A Survey on the Achievements of the Oracle Bone Digitization Projects and Prior Definitions of the Targets

ATSUSHI SUZUKI[†1]   suzuki toshiya[†2]

During 2003-2012, the Old Hanzi Expert Group under ISO/IEC JTC1/SC2/WG2/IRG was working to draft the Oracle Bone character encoding for ISO/IEC 10646. However, the group was disbanded without finishing the draft. It was supposed that the failure was caused by the irregular situation that the group had to work for the industrial and international standard. On the other hands, there were several projects to digitize Old Hanzi materials since the late 1990s, in China mainland, Hong Kong and Taiwan. As a result of the survey, it was found that most text digitization projects had worked with the transliterated text with few Oracle Bone characters. The fonts in their products are investigated and the adequacy for the standardization is discussed.

1. Introduction

During 2003[1]-2012[2], the Old Hanzi Expert Group under ISO/IEC JTC1/SC2/WG2/IRG was working to draft the Oracle Bone character encoding for ISO/IEC 10646[3]. The group was disbanded without finishing the draft. It was supposed that the failure was caused by the irregular situation that the scholastic group had to work for the industrial and international standard. The experts from China mainland and Taiwan were leading the group, but the experts were mainly invited for their Old Hanzi expertise, thus their expertise for the computer was not enough to settle the pragmatic milestones in the standard specification development. In addition, although XML-based data repository was designed in the early phase of the project, the easy client software for the Old Hanzi scholars was not implemented. The participants were forced to work with the hardcopies of the database, and could not check the consistency with machinery support; it caused serious latency of the collaboration. Because of such difficulties, the disbanding was supposed as because of the irregular situation of the project.

However, making a survey of the Oracle Bone digitization projects in China mainland, Hong Kong and Taiwan since the late of 1990s, it was found that the unclear definition of the purpose of the projects often made the projects evaporated without the publication of the result (finished or ongoing). Thus, it could be that the failure of the ISO Old Hanzi Expert Group was not irregular case, and similar failures are regularly repeated. In this report, the behavior of the Old Hanzi digitization projects and the adequacy of the fonts in their products are considered from the viewpoint of the information interchange and the standardization.

2. Classification of the Purposes of the Oracle Bone Digitization

The identification and decipherment of the Oracle Bone is still main topic in the Oracle Bone study, thus the information to be interchanged is dependent with the background of the users. The typical distinctive purposes would be "the interchange of the idea on the glyph shape" and "the interchange of the idea on the contextual meaning". The most fundamental attitude for the former purpose is the utilization of the sketch of the Oracle Bone glyph shape, although the running direction and the line breaking of the text are modified. It is called as "imitated text" in this paper. The common attitude for the latter purpose is the utilization of "modern" Hanzi to interchange the Oracle Bone text. It is called as "transliterated text" in this paper. There are several intermediate attitudes that the transliteration could be used to interchange the idea on the glyph shape, by using especially composed modern Hanzi. In such attitudes, the transliterated texts with the especially composed modern Hanzi and those without such are distinguished and different names are given. However, considering that it is rare to find the materials comparing 2 different texts by the same author, the objective distinction is not easy. In the case of ISO Old Hanzi Expert Group's database, originally "Li-style" and "Kai-style" distinctive items were allocated [4]. But in the late phase of the project, the Kai-style was replaced by "UCS character" and "Li-style" was abandoned. Thus, we give it up to discuss such distinction, and the simple distinction between "imitated text" and "transliterated text" is used in this paper. The transliterated text could include the especially designed modern Hanzi. And, sometimes the imitated glyph could appear in the "transliterated text" when the author could not design the modern Hanzi shape.
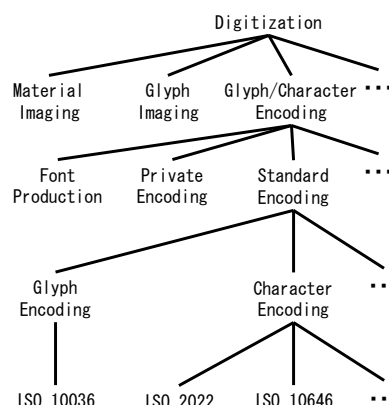
2.1 Purposes of the Oracle Bone Digitization

Figure 1: The structured categories of the possible digitization of the Oracle Bone and the international standards

Although the first action of the Oracle Bone digitization is applied to the ISO/IEC JTC1/SC2/WG2 for the inclusion in

†1 Ibaraki University.
†2 Hiroshima University
Science and Technology

ISO/IEC 10646, there are several international industrial standards that can help the advanced application of the computerization in Old Hanzi studies. As summarized in Figure 1, there are various categories.

For example, in Oracle Bone studies, the discussion based on the shape of the material, the glyph size (in comparison with the material size), the position of the text, and sometimes the position of the cracks are important (because Oracle Bone is used to decide a political decision by the cracks induced by the heat). Such information could not be preserved in the coded text. For such study, the imaging of the whole material is required. It would be the most intra-material purpose of the Oracle Bone material digitization.

For more inter-material purpose, like, the identification of the inscribers (it is assumed that the number of the Oracle Bone inscriber is not so large, and the identification of the inscriber is useful to identify the context of the oracle on the material), the whole imaging of the Oracle Bone material is not essential, but the per-glyph imaging would be required.

For further inter-material purpose, like, the situation that characters are repeatedly used and the loss of the source information is permitted, the abstraction of the character is required to reduce the amount of the information. For such purpose, the character encoding is appropriate solution.

