

Automatic Generation Technology of Oracle Bone Characters Based on the Mechanism of Character-Building

Qinxia Wu^{1,2,3,*}, Hongdan Zhao¹

¹School of Computer & Information Engineering, Anyang Normal University, Anyang 455000, Henan, China

²Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China, Anyang 455000, Henan, China

³Henan Key Laboratory of Oracle Bone Inscriptions Information Processing, Anyang 455000, Henan, China

*Corresponding Author.

Abstract:

Because of the particularity of Oracle-bone inscriptions, the form of the character library is not suitable for Oracle-bone inscriptions. This study proposes a kind of automatic generation method of Oracle-bone characters based on the mechanism of character-building. Firstly, the method research on oracle-bone characters making principles and archetype matching theory, a mathematical model of automatic generation of oracle-bone characters based on the theory of point set topology; at the same time, this paper gives out the automatic generation of oracle-bone characters frame construction design system. Secondly, it studies the modeling methods of the structure, strokes and components of oracle bone characters, and proposes a computational model of oracle bone characters based on semantic components and their spatial positions. Thirdly, By defining the basic meta-strokes, combination meta-strokes and extension meta-strokes of oracle bone character, the structure of oracle bone character component is represented, and the combination of semantic component is realized by space transformation vector. An oracle-bone glyph description algorithm based on structure and components is designed to lay a foundation for oracle's explicated, classified, retrieved and coding, as well as a solid theoretical foundation for oracle's intelligent writing system.

Keywords: *Cognition, Oracle-bone character archetype, Component, The structure of Oracle-bone character.*

I. INTRODUCTION

Writing is the greatest invention of human beings. It is the reaction of human thoughts and the direct record of historical events. Oracle Bone Inscriptions is one of the four ancient characters in the world, the source of Chinese characters and the root of Chinese excellent traditional culture. With the development of computer science and technology, the research of Oracle Bone Inscriptions keeps pace with the times. In recent years, many scholars have applied artificial intelligence, big data analysis, deep learning and other technologies to the research of Oracle Bone Inscriptions. When Oracle Bone Inscriptions is combined with

modern science and technology, the construction of electronic informatization in Oracle Bone Inscriptions has become a reality. At present, a number of Oracle Bone Inscriptions information digitization platforms have been developed, such as "Yinqi Yuan Wen" jointly built by Oracle Bone Inscriptions Key Laboratory of Information Processing, Ministry of Education of Anyang normal University and the Research Center of Inscription, Yin and Shang History of Chinese Academy of Social Sciences, "Handa Library" of Chinese University of Hong Kong and so on, all of which have adopted computer technology to study Oracle Bone Inscriptions, providing great convenience for researchers^[1]. However, most experts and scholars have some misunderstandings in the research of informatization in Oracle Bone Inscriptions: 1) Pay attention to the textual research of the glyph, pronunciation and meaning of Oracle bones. The Oracle bones information processing is regarded as the processing of knowledge, but the Oracle bones itself is not regarded as a kind of information; 2) Pay more attention to differences than unity. There are a large number of variant characters in Oracle bones, so we pay attention to the differences of the shapes and components of each Oracle bone, and despise the unity between the shapes and components of Oracle bones^[2]; 3) Paying attention to theory rather than system. Attaching importance to the application of the research experience of Oracle bones to information processing, and forgetting to consider the systematicness of information processing. In this way, the information processing in Oracle Bone Inscriptions becomes fragmentary and cumbersome, the data information cannot be used universally, and the data stored and processed is huge^[3,4].

