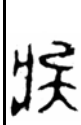
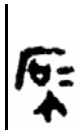
IRGN1119Bronze

Encoding Sample of Bronze Inscription

Serial No.	Rep. Script/ Glyph	Original Shape/ Glyph	Source	Period/ Epoch	Area/ Terrain	Material	Radical	Glyph Determ.	Corresp. Modern Char	Notes
1			叔矢方鼎 文物 2001 年 08 期 9 頁圖一二	西周早期		青铜器	示	礿		
2			宙母丁鼎 集成 4.1704	西周早期		青铜器	中	宙		
3			季宙父乙尊 集成 11.5797	西周早期		青铜器	中	宙		
4			孟卣 集成 10.5399	西周早期		青铜器	中	宙		集成 10.5399；集成 11.5619；集成 11.5576；集成 16.9972；集成 16.9972
5			宰甫卣 集成 10.5395	殷		青铜器	中	宙		集成 10.5395
6			上曾大子鼎 集成 5.2750	春秋早期	上曾	青铜器	口		哀	集成 8.4330
7			哀成叔豆 集成 9.4663	春秋晚期		青铜器	口		哀	集成 9.4650；集成 5.2782
8			豆閉簋 集成 8.4276	西周中期		青铜器	豆		豆	集成 8.4276
9			周生豆 集成 9.4683	西周晚期		青铜器	豆		豆	集成 9.4682
10			宰甫卣 集成 10.5395	殷		青铜器	豆		豆	集成 10.5395；集成 16.10051

11			散氏盤 集成 16.10176	西周晚期		青铜器	豆		豆	集成 16.10176
12			叔猴父簋蓋 集成 7.4069	西周晚期		青铜器	片	猴		集成 7.4068；集成 7.4070
13			應侯見工簋 文物 2002 年 07 期 73 頁圖一、二	西周中期	應	青铜器	网	羹	向	
14			多友鼎 集成 5.2835	西周晚期		青铜器	宀			
15			叔向父簋 集成 7.3849	西周晚期		青铜器	宀		向	集成 7.3850；集成 7.3849；集成 7.3853；集成 7.3851；集成 7.3852；集成 7.3855；集成 8.4242；集成 7.4033；集成 7.4034；集成 12.7306；集成 10.5250；
16			叔向父爲備簋 集成 7.3870	西周晚期		青铜器	宀		向	
17			向方鼎 集成 4.2180	西周早期		青铜器	宀		向	集成 6.3572；集成 16.10567；集成 10.5250；集成 14.9010
18			九年鄭令矛 集成 18.11551	戰國	鄭	青铜器	宀		向	集成 18.11565
19			向斿鼎 集成 3.1349	戰國晚期	待考	青铜器	宀		向	集成 3.1349
20			六年令戈 集成 17.11337	戰國早期	待考	青铜器	宀		向	
21			耳尊 集成 11.6007	西周早期 或中期		青铜器	宀	鼏		

22			仲帶鉤 集成 16. 10405	戰國	待考	青铜器	石			
23			遘鼎乙 考古與文物 2003 年 03 期 9 頁圖一五、一六	西周晚期		青铜器	阜	隩		
24			戈 集成 17. 10683	殷		青铜器				
25									宀	不成字构件
26									宀	不成字构件

Encoding Sample of Bronze Inscription and Propositions to Encoding Principle of Bronze Inscription

1. Source of character selection

1.1 The source of the Bronze Inscription encodings include in all published materials to the present of rubbings and pictures of Bronze Inscription, which amounts to no less than 13000 vessels and a collection of nearly 130,000 glyphs.

1.2 *The Digital Processing System of Bronze Inscriptions of Shang and Zhou Dynasties* (computer program) was developed by the Center for the Study of Chinese Characters and Their Applications (CSCCA), East China Normal University, and published in 2003. It includes all rubbings of *Yin Zhou Jinwen Jicheng* and *Jinchu Yin Zhou Jinwen Jilu*, and is being updated to include the latest materials. It covers all bronze vessels and the collection of glyphs aforesaid and is used in the character selection.

2. Rules of character selection

2.1 The principles of character selection are completeness and authenticity. The encoded characters should be able to map to all existing glyphs and to meet various needs, especially the need of expert's research.

2.2 The rules of selection

2.2.1 Glyphs (non-character components included) of different construct are encoded separately.

2.2.2 glyphs with significant component variance are encoded separately, though the basic construct of the glyph may be retained.

2.2.3 glyphs with same construct and components but with great difference in component's position, direction and line curves are encoded separately.