Even if the character encoding is appropriate, the appropriate encoding scheme is dependent with the stability of the character set and the size of the community. Especially, ISO/IEC JTC1/SC2/WG2 avoids the incompatible update in ISO/IEC 10646 to guarantee the persistency of the data coded by ISO/IEC 10646, thus, the trial or experimental character set should not be added to ISO/IEC 10646.

The advantages and disadvantages of each encoding methods could be summarized as Table 1.

Table 1: The advantages and disadvantages of the encoding methods for the rarely used characters.

	Printing Stability	Search / Database	Co-usability	Development Cost
Existing Ideographs	Stable	Impossible	Hard	Low
Existing non-Ideographs	Instable	Impossible	Hard	Middle
PUA	Unstable	Possible	Possible	Middle
new standard codepoints	Stable	Exact	No Problem	High

2.2 Survey of the Oracle Bone Digitization Projects

Although ISO Old Hanzi Expert Group behaved as all stakeholders should participate, or should have already participated and their decision is a consensus of the all Old Hanzi scholars, the group had never made a survey of the existing projects. The first explicit mention of the existing digitization project was the comments by Adam Smith to WG2

[5], asking for the consideration of the compatibility with CHANT projects run by Hong Kong Chinese University [6].

In the development of ISO 10646, the compatibility to improve the cost of the migration from non-standard encoding system is one of the considerable points. The authors had made the survey of the existing project to digitize the Oracle Bone materials. Some of them had published their results fully or partially (most of them were published just after the start of ISO Old Hanzi Expert Group), and others seem to have faded away or evaporated without publishing the result. In the summary of the survey, it should be supposed that the prior definition of the goal of the project is important to achieve something. In this report, the survey is briefly summarized and the pragmatic goal definition is discussed.

2.3 Method of Survey

The authors made a survey by searching the electronic journal for Chinese papers; "China Knowledge Resource Integrated Database (CNKI)" [7] and Taiwanese papers; "Index To Taiwan Periodical Literature System (Perio)" [8], with basic keywords like "甲骨＋數位化", "甲骨＋数字化", "甲骨＋編碼" and "甲骨＋字庫". In most papers in Old Hanzi studies, the utilized database is rarely referred (e.g. the source of the database is referred; the database itself is bypassed), except of the newsletters or the meeting report on the database itself [9]. Therefore, the survey by the authors might have overlooked the databases that simply constructed and used but not promoted by the papers submitted to the journals. For example, it is mentioned that Beijing Normal University had constructed the fonts and input methods for Oracle Bone and Shuowen [10], there is no official website distributing their products. Such projects could not be investigated consistently.

3. Existing Projects of Oracle Bone Digitization

3.1 Overview

The projects found by the electronic journal databases are summarized in following. The methodology of the projects is classified into 3 categories;
A) the digitization to combine multiple databases published as printed matter,
B) the digitization of the single database published as printed matter,
C) the construction of new database from raw materials.
The achievement status of the projects is classified to 5;
(a) finished and kept updated (or extended)
(b) finished and frozen
(c) unfinished and ongoing
(d) unfinished and not updated.
(e) nothing published

As a brief overview of the projects, it is found that the method C) would be difficult for the projects to publish their products. It could be understood that the method C) might require the careful project management. In the methods A) and B), the expected user is already clear even if the product of the project is not finished. It is possible for them to evaluate the preliminary products of the projects. In addition, it is possible to

define the milestones during the long project, because the existing huge database might have a bias on the utilization in the contents; some contents would be referred quite often, but others are not so often referred. But in method C), it is difficult to evaluate the preliminary products, and difficult to define the milestones during the long project. Therefore, if the project is executed by human resources, it is important to specify the working process to keep the consistency of the data during the long project. However, it is not easy to specify the working process to construct something new from scratch.

Table 2: Oracle Bone digitization projects found in the electronic journal survey