With the continuous cognition of human beings to the objective world, the language and characters come into being. The composition of Oracle Bone Inscriptions's glyph reflects the abstraction, summarization, analysis, induction and reasoning of human cognition of the objective world, and regulates human cognition and psychological process of nature. Oracle bone inscriptions, which are mainly pictographic and ideographic, describe everything in nature in the most concise way, and Oracle bones itself is a prototype of abstract generalization. The emergence and application of Oracle bone characters in daily production have enabled human beings to master the cognitive law of objective things. Oracle Bone Inscriptions already has a complete composition system, which divides Oracle Bone Inscriptions into glyphs and components. Compared with Oracle Bone Inscriptions's work of identification, textual research, conjugation and so on, the research on Oracle Bone Inscriptions components is of more important significance^[5]. Based on the scientific achievements of cognitive psychology, this paper studies and summarizes the mechanism of character-building of Oracle-bone inscriptions. On the basis of analyzing the principle of oracle bone character formation, a calculation model of Oracle-bone inscriptions based on semantic components and their spatial positions is proposed by studying the structure, strokes and prototype (component) modeling method of Oracle-bone inscriptions; By defining the basic pen elements, combined pen elements and extended pen elements of Oracle bone characters, the structure of Oracle bone characters is represented, and the combination of prototypes is realized by spatial transformation vectors, and the scheme of intelligent spelling and word formation of Oracle bone characters is put forward. The mathematical model of automatic generation of Oracle bone inscriptions based on cognitive mechanism is established.

II. PRINCIPLES OF AUTOMATIC GENERATION OF BONE CHARACTERS

2.1 Cognitive Principle of Bone Inscriptions

Cognition is the process of people's understanding of nature and human social activities, and Oracle bones is the product of human cognition and practice of nature in Shang Dynasty. The prototype matching theory in cognitive psychology holds that the representation of the most basic features of a thing is stored in the human brain, and this representation is the abstract form of a certain kind of thing. This abstract form of representation is "prototype". Human beings recognize a kind of things, and the perception of this kind of things is realized by matching the input information with the "prototype" stored in mind.

According to the above analysis, with the deepening of research and archaeological discovery, the object of Oracle bone characters is increasing, and as many cognitive objects as possible can produce as many Oracle Bone Inscriptions characters. The number of the Oracle bones is uncertain, and the prototype of Oracle bones is a limited collection. The word "pictograph" occupies a large proportion in Oracle Bone Inscriptions characters. It refers to drawing the appearance of things with simple abstract lines according to their characteristics, such as day (☉) and month(☽,☾). In addition, a common part of characters in Oracle bone inscriptions are "Zhi Shi", which refers to the character-building method of using abstract symbols to express certain meanings, such as the above (☺) and the next (☹). Therefore, the Oracle bone inscriptions are basically a combination of the basic character sets of "pictograph" and "Zhi Shi" according to certain structural rules, which embodies the process that ancestors knew and described things. For example: 𠄎

(Xi) → 𠄎 (Only), it means catching birds by hand. 𠄎 (claw) + 𠄎 (horn) + 𠄎 (cow) → 𠄎, the hand grabbed the part of the horn, which means to solve the cow, etc [6].

2.2 Composition Principle of Chinese Characters

The character "Jia" is a kind of objective thing, and its prototype consists of "pictogram" and "reference" symbols. The Oracle bones prototype is the most basic component of the Oracle bones character set, which is called "Oracle bones component" when it is mapped to specific Oracle bones. The Oracle character is composed of "Oracle character components" according to a certain structure. When the Oracle character is composed, the relationship between the Oracle character components and between the Oracle character components and the Oracle character as a whole constitutes the hierarchical structure of the Oracle character as shown in Fig 1.

According to the structure of the released Oracle bones characters and referring to the configuration rules of modern Chinese characters, twelve structures (☐☐☐☐☐☐☐☐☐☐☐☐) are currently used to represent the composition positions of Oracle bones characters. Describe the Oracle bone characters according to the hierarchical relationship. For example: 𠄎 can be described as ☐𠄎☐𠄎☐𠄎; 𠄎 can be described as ☐𠄎☐𠄎☐𠄎.