2.2.4 Only the complete glyphs are selected. The fragmentary glyphs and the transcriptions are excluded.

3. Collation of characters

The Bronze Inscriptions are classified into five categories:

3.1. Glyphs which have corresponding characters in Shuowen Jiezi are sorted as they were in Shuowen Jiezi (540 radicals).

3.2 Glyphs which have no corresponding characters in Shuowen Jiezi or the glyphs are still unknown will be assigned a radical and sorted according to the radicals of Shuowen Jiezi and placed to the rear of the corresponding radical and ordered according to their stoke number and the type of the first stroke.

3.3 Glyphs which can not correspond to the radicals of Shuowen Jiezi will be placed after the two categories aforesaid and ordered according to their frequencies of use, the higher comes first.

3.4 Compound graphs (合文) will be sorted according to the Hanyu Pinyin of the two corresponding Kaishu （楷书）characters and placed after the category of 3.3.

3.5 The non-character components（不成字构件）split from the compound graphs are ordered according to the Hanyu Pinyin of their corresponding Kaishu terms and placed after the category of 3.4.

For the first four categories, different glyphs mapping to one modern character will be ordered according to their periods, the earlier comes first.

4．Encoding Procedure

4.1 Isolate the rubbings or pictures of Bronze Inscriptions in terms of glyph into about 130,000 individual units. Then analyze the construct and writing features of the character/glyph one by one according to the rules of character selection, and then label attributes on the basis of analysis for computer processing.

4.2 Scrutinize, unify or separate the glyphs carefully on the basis of rule of character selection to bring forward the scripts/glyphs to be encoded.

4.3 Correlate the encoded scripts/glyphs to all corresponding glyphs in Bronze inscription literature to guarantee that the encoded scripts cover all existing glyphs.

4.4 Fill in the Attribute Table as required

4.5 Order the encoded characters according to the rules of character collation.

5. Notes on the Encoding Sample of Bronze Inscription

5.1 “The representative script/glyph” is attained directly from the corresponding original glyph image with necessary processing.
5.2 In the row of “Original Shape/glyph”, only the original glyph image that the representative script/glyph corresponds to is listed.
5.3 The number and source of the glyphs the “Representative Script/glyph” represent are listed in the row of “Notes”

6. Propositions on the encoding of Bronze Inscription

6.1 A new attribute of “Coverage” should be added in the encoding. This is to avoid the deficiency of correspondence of the encoded scripts to the existing Bronze Inscription glyphs. Coverage means the extent of correspondence of the encoded scripts to the glyphs of existing materials. It can be established by marking the frequencies of use of the encoded scripts in all materials and the source of frequencies. If the coverage of each encoded script/glyph is precisely defined, the complete coverage of the encoded scripts would be ensured. To reach this goal, the complete analysis of existing materials and the scrutiny of glyphs one by one are prerequisite.

6.2 In character selection, the area and period of the characters should not be the basis and standard of encoding classification. In a character set, the uniqueness of the glyph is the basic requirement for the encoded scripts. Therefore, the form of the glyph should be the only principle in character selection. In the encoding of Bronze scripts, there should not be sub-categories as Bronze Inscriptions of Shang, Zhou, or Warring Kingdom Chu and Jin. Otherwise, the Bronze Inscription scripts would include many writing variants without distinctive difference.

6.3 Bronze Inscription scripts characterize in both construct and aesthetic style. Therefore, the fidelity of the representative script to the original shape should be retained. The representative scripts should be obtained directly from the original glyph image, rather than from the manual drawing, which probably lead to distortion. Only for the original glyphs with flaw, is manual modification needed.

6.4 The unknown and hetero-deciphered Bronze Inscription glyphs should also be encoded. This is because the non-disputable Jinwen scripts are in the minority. If the majority of unknown and hetero-deciphered Jinwen scripts are left out, the value and usefulness of the character set will be diminished. The Bronze Inscription scripts, in the first place, are primarily of interests to scholars.

6.5 No limit should be set to the number of the encoded glyphs of one character. All glyphs with distinctive difference should be encoded, to avoid deficiency of the character set and to facilitate the future research of Old Hanzi.

