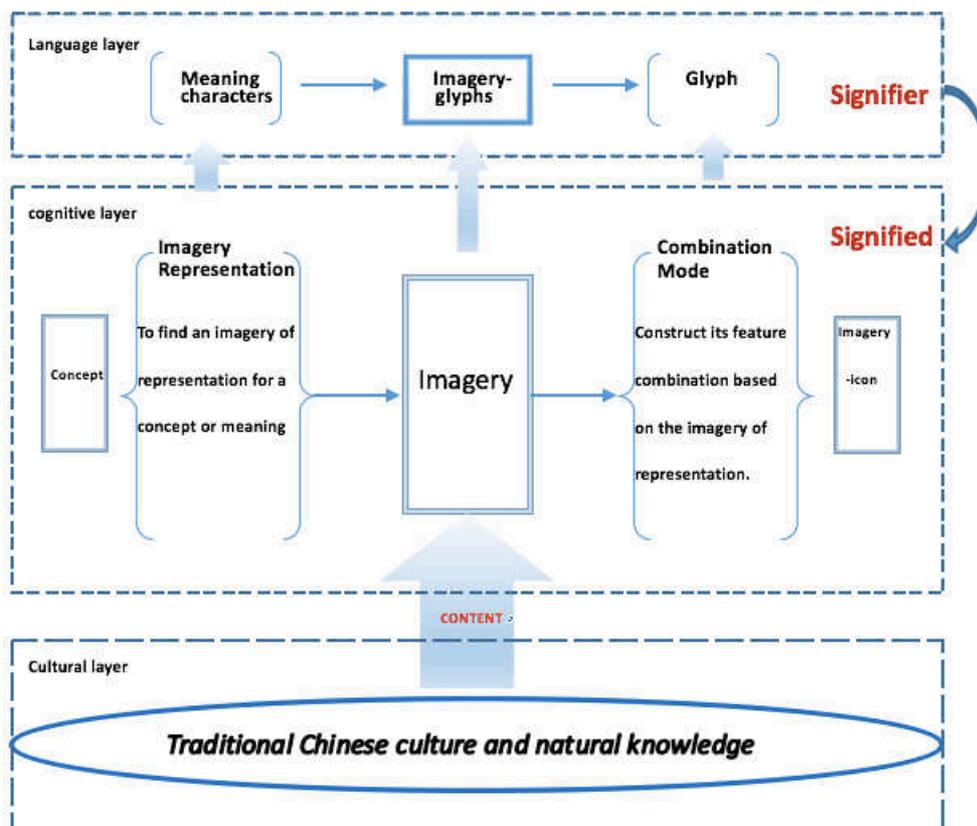


Figure2 the basic principle diagram of Chinese word formation

4 requirements and method

Like any theory, the hypothesis of "imagery word-formation" also needs to be combined with the phenomenon of Chinese characters, and to guide and explain the phenomenon of characters. Chinese characters and their meanings are a kind of literal phenomenon of human behavior. Finding out the basis for encoding configuration according to the meaning of a character is the verification of the theory of "imagery word-formation". However, except a few iconic characters and radicals, for most Chinese characters, we don't know what their imagery-glyphs, imagery-components and systematic imagery-icons are. Therefore, our empirical task is to use scientific methods to explore these imagery, to complete the verification of the theory of "imagery word-formation".

Therefore, we have selected nine thousand commonly used Chinese characters as sufficient systematicness and objectivity, which also allows us to have enough induction³⁰. We must find out the imagery-glyphs, imagery-components and imagery-icons of these nine thousand Chinese characters, including the components and icon that characterize them. This process is a great systematic project, because this system is the result of the combination of icons and their imagery-icons, and it is a multi-level combination, crisscrossing and moving the whole body.

The glyph is the carrier of the imagery, and the imagery determines the style of the glyph. However, when we explore the imagery-glyphs, the imagery-components, and the imagery-icons, we cannot guess these imagery. To meet the scientific requirements, we set four principles that cannot be violated by speculating imagery, namely, consistency, objectivity of configuration, interpretability, and Chinese culture and common sense.