The technology of automatic generation of Oracle bones is to change the prototype of Oracle bones into Oracle bones components, and then combine them into Oracle bones according to a certain structure. The principle of automatic generation of Oracle bones is shown in Fig 2.

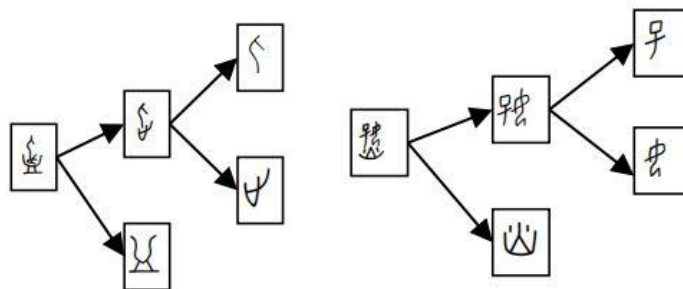


Fig 1: Hierarchy of oracle bones

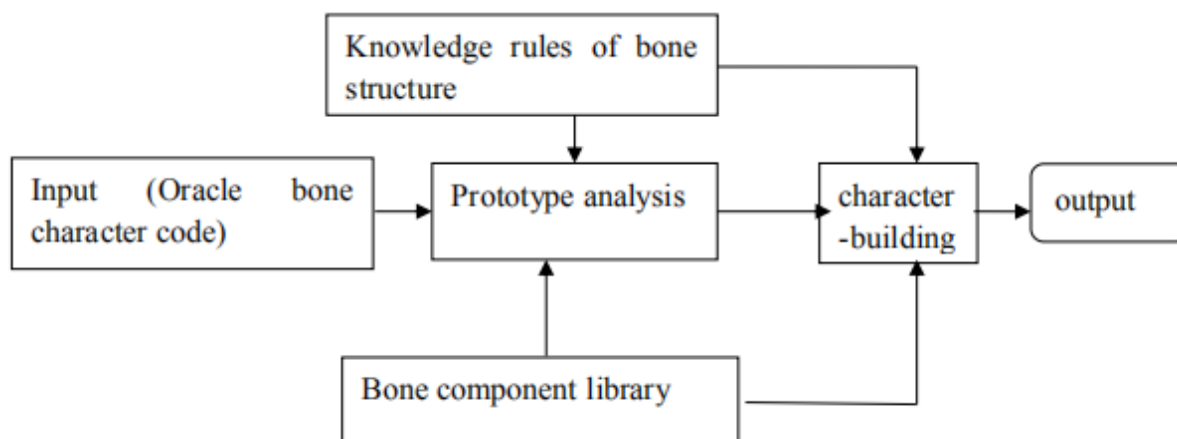


Fig 2: Prototype of oracle bone inscriptions based on cognitive mechanism

To realize the automatic generation of Oracle bones, first of all, the prototype database of Oracle bones and the structure knowledge database of Oracle bones should be established. By analyzing the character-building principle of Oracle bone inscriptions, the character-building knowledge is organized and the intelligent character-building knowledge base is established. When fonts are automatically generated, external stimuli are needed for input at first. According to the rules of the character-building knowledge base, the Oracle bones prototype is mapped into Oracle bones components, and the components are combined into Oracle bones according to a certain Oracle bones structure ^[7,8].

1. Bone coding. The realization tool of automatic generation of Oracle bones is computer. Only by converting the project of automatic generation of Oracle bones into a code that is convenient for computer processing can the automatic generation of Oracle bones be realized. According to the structure of Oracle bone characters and the character-building principle of Oracle bone characters, the coding method of Oracle bone characters is given. The essence of the Oracle character coding processing is to decompose the

Oracle character coding into the corresponding Oracle character structure and Oracle character prototype to establish the Oracle character prototype combination tree.

2. Oracle bones prototype library. The Oracle bones prototype is the basic unit of Oracle bones. When Oracle bones are automatically generated, the Oracle bones prototype is mapped to the Oracle bones structure. The Oracle bones prototype database includes the concrete composition of Oracle bones prototype, the number of Oracle bones prototype and the computer representation of Oracle bones prototype.