Host	Name	Start of Project (for Oracle Bone)	Declared Goal (for Oracle Bone)	Used Resource	Work in Project	Product of Project				URL
						Overview	Font	Input Method	Key to Search DB	
University of Hong	CHinese ANcient Texts Database	1996〜	Combination of HJ, MSZJ and LZ	MSZJ, LZ	Text data of MSZJ	XZB, JSZJ, ZB (A-a)	Yes	No IM	Modern Hanzi, Old Hanzi Index Number	http://www.chant.org/ http://www.cuhk.edu.hk/ics/rccat/research1.html
Academia Sinica	Chinese Document Processing	2006〜	Font for Old Hanzi and their transliteration	LZ	Font for LZ indexing glyphs	2700 glyphs are digitized from 3700 indexing glyphs (B-c)	Yes	Yes	No Search	http://cdp.sinica.edu.tw/cdphanzi/
	甲骨文合集材料来源表資料庫	2012〜	Text search for LYB	LYB	Text data of LYB and web search system	Web Search System of LYB (B-b)	Not For Oracle Bone Character			http://xiaoxue.iis.sinica.edu.tw/obm/Home/
	小學堂甲骨文資料庫	2013〜	Text search of JGWB and glyph preview	JGWB	SVG data of JGWB glyph, source information and web search system	Web Search System of JGWB (B-a)	No Font	No IM	Modern Hanzi, Item Index	http://xiaoxue.iis.sinica.edu.tw/jiaguwen/
	Lexicon of Pre-Qin Oracle, Bronze Inscriptions and Bamboo Scripts	2012〜	Text search of transliterated text for pre-Qin materials	MSZJ	Text data of MSZJ	Web Search System of transliterated text in MSZJ (B-b)	No Font	No IM	Modern Hanzi	http://inscription.sinica.edu.tw/
	Digital Archiving of the Oracle Bone Rubbin_s	2004〜	Combination of the ink rubbings hold by Sinica and their transliteration	XYW, MSZJ	Scanned images of the ink rubbings and text data of XYW, MSZJ	Web Database providing the scanned image and transliterated text (A-d)	No Font	No IM	Modern Hanzi	http://ndweb.iis.sinica.edu.tw/rub_public/System/Bone/home2.htm
National Cheng Kung University	Oracle Bone Inscription Full Text Database	1995〜	Combination of HJ, MSZJ and LZ	MSZJ, LZ	Text data of MSZJ	Web Search System of transliterated text in MSZJ (B-b)	No Font	No IM	Modern Hanzi	http://ttssearch.lib.ncku.edu.tw/ttscgi/ttsweb1.exe
East China Normal University	Huayuanzhuang Dongdi Jiagu Search System	2008〜	Text Search for HD	HD	Text data of HD?	Font for HD database? (A-d)	Yes	No IM	Modern Hanzi	http://www.wenzi.cn/huadong/index.HTM
	Annotated Catalog of Ancient Character Research Search System	2009〜	Collection of existing studies of Old Hanzi	GKTZ	Web Search System	Web Search System of GKTZ (B-a/b)	Yes?	No IM	Modern Hanzi, Item Index, Structure Description	http://www.wenzi.cn/tiyao090104/Index.asp
Anyang Normal University	Oracle Bone Character Collection	2005〜	Input Method of Oracle Bone character, Imitated and Transliterated texts of Oracle Bone materials	?	?	? (C−e?)	?	?	?	http://yswh.aynu.edu.cn/jgwyjsite/jgwzdn.htm

3.2 Detail of Each Project

● CHANT in Hong Kong

The most well known project in China would be CHANT by Hong Kong Chinese University. The project was started as a combination [11] of "Jiaguwen Heji" (甲骨文合集, HJ) [12], "Yinxu Jiagu Keci Moshi Zongji" (殷墟甲骨刻辭摹釋總集, MSZJ) [13] and "Yinxu Jiagu Keci Leizuan" (殷墟甲骨刻辭類纂, LZ) [14]. During the combination of HJ and MSZJ, the scanned images of HJ were compared with the text in MSZJ, and some mistakes were found. The points suspected to be the mistakes in MSZJ were corrected and the proposed corrections were published as a (partial) result of the project; "Xinbian Jiaguwen Zixing Zongbiao" (新編甲骨文字形總表, XZB) [15]. In later, the Oracle Bone materials published after LZ (e.g. "Jiaguwen Heji Bubian" (甲骨文合集補編, HJB) [16], "Huayuanzhuang Dongdi Jiagu" (花園庄東地甲骨, HD) [17] were added to the database. As a result of the additional works, "Jiaguwen Jiaoshi Zongji" (甲骨文校釋總集, JSZJ) [18] and
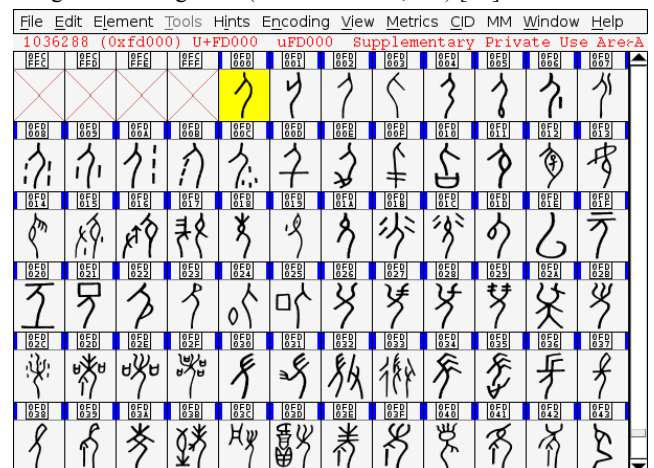
"Jiaguwen Zixingbiao" (甲骨文字形表, ZB) [19].

Figure 2: PUA Glyphs in CHANT combined font.

CHANT published several TrueType fonts (ics1.ttf 〜 ics7.ttf) and there is an Oracle Bone font, named ics3.ttf. The latest version is based on ZB, and the characters are coded at

PUA codepoints (U+E000～) in BMP. Although it is reasonable design, all ICS fonts use BMP PUA area and conflict with each other. The user could not use the multiple ICS fonts easily. To reduce the difficulty, CHANT also provides a unified font (chant.ttf) collecting almost all glyphs required for their database (modern Hanzi for the transliteration, Oracle Bone, Bronze Inscriptions, Bamboo scripts, Shuowen Jiezi, etc). It is remarkable feature that all non-standardized characters are coded in surrogate PUA planes; U+F0000 ～ U+10FFFF. However, it means that the text coded with ics3.ttf could not be used by the unified font.
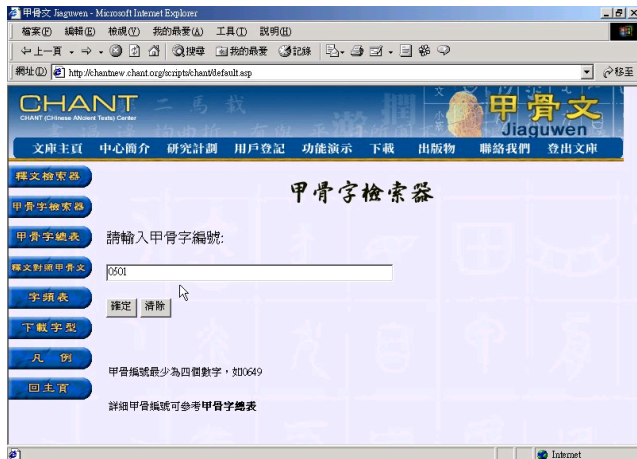
Figure 3: Interface to Search the Materials by Oracle Bone, in CHANT