4.1 consistency principle

The principle of consistency means that a certain imagery-icon or imagery-component or imagery-glyph must have the same imagery in the entire Chinese character system. Because the entire Chinese character system is the result of the repeated combination of icons including their combined components, this requires that each imagery-icon and imagery-component must be consistent in the imagery no matter how many times it is repeated, that is, the meaning of the imagery should be stable. In other words, an imagery-component, viewed from the horizontal, should maintain consistent imagery in other Chinese characters containing the component. Within a Chinese character, it should also maintain the same imagery as a component of the upper-level imagery that participates in the organization in the same plane. But it is the result of the combination of imagery-component or imagery-icon in the lower-level. In this way, an imagery-component maintains consistent imagery with the upper, lower, plane and other Chinese characters, and finally contributes its imagery to the imagery-glyph until it is connected with the meaning of the character. From the perspective of symbols, a pair of signifier and signified whether they are part of other symbols or in other Chinese characters, should maintain the consistency of signifier and signified. This is the case for an imagery-component, and it is the same for all layers, including every imagery-icon, to maintain the consistency of the imagery in the entire Chinese character system. Therefore, only by satisfying the Omni-directional imagery consistency of this system, can the whole system be closely connected and meet the scientific and systematic requirements. At the same time, this is also a requirement to avoid the interpretation of "vulgarization" of Imagery-glyph and to ensure the high efficiency of learning Chinese characters. For example, "良" is composed of the icon "日" and "廾". Their imagery-icon represents the sun and walking on both feet, and they combine to represent "such as

³⁰ Zheng Jinquan (鄭錦全 1998) checked various classics, history books, and dictionaries in the article "Understanding Language Cognition from Measurement" and found that different authors used no more than 8000 morphemes and non-derivative words.

Kuafu(夸父) chasing the sun to Yugu (boundary, limit, prehistoric) " This mythical imagery. In the westward run, Kuafu will meet a high mountain (Chinese Geography). when combined with other characters, such as "根, 即, 既, 艱, 良, 垠, 很, 恨, 狠, 限, 眼, 退, 萋, 痕, 眼, 硯, 裨, 跟, 銀, 簋, 齷, 諛, 琅", it all plays the same "initial" role.

4.2 configurational objectivity principle

The principle of configurational objectivity has two requirements. On the one hand, it refers to the result that an imagery-glyph or imagery-component must be composed of its imagery-component or imagery-icon. in other words, it cannot be separated from its imagery-component or imagery-icon and subjectively talk about imagery-component or imagery-glyph. For example, the imagery-glyph of "意" cannot be separated from the imagery-glyph of "音" and the imagery-glyph of "心", but is the result of their joint construction according to the combination mode. similarly, the imagery-glyph of "音" cannot be separated from the imagery-glyph of "立" and "日" are combined by them. the imagery-glyph of "心" cannot be separated from the imagery-icon of "乚" and "㇇", but is formed by them.

On the other hand, the imagery-glyph, imagery-component, and the imagery-icon are presented with its carrier, and the combination of imagery is also a combination of configurations. The regular script glyphs are already standardized, and more, less, or arbitrary movement to increase or decrease the strokes is not allowed and unacceptable. Because they represent imagery behind them, any irregular shape changes will destroy these imagery and their construction, and affect the consistency of the system. Therefore, the principle of configurational objectivity here again refers to taking these regular script configurations as objective phenomena themselves, and changing them will result in a change of rationale.

4.3 interpretability principle

Interpretability mainly refers to the analyzable interpretation relationship between the imagery-glyph and its meaning of a character. specifically, the imagery of a Chinese character (imagery-glyph) has a cognitive and empirical representational connection such as analogy, category, gestalt, and metaphor to the meaning of the character it refers to. This representational connection is the interpretability of imagery-glyph and its meaning of the character.

We know that there are three main forms of mental representation in our minds about things, ideas, and events in the external world: imagery, text, and propositional form. Since Pinyin text is phonetic, there is an arbitrary relationship between phonetics and the objects represented by vocabulary. Therefore, from the perspective of cognitive psychology, mental imagery and text are two opposite representations (dual-code theory): the former is an analogy metaphorical relationship, while the latter is an arbitrary relationship (or a conventional relationship). The former is more specific and the latter is more abstract. However, Chinese characters are different. Chinese characters are ideographic, which is achieved through imagery. This imagery is the basis for Chinese character configuration coding. The five combination modes we have compiled are the result of combining and constructing this imagery, and the result of constructing a glyph. Therefore, Chinese characters use categories (concept classification), analogies (imagery), metaphors, gestalt features, and other cognitive methods through imagery to form an empirical connection with the concept of the meaning of the character. This kind of connection is precisely the imagery representation of human concepts that originate from experience situations. The ideographic relationship between the glyph and its meaning of Chinese characters is not an arbitrary relationship, but the glyph itself contains rich cognitive information, which is the representation of the experience imagery produced by its own meaning. This is the unique ideographic characteristic of Chinese characters which is different from Pinyin characters.