3. Structural knowledge of Oracle bones. The composition of Oracle bone characters is the process of generating Oracle bone characters by nail bone characters according to certain combination rules, and describes the positional relationship of Oracle bone characters primitives in Oracle bone characters. Only with the structure of Oracle bones can all Oracle bones character sets be generated with a limited number of Oracle bones prototypes.

4. Prototype mapping of Oracle bones. The transformation from prototype to object is a kind of transformation that keeps the topology unchanged. The automatic generation process of Oracle bones is the composition of Oracle bones components in the plane according to the rules of Oracle bones composition. From the Oracle bones prototype to the concrete Oracle bones component is the mapping from abstract to concrete, which keeps the topology of the Oracle bones prototype unchanged.

III. MATHEMATICAL MODEL OF AUTOMATIC GENERATION OF ORACLE BONE INSCRIPTIONS

According to the theory of automatic generation of Oracle-bone characters based on cognitive mechanism, in all released Oracle-bone characters, the character-building unit with independent semantics is the Oracle-bone character prototype, and each Oracle-bone character in the Oracle-bone character set consists of one or more Oracle-bone components, which are the mapping of the Oracle-bone character prototype in the Oracle-bone character structure space.

3.1 The Extraction Principle of Oracle Bones Prototype

The principles to be adopted for the prototype extraction of Oracle bones are as follows: 1) Only basic components with independent semantics are stored in the prototype library. For example, 𠄎 can split the components 𠄎 (people) and 𠄎 (Yin) according to the left and right structures, while the components 𠄎 (Yin) will not be split. 2) Count the word-building frequency of prototype components. Select components with high word-building frequency to join the prototype component library. 3) Point components are not included in the prototype component library. Because Oracle Bone Inscriptions's writing is different in simplicity and complexity, a little more and a little less does not affect the expression of meaning, such as 𠄎, 𠄎, 𠄎, 𠄎, 𠄎, 𠄎. The treatment of dots is only used as the basic strokes of Oracle bones. 4) Treatment of identical components. Due to the immature development of Oracle Bone Inscriptions's

strokes, there are many components with the same shape. This kind of component has different shapes but similar word-building functions, so it is regarded as the same component to join the component prototype library.

The prototype library is the basic component unit of Oracle bone inscriptions. When Oracle bone inscriptions are automatically generated, they are composed of images mapped from prototype components to Oracle bone inscriptions. The component library includes the concrete composition of the Oracle bone inscriptions prototype, the number of Oracle bone inscriptions prototype components and the computer representation of Oracle bone inscriptions prototype components. The prototype library is the cornerstone of the automatic generation of Oracle bone inscriptions.

3.2 Mathematical Description of Bone Characters

Because the number of Oracle bones characters can't be quantified, according to the cognitive principle, a limited number of Oracle bones prototypes can generate all Oracle bones characters. According to this principle, the following definitions are given:

Definition 1: Let Ω represent all Oracle bone character sets in a certain period of time, and C_i is the element in Ω . R is a set of real numbers, and the mathematical expression of Ω is as follows:

$$\Omega = \bigcup_{i=1}^N \{C_i\} \quad (m \text{ a } \mathbb{N}() \leq |C_i|, \varepsilon \in R) \quad (1)$$

Definition 2: Let P be the set of all prototypes of Oracle bone character set, because the absolute value of P is less than the absolute value of Ω , so P is a finite set; Let p_i be the element in the P set, and the mathematical expression of the set composed of all Oracle bone character prototypes is as follows:

$$P = \bigcup_{i=1}^N \{P_i\} \quad (\sup(N) \leq |M|, M \in R) \quad (2)$$

According to definitions 1 and 2, we can get the automatic generation mechanism of Oracle bones, that is, in the Oracle bones prototype set P , we can find a limited cover of Ω . That is, all Oracle bones are composed of a limited number of Oracle bones prototypes through the structural mapping of Oracle bones prototypes in Oracle bones. An Oracle bone inscription can be expressed as the distribution of one or more Oracle bone inscription components in the Oracle bone inscription space; The Oracle-bone component is the topological mapping of the Oracle-bone prototype on the Oracle-bone structure. The specific appearance, size and position of the Oracle-bone component in the Oracle-bone structure will be jointly determined by the Oracle-bone prototype and the Oracle-bone structure.