Figure 4: Transliterated Texts Comparison in CHANT

However, no input method for their Oracle Bone font is provided. Thus, to search the Oracle Bone texts without the transliteration, the Oracle Bone character indexing number is required, and the "phrase" of Oracle Bone characters could not be used (Figure 3, Figure 4). On the other hand, a Cangjie-based input method is provided for special Hanzi in the transliteration is provided [20]. Considering that the Oracle Bone character indexing numbers in XZB and ZB are incompatible, and the structure of JSZJ was similar to "Jiaguwen Heji Shiwen" (甲骨文合集釋文, HJSW) [21] than to MSZJ; it does not include the Oracle Bone texts, still the Oracle Bone character set in CHANT has a room of the further stabilization and improvement as a

platform to interchange the Oracle Bone text information.

● Various Projects in SINICA, Taiwan

Academia SINICA in Taiwan has many laboratories; some of them are working for the digitization of the Oracle Bone materials. Some projects are led by the literal, historical or archaeological experts, and others are led by the information technology experts. Also there are collaborations of these two groups [9].

The most long running project would be "Chinese Document Processing" project (CDP). It started as a project to automate the production of the font for the rarely used modern Hanzi in Hanyu Da Zidian. The "Character Description Language" (CDL) by Wenlin Institute is a diversion of CDP, and it had some impact to Unicode. In later, the project was asked to support the production of the modern Hanzi used for the transliteration of Old Hanzi materials. Now the target of CDP includes various Old Hanzi scripts (JiaGuWen, JinWen, ZhouWen, etc). For Old Hanzi, a part of LZ referential glyphs are digitized (LZ has 3600 indexing characters, and CDP digitized 2700 glyphs from them), and the input methods are also developed. However, it seems that the coverage of the digitized LZ glyphs would be defined by the request to digitize MSZJ, not LZ itself. It seems that CDP had ever collaborated with Beijing Normal University to produce Old Hanzi fonts, but detail is not reported[a].

Another remarkable project would be "Digital Archiving of the Oracle Bone Rubbings" (甲骨文全文影像資料庫, DAOBR) and "Lexicon of Pre-Qin Oracle, Bronze Inscriptions and Bamboo Scripts" (先秦甲骨金文簡牘詞彙資料庫, LPQ). The DAOBR was a project to digitize the ink rubbings of Oracle Bone and combine them with the transliterated texts. The remarkable point was that SINICA holds the copyrights of the important referential materials ("Xiaotun Yinxu Wenzi" (小屯殷虛文字)) [22][23][24] published in Republic of China, and the original inked rubbings used by the older referential materials (like "TieYun CangGui", (鉄雲蔵亀))[25], so SINICA is the most appropriate organization to disclose the combined database of the ink rubbing images and the transliterated texts[b]. However, the project seemed to be stalled after the combination of the images in XYW and the texts taken from XYW[c] and MSZJ. In LPQ, the transliterated texts in MSZJ only are digitized, without the inked rubbing images. DAOBR seemed to have an intention to compare the transliterated texts among the different scholars, although the digitization was not finished. In LPQ, the scope is limited to the digitization of MSZJ and no intention to compare or correct the mistakes in MSZJ. In the SINICA's design of the metadata for Old Hanzi materials, only the transliterated texts are considered. Although 2 kinds of the transliterations are

---

a The font and Input Method package by CDP includes the fonts titled bnucdp.ttf, 北師大說文小篆, which the last update was on 2012, but no background is documented in the package.
b The simple combination of ink rubbing collection (e.g. HJ) and the transliteration texts (e.g. HJSW) would be the most primal database expected by the scholars. However, it is impossible to publish such database without the permission of the copyright holders, because such combinations are not "quotation".
c It seems that Japanese transliteration material for XYW, 小屯殷虛文字乙編考釋 by 甲骨学会, was used, but further investigation is needed to identify.

considered (transcription and punctuated transcription), it is not expected to fill the imitated texts are stored in the database.

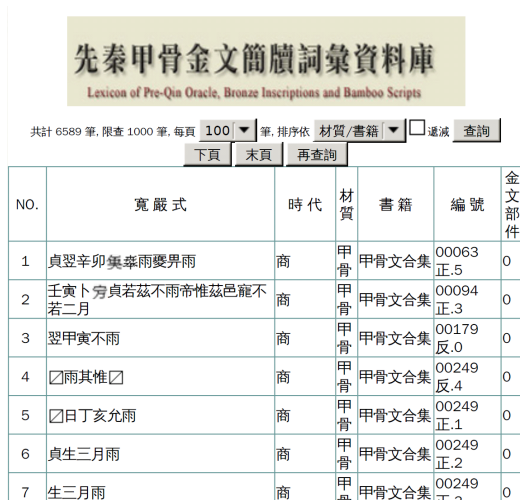Figure 5: Transliterated Texts Comparison in DAOBR

Figure 6: MSZJ Texts in LPQ

Other notable projects from SINICA are the simple digitization of "Jiaguwenbian" (甲骨文編, JGWB) [26] and "Jiaguwen Heji Cailiao Laiyuanbiao" (甲骨文合集材料来源表, LYB) [27]. HJ provides the inked rubbing reprints only without source information, thus the scholars had serious difficulty to identify the material and search the existing research on the material before the publishing of HJ. LYB was published about 10 years later of HJ, to solve the problem. However, the mapping table in LYB was too large to use as handbook, and difficult to check "the material excluded from HJ". The web based interface makes LYB quite useful.