Chinese characters are the double-code unification of the imagery formed by its meaning and the symbol encoded by the imagery. This double-code unification relationship is interpretability. For example, the configurational imagery of the character "水"(water) is exactly "the swaying of a fish hook causes water to ripple." This configuration of "水" is empirically connected with the concept of "水", which is more pronounced emotional than the phonetic code of "water". Because this kind of experience is a common sense of public experience, this kind of common sense has universal resonance for understanding the meaning of characters. This is the greatest advantage of the direct representation of ideographic characters from experience. As long as you have this experience, following this code, understanding the meaning of a character is a natural process. This interpretability of Chinese characters is the basic principle of the theory of "imagery word-formation".

4.4 Chinese culture and common sense principle

Cognitive theory is the theoretical basis of the theory of "imagery word-formation", while these content of the imagery are the observation of traditional culture and common sense. After all, people and the environment produce cognition through the interaction of physical and mental perception. The representation of the cognitive result is language, and the sum of language and action is culture. Language, cognition, and culture jointly support the theoretical framework of the theory of "imagery word-formation". Therefore, returning to the history and culture of the creation of Chinese characters can be used as a verification of the imagery, but also as a basis for finding the imagery. Because the empirical connection between the imagery of the formation of the character and the meaning of the character mainly refers to conformity with the specific traditional culture and common-sense experience. Imagery that is inconsistent with traditional culture and common sense should be regarded as inappropriate.

The above four principles are necessary and sufficient conditions for the establishment of the imagery of Chinese characters, and also the guarantee of the scientificity and efficiency of the Chinese character system.

Operational strategies and methods on how to obtain imagery-glyph, imagery-component and imagery-icon, On the whole, we adopt the strategy of "reverse engineering", that is, through the way of tracing the cause through the results, we examine the imagery-glyph and imagery-component from the known meaning of the character to the inside of the Chinese character, layer by layer, down to the icon and its imagery-icon. Specifically, we used four methods.

One is to follow the commonly used component analysis or structure analysis in linguistics, that is, to find elements, hierarchical relationships, and structural patterns, and we use this method to sort out the aforementioned three-level two-code structure of Chinese characters.

The second is the induction. This is to use the same component to appear in many different characters in the system, or the same character to appear in many other characters, and most of the characters are polysemous. Therefore, to find out the imagery of a certain component, we need to find out all the other Chinese characters that contain this component and gather all their meanings, including all polysemous meanings, and then sum up the imagery of the component under the four principles of inferring the image. For example, if we want to find out the imagery of the character "𦰇" on the right side of the character "漢", we need to find the characters "漢, 難, 艱, 嘆, 歎, 嘆" from the system (nine thousand commonly used Chinese characters). Because they all contain "𦰇". then, based on all the meanings of these characters (including their polysemous meanings), the imagery of "𦰇" is summarized from these many meanings as a raft: an aquatic bamboo raft made by inflating animal skins, which is popular in the Hanjiang River Basin in Northwest China. Of course, this raft must conform to the four principles of speculating imagery

mentioned above. It must not only conform to historical facts, but also explain the numerous meanings of the aforementioned characters, and conform to the imagery-component of the lower component by its own "革夫". From here, we can see that the imagery-glyph being explored is the imagery of empirical culture, and it has rich content references. At the same time, it is also summarized from the multiple meanings of many related Chinese characters. Therefore, the imagery of the character formation can explain its the polysemy of character meaning. For example, the character "漢" has imagery of "the waters of the prevalent raft". Although this imagery directly reflects the waters of rivers, it is also a geographical concept, so the people here are called 漢人(the Hans), and the language used by these people is called 漢語(Chinese), the characters used are called 漢字(Chinese characters). even because the founder of the dynasty was a figure in this area, it was designated as the name of the dynasty (漢朝 Han Dynasty). That's how we are, we go deep into the nine thousand regular script Chinese characters to find out their unknown imagery-components using this induction method. Therefore, this is a huge, complicated, and long-term patient task, because it affects the whole body and requires constant adjustment.

Third, a large number of classics have appeared in Chinese histories, such as "ShuoWenJieZi", "KangXi dictionary" and other historical documents. They will have some predecessors' interpretations, including re-annotations of the classics. If some explanations conform to the above-mentioned four principles of inferring imagery, we may wish to use them. This literature analysis method is the third of our research methods.