According to the above-mentioned mechanism of automatic generation of Oracle bone inscriptions, the mathematical model of automatic generation of Oracle bone inscriptions is as follows:

Definition 3: Let S be a set composed of all the Oracle bone structures, the absolute value of S is less than the absolute value of Ω , and S is a finite set; The collection of bone structures can be described by the following formula:

$$S = \bigcup_{j=1}^N (S_j) \quad (\max(N) \leq |M|, M \in R) \quad (3)$$

S represents all subsets of the structure set S of the Oracle bone inscription, then the structure set S of the Oracle bone inscription is a topological space (S, τ) relative to its subset S . Assuming that P represents all subsets of the Oracle bones prototype component set P , then P is a topological space relative to its subset P , which is denoted as (P, τ) . Then the subset family $\mathcal{x} = \{U_1 \times U_2 \mid U_1 \in S, U_2 \in P\}$ of cartesian product $X = S \times P$ is a base of X , and the topological space (X, \mathcal{x}) is the product of topological space (S, τ) multiplied by topological space (P, τ) . Suppose a subset of X can be expressed as: $x_i = \{(s_i, p_i) \mid s_i \in S, p_i \in P\}$. Since the structure set S and prototype component set P of Oracle bone inscriptions are bounded closed sets, according to the compact space theorem, we can wait until S and P are both compact space sets, and then the cartesian product space X is also a compact space set, then there is a continuous space mapping $F: X \rightarrow C_i$ that enables the prototype component of Oracle bone inscriptions to generate new Oracle bone inscriptions according to the structure mapping of Oracle bone inscriptions, and The new Oracle character C_i can be expressed as: $C_i = \bigcup_{x_i \in X} F(x_i) = \bigcup_{x_i \in X} F(s_i, p_i)$, which is a mathematical model automatically generated by Oracle bone prototype components according to its composition, where F represents homeomorphism mapping, that is, F can map the Oracle bone prototype to the structure of specific Oracle bone characters. However, in Oracle bones characters, there are many differences in each word-forming component, but the essential semantic information it represents is always. In order to keep the topological properties of the Oracle bones prototype unchanged; $F(s_i, p_i)$ is the component of Oracle bone, which is determined by Oracle bone prototype p_i and Oracle bone structure [9].

3.3 Mathematical Model of Automatic Generation of Bone Characters

The mathematical model is automatically generated according to Oracle bone inscriptions, and the steps of automatic generation of Oracle bone inscriptions are as follows:

1. According to the mathematical formula description of the structure of Oracle bones, select prototype of Oracle bones.

2. Map the Oracle bones prototype p_i to the Oracle bones structure s_i to obtain the Oracle bones component $F(s_i, p_i)$;







3. In the figure space of Oracle bones, use the union of one or more Oracle bones components $F(s_i, p_i)$ to automatically generate Oracle bones C_i .

4. Unified treatment of glyphs. To select the fonts and glyphs of the automatically generated Oracle bone inscriptions, it is necessary to construct a function Tfs (F is glyph, S is font size) to uniformly process the generated Oracle bone inscriptions.

On this basis, the prototype theory in cognitive psychology is deeply analyzed, and it is concluded that the object of Oracle bone inscriptions cannot be quantified, but the prototype of Oracle bone inscriptions is limited. Prototype component is the basic component of Oracle bone inscriptions. On this basis, according to the particularity of Oracle characters and the efficiency of computer processing, the prototype extraction criteria of Oracle characters are summarized. Taking 1024 Oracle bones characters that have been released as experiments, the prototype component extraction experiment is carried out, which lays a foundation for the automatic generation of Oracle bones characters.