The projects in SINICA are mainly working for the transliterated texts and it seems that direct text encoding of Oracle Bone script is not worked yet. In comparison with CHANT font, the implementation of CDP fonts is fragile. The encoding of the fonts is declared as Big5. The Oracle Bone glyphs are basically coded at the corresponding modern Hanzi.

If no corresponding modern Hanzi is found in Big5, PUA codedpoint of Big5 is used. Because the PUA codepoints in Big5 could not be mapped to UCS, and CDP fonts lack UCS cmap tables, it has a serious difficulty to interchange with Unicode centric software.

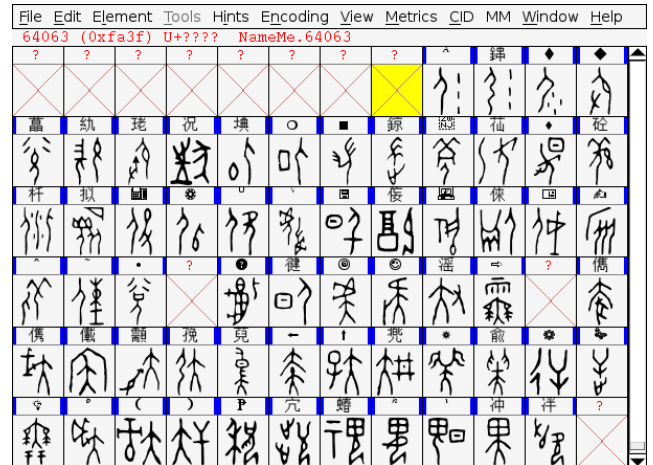Figure 7: CJK Unified Ideograph Area in CDP font

Figure 8: Big5 PUA Area in CDP font

● Other Achieved Projects

Figure 9: MSZJ Texts in NCKU's Oracle Bone DB

As described in above, the digitization of the existing printed matter database was the typical and pragmatic

milestones for the Oracle Bone digitization. Although the projects are not active anymore, there were other projects digitizing existing printed matter databases.

National Cheng Kung University in Taiwan had a project "甲骨文全文影像資料庫" (Figure 9) which is similar to CHANT, planned to combine HJ, MSZJ and LZ. Although it was stopped after the whole digitization of the transliterated text in MSZJ, it seems that the web based search system is disclosed to the internal of the university.

The Center for the Study and Application of Chinese Characters (CSACC) of East China Normal University in China mainland (ECNU) is known by their expertise on the Bronze Inscription studies [28], but they constructed a few Oracle Bone databases; the earlier was a database for HD around 2008. HD was the one of the post-HJ collection, so the HD digitization was a distinctive project ("Huayuanzhuang Dongdi Jiagu Keci Leizuan" (花園庄東地甲骨刻辭類纂) [29], the analysis of HD with LZ methodology was published in 2011).
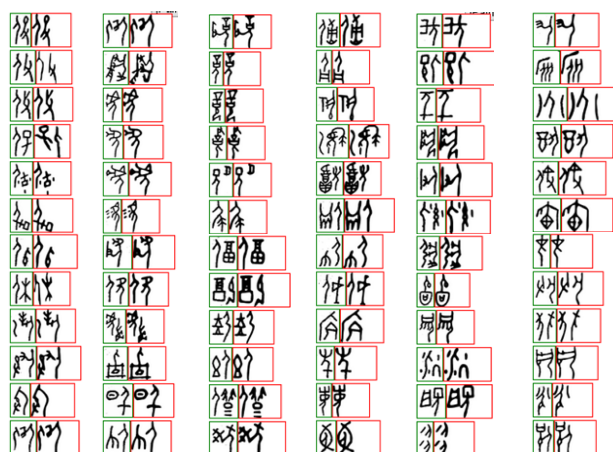

Figure 10: Glyph Comparison between LZ (Green) and HD font by CSACC (Red)
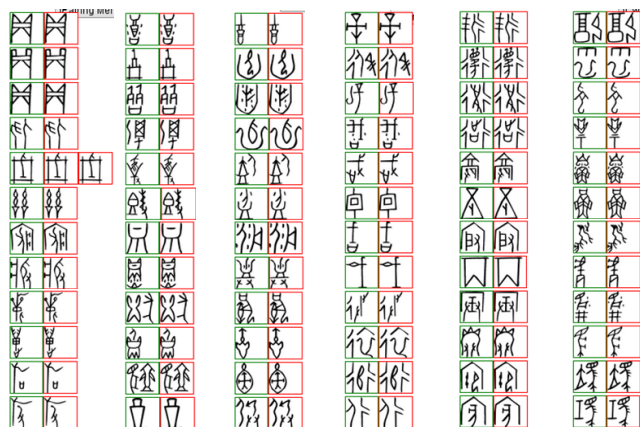

Figure 11: Glyph Comparison between CHANT font (Green) and HD font by CSACC (Red)

Considering that they provided the font to display the database but without the input methods, it is expected that the transliterated text is the core of the database. The font designed for HD database was a mixture of the modern Hanzi for the transliteration, Oracle Bone glyph machinery converted from the

inked rubbings, Oracle Bone glyph made by the scanned images of LZ (Figure 10)[d], and the duplication of CHANT font (Figure 11).