Center for the Study of Chinese Characters and Their Applications, (CSCCA)
East China Normal University,

Shanghai, China

Jan. 2005

金文编码字符样例说明

一、选字的材料范围：

1．金文编码字符集选样范围为迄今已有的金文铭文拓片，计 13000 多器，文献本体总字数近 130000。

2．金文字样的选样具体利用了《商周金文数字化处理系统》（华东师范大学中国文字研究与应用中心研制，广西金海湾电子音像出版社、广西教育出版社 2003 年版）。该系统包括了《殷周金文集成》、《近出殷周金文集录》所收材料及其他最新发表的材料。

二、选字原则：

依据选字求全，满足各个层次用字需求的原则，选收字样为：

1．形体结构不同的字形（所谓“字形”包括不独立成字的构件，下同）。

2．形体结构相同，但组字的各个构件形体有较大差别的字形。

3．形体结构相同，构件也相同，但构件有方向、部位、线条曲折等重大差异的字形。

4．原则上只选择比较清晰的字形，不收残泐字形及传世摹本字形。

三、字形排序：

字头的排列顺序分五个层次。

1．字形能与《说文解字》相对应的字，按《说文解字》的顺序排列。

2．《说文解字》未收或该字尚未认识，因此无法与《说文解字》相对应的字，归入相应《说文解字》部首，排列在见于《说文解字》的字之后，按笔画数及起笔类型排序。

3．无法归入《说文解字》部首的字，排在部首收字之后，并按其使用频率从高到低排列。

4．合文字形以所对应的两个楷书字的汉语拼音顺序排列在无法归入《说文解字》部首字之后。

5．从金文合体中切分出的金文不成字构件依据构件对应的楷书名称的汉语拼音顺序排列在合文字形的后面。

前四个层次的形体中，属于同一个楷书字头的不同字形以字形所在时代顺序从早到晚排

列。

四、字样编码过程：

1．对已有近 13 万字金文铭文拓片进行逐字切分，根据上述选字原则对切分的拓片逐字进行结构分析和笔势特征分析，并对分析结果进行可以作计算机统一处理的标注。

2．根据 1 所完成的标注，按选字标准进行细致地认同归并，从而产生编码字样。

3．对被选定的编码字样所代表的实际铭文用字逐个进行对应关联，最终达到编码字样对实际铭文用字的完全覆盖。

4. 将所有编码字样的有关属性填入表格相应的栏中。

5．将所有编码字样按照规则排序。

五、对本次提交字样的若干说明：

1．“代表字”直接通过对所出金文字形拓片的图象处理获得。

2．“原形字”栏中只列出“代表字”所出字的那张拓片。

3．“代表字”所代表的实际铭文用字的数量及出处在“备注”栏中列出。

六、关于古文字编码的几点建议：

1．新增字样覆盖情况的标注，主要是为了避免字符集相对实际文献材料用字产生缺漏，即在金文输入中发生无字调用的情况。覆盖情况，具体指编码字样与实际铭文用字的对应，可用注明字样在实际文献中的出现频次，及各频次的出处来完成该项属性的标注。每个字样都如此确定其准确的覆盖情况，就能保证字符集中所有的编码字样覆盖对应的实际文献材料的所有用字，从而实现字符集应有的功能。当然，完成覆盖情况标注，需要在穷尽一手材料进行逐字细致整理的基础上才能完成。

2．不以时空类型划分作为编码单位整理筛选的依据。字符集强调的是唯一性，在编码字样整理筛选中，必须坚持以形为本的一元筛选原则，如脱离构形特征，仅仅因为时代或地域差异在金文内部再划分诸多小类，独立确定编码字样，则形成了多元筛选原则，很可能导致字样的重复选定，影响字符集的数字化功能。具体来说，不应在金文编码中设立诸如“商金文”“两周金文”“战国楚系金文”“战国晋系金文”之类的划分原则，以避免字符集中进入一些无区别意义的字形。

3．坚持保真原则，所提交字样尽可能直接依据原始图片制作，而不应依据手写字形制作。原始图片质量实在有问题的再考虑用摹写方式加以完善。

4．未识字、歧释字也予编码。金文中没有争议的已识字只占少数，只给少数字编码而将大多数未识字、歧释字排除在编码范围之外，不合字符集的本来意义。

5．不限定每个字的编码字样数量。根据以形为本的一元筛选原则，每个字有多少字样

就选定多少字样。如果限定每个字的编码字样数量，可能导致编码字样缺漏。

华东师范大学中国文字研究与应用中心

2005 年 1 月

IRGN 1119Oracle Bone

Encoding Sample of Oracle Bone Inscription (甲骨文)