When the shape of the prototype component changes continuously during word formation, the information of points, lines, blocks, etc. of the geometric figure composed of points of the storage component should remain unchanged from the original data information. As shown in TABLE I, the Oracle bones stored in the prototype library are, and the appearance of the components in different Oracle bones can be obtained through topology change.

TABLE I. Components corresponding to prototype

Bone prototype	compon ent 1	compon ent 2	compon ent 3	compo nent 4	compo nent 5
					

IV. BONE CHARACTER GENERATION PLATFORM

4.1 Introduction to the Function of Oracle Bone Character Generation Platform

According to the mathematical model of automatic generation of Oracle-bone characters, Oracle-bone characters are composed of the mapping Oracle-bone character components from Oracle-bone character prototype to Oracle-bone character structure. The platform for automatic generation of Oracle bones should have the kinetic energy of input, display, editing and saving, as shown in Fig 3.

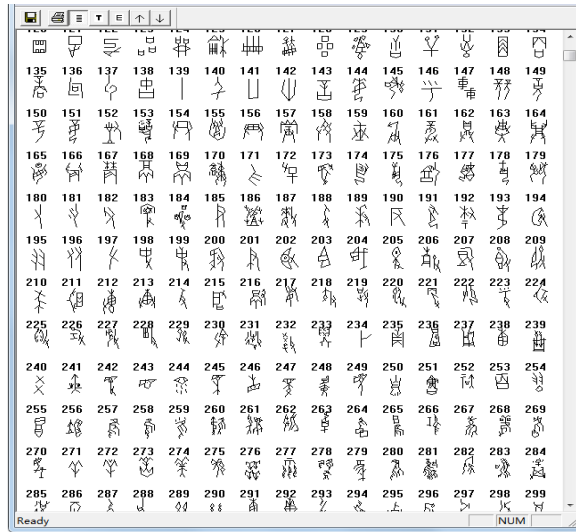


Fig 3: Oracle bone character generation platform

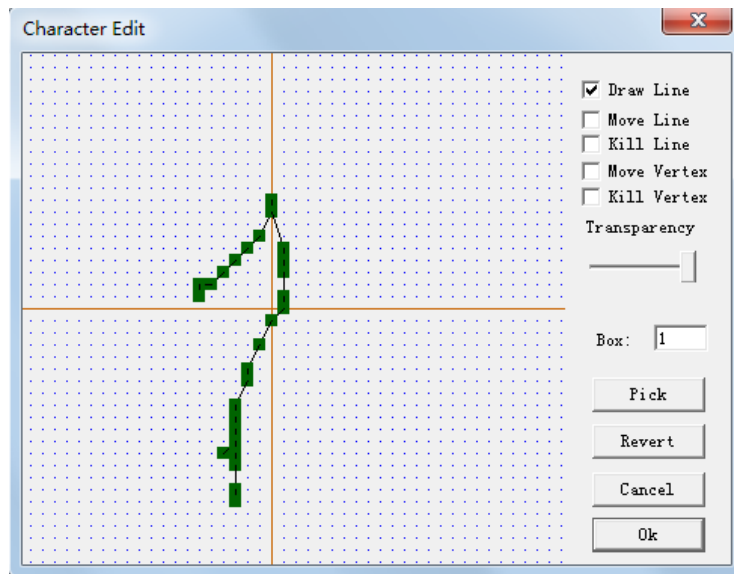


Fig 4: Prototype components of Oracle bones

Keyboard input function: the Oracle bone characters are encoded, and the external code of Oracle bone characters maps the structure information of Oracle bone characters to 26 letter keys on the keyboard.