The coverage of the glyphs from LZ would be almost complete, but that from CHANT is partial. It seems that the motivation of the font architecture would be the analysis of the HD glyphs; the glyphs known already in LZ, the glyphs known in XZB, and the glyphs unknown before. It might be related with the study by Chen Tingzhu [30] collecting the "new" glyphs in the materials published after XZB. For such purpose, even if the qualities of the glyphs are varied because of the multiple source fonts, the design of the font is quite reasonable. However, the font encoded the Oracle Bone glyphs at existing CJK Unified Ideograph codepoints, in BMP (Figure 12), without consideration about the meaning correspondence. The font is useful for the purpose to find the first collection (LZ, XZB, ...) including a specified glyph, but inappropriate to use for the information interchange.
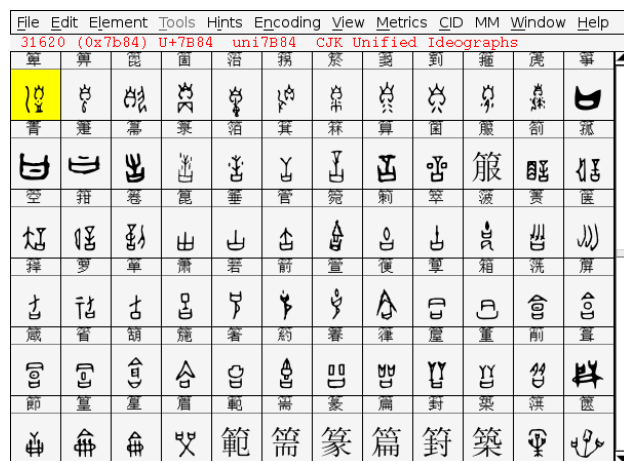

Figure 12: CJK Unified Ideograph Area in HD font by CSACC.

Another database by CSACC was the digitized index of "Guwenzi Kaoshi Tiyao Zonglan" (古文字考釋提要總覽, GKTZ) [31]; which is a compilation of the existing studies on Old Hanzi (for all pre-Qin materials). The compilation was for per-character level in Shuowen order, thus, the fonts for the digitized index is not for the digitization of raw Oracle Bone text. It is questionable if their product is suitable for direct encoding of Oracle Bone script, but the project was successfully achieved.

● Other Unfinished or Stopped Projects

In the survey of electronic journal, the research groups from Anyang Normal University (ANU) publish many papers for the Oracle Bone digitization. However, the various technologies to help the digitization (the font synthesis from the common graphical parts, the conversion from the scanned images to the outline fonts, the input methods, the character

---

[d] "Shuowen Jiezi Quanwen Jiansuo (説文解字全文検索)" by CSACC on 2004 [36] had included 3 TrueType fonts; SURSONG.TTF (a font including Ext B characters for GB18030 support), SW.TTF (Shuowen glyphs) and jiaguwen.ttf includes LZ glyphs at the CJK Unified Ideograph codepoints.

recognition, etc) are discussed, but the products like the font, the input method, the text or the database are not published yet. Furthermore, they have not stabilized the character set for the Oracle Bone or the encoding architecture. In the website supposed to be since around 2005, the input method and the mapping table between Oracle Bone and modern Hanzi are based on ZB. But in the recent papers for the font synthesis, the character set and ordering is supposed to be based on Shuowen Jiezi [33]. Although the quality of the font by CHANT is not sufficient for the authenticated printing, the quality of the sample in their paper is more primitive than CHANT.
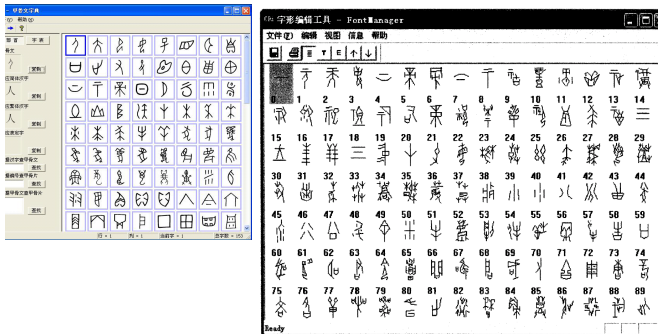

Figure 13: "Proof of Idea" fonts by ANU on website (left) [32] and paper (right) [33]

4. Conclusion

Among the fonts found in our survey, the fonts by CHANT and CDP would be remarkable results because they are based on existing database, LZ. However, there are rooms to be improved for better stability and coverage. It should be noted that no attempt to construct an Oracle Bone glyph collection from scratch was achieved within several years.