Serial No.	Rep. Script/ Glyph	Original Shape/ Glyph	Source	Period/ Epoch	Area/ Terrain	Material	Radical	Glyph Determ.	Corresp. Modern Char	Notes
1			14822	Shang Dynasty		Oracle Bone	一		元	
2			19790	Shang Dynasty		Oracle Bone	一		元	
3			2164	Shang Dynasty		Oracle Bone	八		介	
4			816 正	Shang Dynasty		Oracle Bone	八		介	
5			248 正	Shang Dynasty		Oracle Bone	人		人	
6			7291	Shang Dynasty		Oracle Bone	人		人	
7			35361	Shang Dynasty		Oracle Bone	人		人	
8			27909	Shang Dynasty		Oracle Bone	匕		匕	
9			27908	Shang Dynasty		Oracle Bone	匕		匕	
10			28598	Shang Dynasty		Oracle Bone	匕		匕	
11			831	Shang Dynasty		Oracle Bone	尸		尸	
12			830	Shang Dynasty		Oracle Bone	尸		尸	
13			33112	Shang Dynasty		Oracle Bone	尸		尸	

1. Data attributes:

1.1 Serial No.: used to identify individual script.

1.2 Rep.Script/Glyph: Provides a reference script for comparison based on emulating drawing from the original material. The script is source in Peking Normal University oracle bone inscription font.

1.3 Original Shape/Glyph: Original scripts on archaeological finds or their rubbings.

1.4 Source : Sequence number of rubbings in "The Oracle Bone Inscription Rubbings To Gather"(《甲骨文合集》郭沫若主编 中华书局 1982 年版)

1.5 Period/Epoch: The period the carrier of the script was used or created.

1.6 Area/Terrain: The place the carrier was used or created.

1.7 Material: The material of the carrier of the script.

1.8 Radical: The radical of glyph corresponding to 540 radicals in "Shuowen Jiezi."

1.9 Glyph determ. : Converts a script into a corresponding Kaishu glyph according to the rules of glyph determination.

1.10 Corresp. Modern Char.: Lists one or more possible corresponding modern character(s) of the script.

1.11 Notes: Related comment.

2. Rules of character selection

2.1 The principles of character selection are completeness and authenticity. Well-recognized characters are collected; uncertain characters are left out. The encoded characters should be able to map to all existing glyphs and to meet various needs, especially the need of expert's research.

2.2 The rules of selection

2.2.1 Glyphs of different construct are encoded separately.

2.2.2 Glyphs with significant component variance are encoded separately, though the basic construct of the glyph may be retained.

2.2.3 Glyphs with same construct and components but with great difference in component's position, direction and line curves are encoded separately.

3. Rules of collation

3.1 Glyphs found in Shuowen Jiezi will be sorted as they were in Shuowen Jiezi (540

radicals).

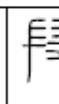
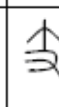
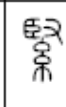
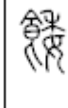
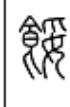
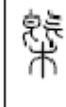
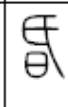
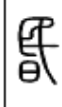
3.2 Glyphs not found in Shuowen Jiezi will be assigned a radical according to their composition for sorting in according to the Shuowen Jiezi (540 radicals).

Peking Normal University

2005-1-29

Encoding Sample for Xiaozhuan

Serial No.	Rep. Script/ Glyph	Original Shape/ Glyph	Source	Period/ Epoch	Area/ Terrain	Material	Radical	Glyph Determ.	Corresp. Modern Char	Notes
1			《说文》大徐本			纸本文献	丄	上	上	大徐本以古文为正篆
2			《说文》段注本			纸本文献	二	上	上	段注改古文正篆
3			《说文》大徐本			纸本文献	丄	上	上	篆文形体
4			《说文》大徐本			纸本文献	示	社	社	
5			《说文》大徐本			纸本文献	示	社	社	古文
6			《说文》段注本			纸本文献	示	社	社	段注改古文
7			《说文》大徐本			纸本文献	又	叟	叟	
8			《说文》大徐本			纸本文献	又	叟	叟	籀文
9			《说文》大徐本			纸本文献	又	俊	叟	或体
10			《说文》大徐本			纸本文献	又	叚	叚	
11			《说文》大徐本			纸本文献	又	叚	叚	古文