Coding processing function: analyze the Oracle bone characters coded according to the "Oracle bone character prototype +Oracle bone character structure", and transform the analysis results into specific data, namely the structure of Oracle bone characters and the prototype of Oracle bone characters, which is convenient for the automatic generation of Oracle bone characters.

Automatic generation module: extract the mapping knowledge of each primitive from the knowledge base, read the characteristic data points of each primitive from the hard disk of the computer, and make a new Oracle bone inscription after affine change.

Opening and saving function: the newly-created Oracle bone inscription only needs to save the composition code of the inscription. At the same time, the file opening function is designed to open the data saved by the user, and the program will deduce the word according to the coding of the word and make it up and display it on the screen.

4.2 Results and Analysis

According to the structure coding information of Oracle bones, the Oracle bones automatic generation system generates new Oracle bones under the guidance of prototype component library. The platform for automatic generation of Oracle bones provides three functions: input, output and editing of Oracle bones information. Every prototype in the Oracle bones prototype database is saved by feature points. Now, more than 400 Oracle bones prototypes have been summarized, with a storage capacity of 156KB. By applying this principle, 1,202 commonly used Oracle bones have been generated. The storage of the prototype components of the Oracle bones is shown in Fig 4.

The composition of Oracle bones is the topological mapping of prototype components on Oracle bones structure, that is, the position, size and shape of prototype components in different structural Chinese characters may be different. However, the corresponding components of the same prototype component in different Chinese characters have the same topological structure as the prototype component. As shown in Fig 5, the Oracle bone character "north" can be composed of the prototype component of "human" being according to the left and right structures.

The geometric transformation method of component mapping from prototype to specific Oracle bones is a problem that must be solved. The experimental platform adopts the automatic acquisition method of affine transformation coefficients based on region and feature extraction, and obtains the affine transformation coefficients of prototype components from different angles and different technical methods, so as to form complementary advantages and acquire more accurate mapping knowledge of prototype components^[10].