4.1 Definitions of "Digitization of Oracle Bone"

In this survey, most finished projects had achieved the digitization of the transliterated texts in MSZJ. The imitated texts in MSZJ (or the content in LZ) are not dealt, except of CHANT. There might be a question; if the transliterated text is the original target of the project, why HJSW, the transliteration by the HJ editors (thus more authorized than MSZJ), was not chosen? The case of NCKU would be the suggestive evidence. Considering that CHANT and NCKU had declared the goal of the project in 1990s, it was a combined database of HJ, MSZJ and LZ, the original ink rubbing (HJ), imitated glyphs (MSZJ imitation), transliterated texts (MSZJ transliteration), and the search system (LZ) is expected. But when the project is started with finite time period, the pragmatic goal was only found in the digitization of the transliterated text. But the projects started since 2000s, like SINICA DAOBR and LPQ are no longer declaring as their goals include the imitation texts, although SINICA CDP had supported these projects by the production of the fonts and the input methods. It seems that the difficulty to digitize the imitation texts are already is high. Also it coincides with the situation that the publications from CHANT, the transliterated texts are published, and the glyph collections are updated, but the collection of the imitated texts and the database

like LZ are not published. The exceptional work would be Chen Nianfu's "Jiaguwen Moshi Quanbian" (甲骨文摹 釋全編, QB) [34], the collection of the imitated texts with the fonts based on LZ indexing glyphs. However, QB is still in the phase of MSZJ. The LZ-like database by QB is expected.

4.2 Future Issues

As written in above, the expected advance by the utilization of the computer in Old Hanzi study would be the reduction of the cost to search, index, and compare the paper materials. It is found that the digitization of single referential material is the stable definitions of the goal for Oracle Bone digitization projects. But the digitization of the existing imitated texts is still too expensive for most projects, because the typewriting of the existing imitated texts in the printed matter (like MSZJ) requires highly skilled workers. The per-glyph rearrangement of the imitated texts, like LZ, would be useful to reduce the cost of the digitization. Also it would be useful to consider the stable glyph set of the Oracle Bone for general purpose.

Although there are the resources based on digital technology; like CHANT and QB, their imitation texts are difficult to validate, because they are arranged by the source material, not by the referential glyphs. As a result, to discuss the stable glyph collection for Oracle Bone, we should start from LZ or its ancestor; "Inkyo Bokuji Sourui" (殷墟卜辭綜類)[35].

Reference

1) Li Guoying (李國英), Tom Bishop: "Draft Agreement on Old Hanzi Encoding", ISO/IEC JTC1/SC2/WG2/IRG N1014 (2003).
2) ISO/IEC JTC1/SC2/IRG: "Resolutions of IRG Meeting #38", ISO/IEC JTC1/SC2/IRG N1870 (2012).
3) ISO/IEC 10646: Information technology -- Universal Coded Character Set (UCS), International Standard Organization, Switzerland (2012).
4) ISO/IEC JTC1/SC2/IRG: "Report from the Old Hanzi Expert Group", ISO/IEC JTC1/SC2/IRG N1215 (2006).
5) Adam Smith: "Comments on the work of the Old Hanzi Group towards an encoding of OBI script", ISO/IEC JTC1/SC2/WG2 N4236 (2011).
6) CHinese ANcient Texts, http://www.chant.org/
7) Chinese Knowledge Resource Integrated Database, http://www.cnki.net/
8) Index To Taiwan Periodical Literature System, http://readopac.ncl.edu.tw/nclJournal/
9) Pham Lee-moi (范麗梅), Zhang Zaixing (張再興), Hung Imei (洪一梅), Ho Chewah (何志華), Cheng Shaoxuan (程少軒), Lu Jialiang (魯家亮), Li Jing (李靜), Shan Yuchen (單育辰) : "A Special Issue on "Newly Excavated Texts in the Digital Age (「數位時代的出土文獻」專輯)", Newsletter of the Institute of Chinese Literature and Philosophy (中國文哲研究通訊), Vol. 21, No.2 (2011) p. 1-113.
10) Lo Feng-ju (羅鳳珠), Zhou Xiaowen (周曉文): "An Important Milestone in the Digitalization of Ancient Chinese Documents-The Establishment of a Seal Character Font Database by Beijing Normal University (古籍數位化的重要里程--大陸北京師範大學完成小篆字型字庫的建立)", Newsletter for Research in Chinese Studies (漢學研究通訊), Vol. 17(3), No. 67, (1998), p.302-304.
11) Choi Sai Bun (蔡世彬): "甲骨文全文資料庫--現代科技与古老文字的結合", Chinese University of Hong Kong Newsletter (中大通訊), Vol. 12 (10), No. 185, p.4 (2001-6-4)
12) Guo Moruo (郭沫若): "Jiaguwen Heji (甲骨文合集)", Zhonghua Shuju (中華書局), Beijing (1978-1982) ISBN 7101016537
13) Yao Xiaosui (姚孝遂): "Yinxu Jiagu Keci Moshi Zongji (殷墟甲

骨刻辭摹 釋總集)”, Zhonghua Shuju (中華書局), Beijing (1988) ISBN 9787101003451

14)　Yao Xiaosui (姚孝遂): “Yinxu Jiagu Keci Leizuan (殷墟甲骨刻辭類纂)”, Zhonghua Shuju, Beijing (1989) ISBN 9787101004779

15)　Shen Jianhua, Cao Jinyan (沈建華，曹錦炎): “Xinbian Jiaguwen Zixing Zongbiao (新編甲骨文字形總表)”, Zhongwen Daxue Chubanshe (中文大學出版社), Hong Kong (2001) ISBN 9789629960476

16)　Zhongguo Shehuikexueyuan Lishi Yanjiusu (中國社会科学院歴史研究所): “Jiaguwen Heji Bubian (甲骨文合集補編)”, Yuwen Chubanshe (語文出版社), Beijing (1999) ISBN 9787801264961