12			《说文》大徐本			纸本文献	又	叚	叚	《说文》引谭长说
13			《说文》大徐本			纸本文献	支	支	支	
14			《说文》大徐本			纸本文献	支	支	支	古文
15			《说文》大徐本			纸本文献	臤	緊	紧	
16			《说文》大徐本			纸本文献	食	餧	馁	
17			《说文》段注本			纸本文献	食	餒	馁	段注改篆
18			《说文》大徐本			纸本文献	木	槩	概	
19			《说文》唐写本			纸本文献	木	槩	概	依今存唐写本
20			《说文》大徐本			纸本文献	口	圖	图	
21			《说文》小徐本			纸本文献	口	圖	图	小徐本与大徐本有较大差异
22			《说文》大徐本			纸本文献	日	昏	昏	
23			《六书故》			纸本文献	日	昏	昏	《六书故》引唐写本

Notes on the Encoding Sample of Xiaozhuan

1. Source of character selection

1.1 The source of the Xiaozhuan encodings include all the materials we can get, which amounts to more than 20 books and a collection of nearly 20000 glyphs.

1.2 *Xiaozhuan Font* developed by Beijing Normal University and other materials concerned was used in the character decoration and selection.

2. Rules of character selection

2.1 The principles of character selection are completeness and authenticity. The encoded characters should be able to meet various needs.

2.2 The rules of selection

2.2.1 All the glyphs in *Shuo Wen(Daxu Edition)* will be selected. No matter the glyphs is correct or not.

2.2.2 We select the glyphs in *Shuo Wen(Xiaoxu Edition)* only when they are different from the glyphs in *Shuo Wen(Daxu Edition)* obviously.

2.2.3 The glyphs in *Shuo Wen(Writing Edition in Tang Dynasty)*, both the exiting edition and its materials in other books, will be selected if they are different from the glyphs in *Shuo Wen(Daxu Edition)* and *Shuo Wen(Xiaoxu Edition)* obviously.

2.2.4 Some glyphs in *Shuo Wen(Daxu Edition)* had been modified by *Duanzhu*. We only select the correct ones of them.

2.2.5 We also select the glyphs in other books use our discretion. Those glyphs can't been found in above-mentioned books.

3. Collation of Characters

3.1 The glyphs of different characters are sorted as its place in *Shuo Wen*.

3.2 The glyphs of the same character are sorted as this order: (1) the glyphs in *Shuo Wen(Daxu Edition)*; (2) the glyphs in other books.

3.3 The glyphs in *Shuo Wen(Daxu Edition)* are sorted as this order: (1) Zheng Zhuan; (2) Gu Wen; (3) Zhou Wen; (4) Huo Ti.

3.4 The glyphs in the other books are sorted as this order: (1) *Shuo Wen(Xiaoxu Edition)*; (2) *Duanzhu*; (3) *Shuo Wen(Writing Edition in Tang Dynasty)*; (4) the others.

4. Notes on the glyphs' table

4.1 "The representative script/glyph" is the result of decorating the corresponding original image with *Xiaozhuan Font System* developed by Beijing Normal University.

4.2 "The original shape/glyphs" is attained directly by copping the images in the books.

4.3 The special capacity of the glyphs are given in the row of "Notes'.

Beijing
Jan. 2005

小篆编码字符样例说明

一、选字的材料范围：

1．小篆编码字符集选样范围为迄今所能搜集到的所有小篆字形材料，计 20 余种文献，所得字样总数 2 万多个。

2．小篆字样的选择与加工利用了北京师范大学研制的《小篆字库》（北京出版社 2000 年出版）及相关资料库。

二、选字原则：

根据收字尽可能全面，能满足各种不同需求的宗旨，确立以下选字原则：

1.《说文》大徐本中的所有正篆、古文、籀文、或体，不论是否存在问题，均予收录。

2.《说文》小徐本中与大徐本有明显差异者，予以收录；如果只是细微差异，则只收大徐本字形。

3.对于今存唐写本以及其他文献所引唐写本中与大徐本不同的字形，均予收录。

4．段注所改大徐本字形，如符合汉字构形实际，可予收录；否则，不予收录。

5．对于其他文献中所保存的篆文字形，如系上述文献中所无，也酌情收录。

三、字形排序：

1．不同字种之间按《说文》中的各字的实际顺序排列。

2．同一字种的不同字样之间先列大徐本的各种字样，次列其他文献中的字样。

3．大徐本中的字样先列正篆，然后依次列古文、籀文、或体。

4．其他文献之间如果同时在某一字种下出现，其顺序为《说文》小徐本、段注本、唐写本、其他。

四、关于字样表的说明：

1.“代表字”是运用北京师范大学小篆字库处理系统对扫描图象进行加工获得。

2.“原形字”直接由所出自的文献扫描获得。

3.“备注”标明该字样的特殊身份。