Fourth, Chinese characters are ancient characters from their origin, such as antiques, which have historical textual value. Therefore, understanding various historical and cultural contexts and events is another way to explore the imagery-glyphs, imagery-components, and imagery-icons, or textual research. This textual research method is the fourth of our research methods. For example, when we understand the structure and performance of the Guqin, it is not difficult to imagine the imagery of the "彳山女" component in the characters "徽, 徵, 微, 徽": the wonderful fingers touch silk strings on the Guqin YueSan(古琴岳山). After exploring this imagery-component, this series of imagery-glyphs "徽, 徵, 微, 徽" can be easily solved.

In short, we have found another way to summarize the imagery-components and imagery-glyphs of Chinese characters, as well as the icons and imagery-icons of the whole system, and clarify the internal structure of Chinese characters, and the four principles of speculative imagery introduced ensure the scientific nature of the results, including interpretability is the core principle behind the theory of "imagery compose character", because interpretability ensures the empirical and cognitive connection between imagery and concept or meaning.

Figures

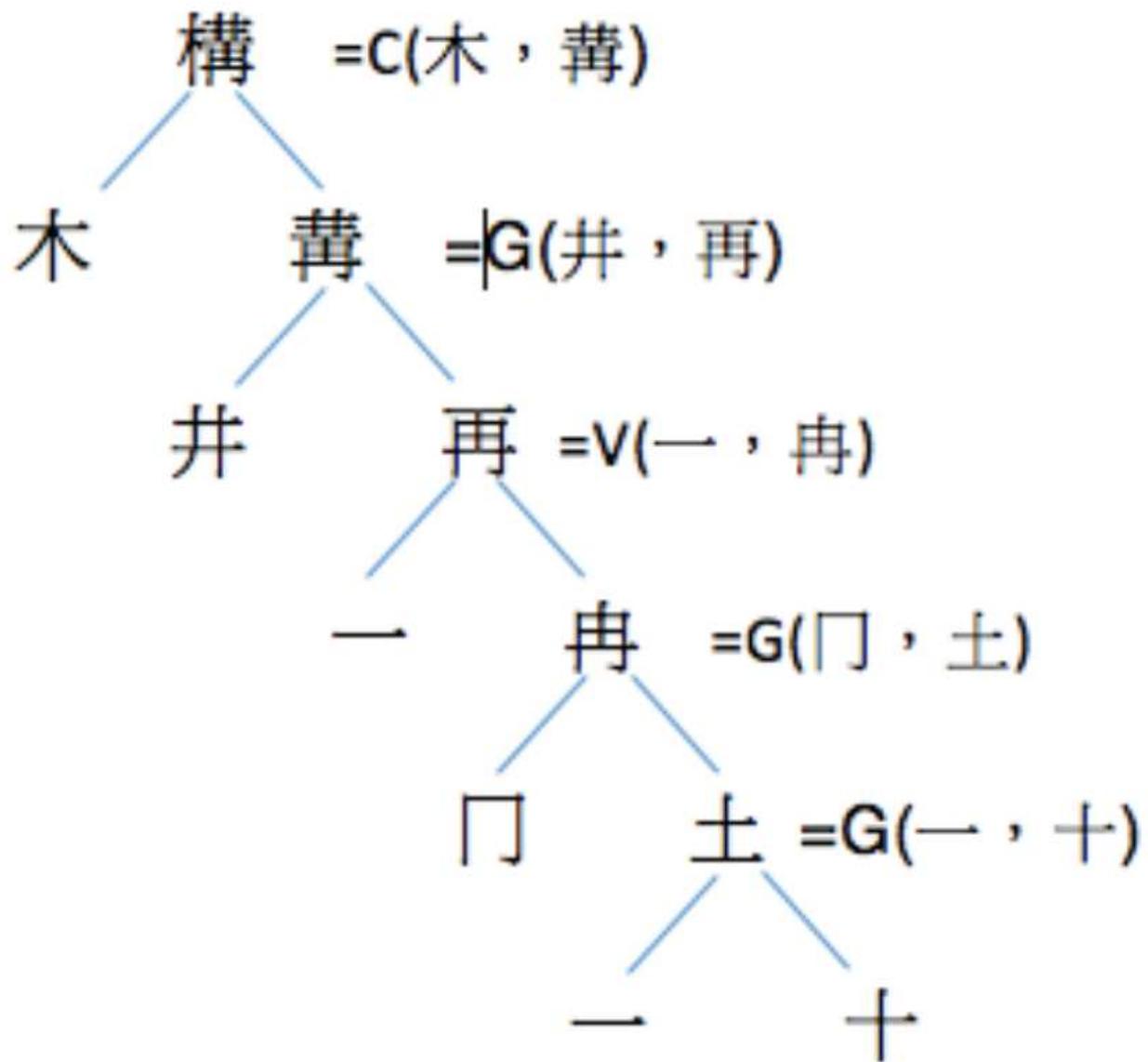


Figure 1

the structure of the character “構”