17)　Zhongguo Shehuikexueyuan Lishi Yanjiusu (中國社会科学院歴史研究所): “Yinxu Huayuanzhuang Dongdi Jiagu (殷墟花園庄東地甲骨)”, Yunnan Renmin Chubanshe (雲南人民出版社), Kunming (2003) ISBN 9787222038776

18)　Cao Jinyan, Shen Jianhua (曹錦炎，沈建華): “Jiaguwen Jiaoshi Zongji (甲骨文校釋總集)”, Shanghai Cishu Chubanshe (上海辭書出版社), Shanghai (2006) ISBN 9787532621293

19)　Shen Jianhua, Cao Jinyan (沈建華，曹錦炎): “Jiaguwen Zixing Biao (甲骨文字形表)”, Shanghai Cishu Chubanshe (上海辭書出版社), Shanghai (2008) ISBN 9787532624317

20)　Che Wah Ho: “CHANT (CHinese ANcient Texts): a comprehensive database of all ancient Chinese texts up to 600 AD”, Journal of Digital Information, Vol. 3, No.2 (2002).

21)　Hu Houxuan (胡厚宣): “Jiaguwen Heji Shiwen (甲骨文合集釋文)”, Zhongguo Shehui Kexue Chubanshe (中國社会科学出版社), Beijing (1999) ISBN 9787500425212

22)　Dong Zhuobin (董作賓): “Hsao-Tun, Vol. 2, Inscriptions, Part 1-2 (小屯，第二本，殷虚文字，甲編・乙編)”, Academia Sinica (中央研究院), Taipei (1948)

23)　Jung Bor-sheng (鐘柏生): “Hsao-Tun, Vol. 2, Inscriptions Part 2, Supplement (小屯，第二本，殷虚文字，乙編補遺)”, Academia Sinica (中央研究院), Taipei (1995)

24)　Chang Ping-chuan (張秉權): “Hsao-Tun, Vol.2, Inscriptions Part 3 (小屯，第二本，殷虚文字，丙編)”, Academia Sinica (中央研究院), Taipei (1957-1972)

25)　Liu E (劉鶚): “400 Pieces From Tieh-yun Tsuang-kuei (鐵雲藏龜四百種)”, Academia Sinica.

26)　Haibo Sun (孫海波), Zhongguo Shehuikexueyuan Kaoguyanjiusuo (中國社会科学院考古研究所): “Jiaguwenbian (甲骨文編)”, Zhonghua Shuju (中華書局), Beijing (1965) ISBN 7101005233

27)　Hu Houxuan (胡厚宣): “Jiaguwen Heji Cailiao Laiyuanbiao (甲骨文合集材料来源表)”, Zhonghua Shehui Kexue Chubanshe (中國社会科学出版社), Beijing (1999) ISBN 7500425228

28)　Zhang Zaixing (張再興): “The Intermational Symposium On Computer-processing of Ancient Chinese Characters (古文字信息化処理國際学術研討会)”, Journal of Chinese Information Processing (中文信息学報), Vol.16, No. 3 (2002) p.63-64

29)　Qi Hangfu (齊航福), Zhang Xiuxia (章秀霞): “Yinxu Huayuanzhuang Dongdi Jiagu Keci Leizuan (殷墟花園庄東地甲骨刻辭類纂)”, Xianzhuang Shuju (線裝書局), Beijing (2011) ISBN 9787512002982

30)　Chen Tingzhu (陳婷珠): “Yinshang Jiaguwen Zixing Xitong Zaiyanjiu (殷商甲骨文字形系統再研究)”, Shanghai Renmin Chubanshe (上海人民出版社), Shanghai (2010) ISBN 7208083266

31)　Liu Zhiji (劉志基): “Guwenzi Kaoshi Tiyao Zonglan Vol. 1-3 (古文字考釋提要總覧)”, Shangai Renmin Chubanshe (上海人民出版社), Shanghai (2008) ISBN 720807982X, 7208092117, 7208104700

32)　Anyang Normal University (安陽師範学院): “Jiaguwendian (甲骨文字典)”, http://yswh.aynu.edu.cn/jgwyjsite/jgwzdn.htm browsed on 2013-05-01.

33)　Li Qingsheng (栗青生),Wu Qinxia (吳琴霞), Wang Lei (王蕾): “An Input Method of Oracle-bone-script Based on Dynamic Description of Oracle Strokes (基于甲骨文字形動態描述庫的甲骨文輸入方法)”,
Journal of Chinese Information Processing (中文信息学報), Vol. 26, No. 4 (2012) p.28-33

34)　Chen Nianfu (陳年福): “Yingxu Jiaguwen Moshi Quanbian (殷墟甲骨文摹釋全編)”, Xianzhuang Shuju (線裝書局), Beijing (2010) ISBN 7512002920

35)　Shima Kunio (島邦男): “Inkyo Bokuji Sourui (殷墟卜辭綜類)”, extended edition, Kyuko Shoin (汲古書院), Tokyo (1971)

36)　Huadong Shifan Daxue Zhongguo Wenzi Yanjiu Yu Yingyong Zhongxin (華東師範大学中國文字研究與應用中心): “Shuowen Jiezi Quanwen Jiansuo (説文解字全文檢索)”, Nanfang Ribao Chubanshe (南方日報出版社), Guangzhou (2004) ISBN 7806523278.