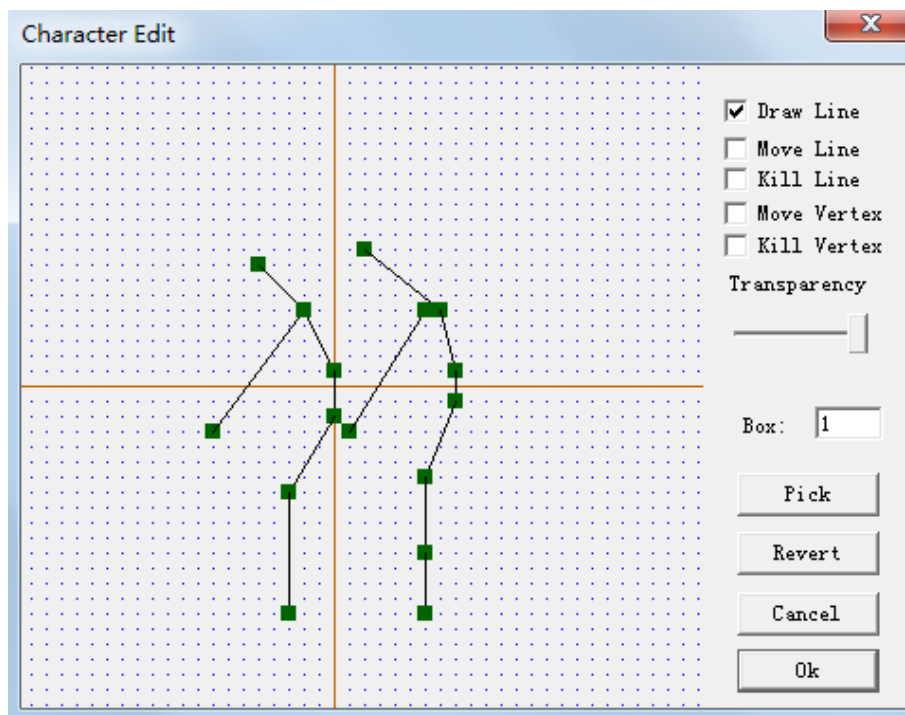


Fig 5: Oracle bone characters generated by components and pen elements

V. CONCLUSION

Based on the analysis of the character-building principle of Oracle-bone inscriptions, by studying the modeling method of the structure, strokes and prototype components of Oracle-bone inscriptions, an automatic generation method of Oracle-bone inscriptions based on the character-building mechanism is put forward to replace the word bank, instead of taking Oracle-bone inscriptions as the basic unit of information processing, the prototype components of Oracle-bone inscriptions are used as the basic unit of word grouping. It solves the problems of inputting, displaying and editing Oracle bones without uniform Unicode code. This method can solve the problem that Oracle bones can't be quantified and finalized, sum up the rules of intelligent character-building, and the new Oracle bones will be generated according to the rules of character-building, so that a stable informatization standard of Oracle Bone Inscriptions can be formulated. Both the Oracle bones prototype database and the intelligent character-building knowledge base are based on the cultural characteristics of Oracle Bone Inscriptions, which facilitates the learning and cognition of Oracle bones by Oracle Bone Inscriptions fans. The number of Oracle bones prototype database is far less than the number of Oracle bones characters. The information entropy in information processing will be greatly reduced, the information efficiency will be greatly improved, and the Web transmission of Oracle Bone Inscriptions will be facilitated.

ACKNOWLEDGEMENTS

This research was supported by Science and technology Development project of Henan Province (Grant No. 212102310300; Grant No. 222102320189) and "Oracle Information Processing" Innovation team of Ministry of Education (Grant No. 2017PT35)

REFERENCES

- [1] Song Zhenhao. Carry forward the research work of Oracle Bone Inscriptions in the new era. *Language strategy research*, 2019(6):7-8.
- [2] Jiang Minghu, Deng Beixing, Liao Panpan, et al. The establishment of Oracle Bone Inscriptions font and intelligent knowledge base. *Computer engineering and application*, 2004, 40(4):45-47.
- [3] Ma Xiaohu, Yang Yiming, Huang Wenfan, et al. Research on the generation technology of outline glyphs in Oracle Bone Inscriptions and the construction of general Oracle Bone Inscriptions font. *Language application*, 2004(3):105-111.
- [4] Shi Lei, Han Dezhi, Zhang Xiafen, et al. New style calligraphy character synthesis based on contour fitting. *Computer applications and software*, 2017, 34(6):46-51.
- [5] Gu Shaotong, Ma Xiaohu, Yang Yiming. Research on Oracle Bone Inscriptions input coding based on font topology. *Journal of Chinese information*, 2008, 22(4): 123-128.
- [6] Wu Qinxia, Li Qingsheng, Gao Feng. Research on automatic generation technology of Oracle Bone Inscriptions font based on semantic components. *Journal of Peking University: Natural Science Edition*, 2014, 50(1):161-166.
- [7] Gu Shaotong. Oracle Bone Inscriptions glyph recognition method based on fractal geometry. *Journal of Chinese information*, 2018, 32(10): 138-142.
- [8] Jun Guo, Changhu Wang, Edgar Roman-Rangel, Hongyang Chao, Yong Rui. Building Hierarchical Representations for Oracle Character and Sketch Recognition. *IEEE Transactions on Image Processing*, 2016, 25(1):104-118.
- [9] Han Gefei, Gu Shaotong, Yang Yiming. Method for extracting glyph features of Oracle bone rubbings based on mathematical morphology. *Journal of Chinese information*, 2013, 27(2):79-85.
- [10] Chengdong Liu, Lian Zhouhui, Tang Yingmin, et al. Automatic generation system of high-quality Chinese fonts based on component splicing. *Journal of Peking University (Natural Science Edition)*, 2018, 54(1):35-41.