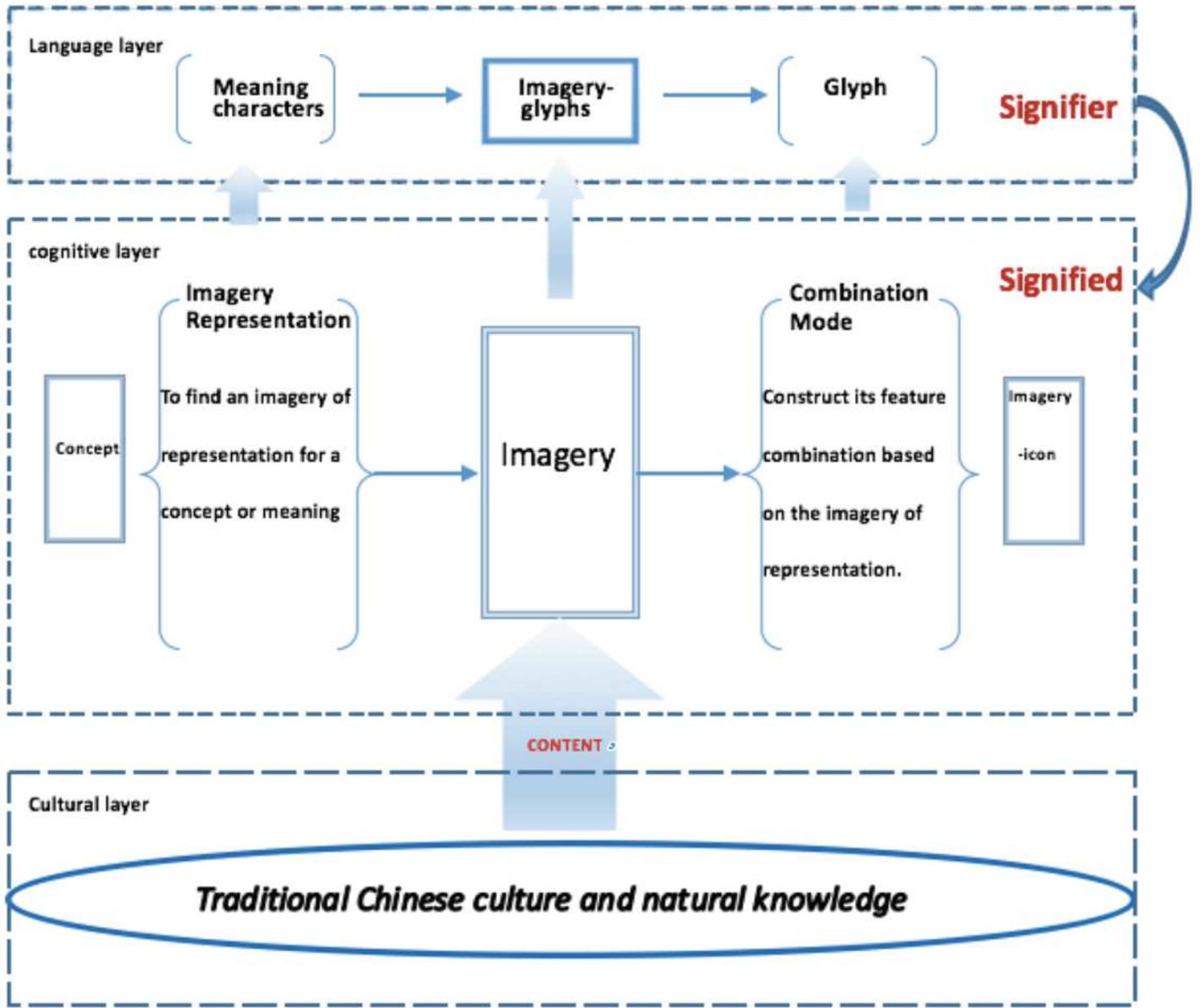


Figure 2

the basic principle diagram of Chinese word formation