

Part C Project Details

1. Goals and Objectives

The project aims at constructing a Chinese character database that provides rich linguistic content and useful functions in support of quality education of the Chinese language and script for the future generation. With major features such as the listing of archaic script forms (*inter alia*, oracles bone scripts, bronze scripts, bamboo scripts, silk scripts, and seal scripts), cognitive analysis of semantic and phonetic script components, explanation of related etymologies, word-formations, English-Chinese lookup etc., the project aims at helping teachers and students on different levels, either through guided instruction or through self-learning, to acquire a deeper understanding of Chinese language and script. By integrating modern information technology with classical Chinese scholarship, the project can be compared to the provision of an infrastructure in basic Chinese language education which not only helps improve the overall standard of Chinese language education but also renders Chinese more competitive in a globalized world full of new challenges.

2. Needs Assessment and Applicant Organization's Capability

- (a) Background: The project is in the first place a self-initiated project building upon the good practices of two projects previously funded by QEF, namely,
- i) Education towards Biliteracy & Trilingualism (QEF 1998/5073)
 - ii) Internet Resource Project for Chinese Language Education (QEF 2003/0756)

In view of our peculiar and complicated language environment, Hong Kong students' overall standard in Chinese language has always been questioned. After the repatriation of Hong Kong in 1997, wide-ranging debates over language policy issues have raged. Recently, such debates have developed to such a disquieting extent that even the value of native language teaching was much doubted among educators in the school sector. On the other front, with the global dominance of English, there are signs in HK's tertiary education sector for Chinese to be further marginalized so that the status of Chinese as an academic language is further endangered. Consequently, the need to reinforce basic Chinese language education, while sustaining Hong Kong students' traditional edge in English, becomes a matter of grave concern. With this in mind, we plan to implement a quality web database in support of the education of Chinese language and script. With the use of IT we aim at making traditional scripts and related etymological explanations easily accessible and effectively navigable. With this need in view, this project fits into the special initiative of "*Effective Learning and Teaching of Languages*" as designated by the QEF.

(b) Applicant Organization's track record and capabilities

This application is initiated by the Research Centre for Humanities Computing (RCHC), founded since 1994 at the CUHK by the PI. Over the past decade, Besides having a notable funding record¹, the RCHC team has developed a number of web pages, many of which are related to language teaching and learning, and are proven useful through

¹ For a detail account of the funding record of the project leader in his capacity as head of the applying unit, see: <http://humanum.arts.cuhk.edu.hk/~twkwan/kwan-grants-f.htm>

the tens of millions of accesses recorded (see section 12). More importantly, the RCHC team has accumulated valuable working experience with hard and software, has assembled a small but capable crew, and has set up an infrastructure which can be applied to yet another round of ambitious project to serve the future need of Hong Kong school education.

(c) Team members' track record and roles to play in project

- i) Professor and Chairman designate of Philosophy Department, CUHK, founding director of RCHC, and appointed member (*ad personam*) of the CUHK IT Policy Committee since its establishment -- although trained in philosophy, Professor shows a sustained interest in language studies. He has been invited to speak on language related topics in Zurich (ETH Switzerland), Dublin (UCD), Leuven (KUL), Berlin (FU), Seoul (SNU), Honolulu (U Hawaii), mainland China (Tongji U, Nanjing U, Fudan U), Taiwan (NCU, NCCU), and Hong Kong (CUHK). Over the past years he took part actively in discussions on and debates over language education and language policies at both the tertiary and secondary levels. Of the many research projects he undertook as PI, many were related to the support of Chinese language, including two projects funded by the QEF. In this application he will again be responsible for the integration of the theoretical, linguistic as well as computational aspects of the project.
- ii) Two experts in classical Chinese scholarship have a crucial role to play in safeguarding the quality of the project. Chairman of the Department of Chinese Language and Literature was a member of the Moderation Committee for Chinese of the Hong Kong Examinations and Assessment Authority. In his concurrent capacity as Director of the Research Centre for CHinese ANcient Texts (CHANT) and as leader of the celebrated CHANT projects, he will give advice to the PI on all issues related to classical Chinese textual and lexical studies. It must be noted that CHANT has in fact provided the RCHC the right to use graphical images of many archaic Chinese scripts. The CHANT team has been working very smoothly with the RCHC team, and this precious working relationship will be strengthened so that further collaborative efforts of the two teams will be explored.
- iii) An expert in Chinese etymology and semasiology -- well versed in Wang Niansun's (王念孫, 1744 -1832 CE) *Guangya Shuzheng* (廣雅疏證) which marked the pinnacle of classical philological- phonological scholarships of the Qing Dynasty. This expert was recipient of the VC Exemplary Teaching Award (2006) and is much engaged in a host of radio/TV programmes related to Chinese language education, including 最緊要正字, 懶音逐個捉. As a new blood to our team, his involvement will be an additional guarantee for the sustainability of the project over the foreseeable future.
- iv) A dedicated scholar in Chinese language and Chinese linguistics -- being a member of HKIEd and the Editor of *Newsletter on Chinese Language*, this professor is in a favourable position to establish relations with the school sector. Given the nature of our project, interface with and impact on the educational sector (including the school sector) is a matter of special importance. In his daily interaction with school teachers, He can play the role of a spokes-man for the project as well as a conveyor of outsider opinions.
- v) A professor with strong IT backgrounds -- after earning a basic degree in information engineering, he completed his *Dr. phil.* degree in philosophy at the University of Heidelberg, Germany, with double minors in Computer Engineering and Information Systems. With his combined competence in the humanities and in IT, he has





contributed substantially to the development of projects of the RCHC. In the current project, he will be the PI's immediate backup in regard to computation and information processing.

- vi) Professor and Chairman of Department of Computer Science and Engineering, CUHK -- while being research active in his own field and the head of numerous computation and IT related research projects, this Professor has shown a great interest for humanities scholarship including philosophy and linguistics, and is for this reason a long-time acquaintance of the PI and a supporter of his humanities computational programmes.

3. Targets and Expected Number of Beneficiaries

Besides quantitative proof of the popularity of our work, it was reported that some existing databases of our Centre are deemed indispensable by many Chinese language teachers in their teaching and research. Since our project aims at providing a much upgraded version of our existing work, which is much richer in content and is equipped with a lot more features, we expect its impact on Chinese education to surpass that of its predecessors. Although this applies to Chinese education of all levels, there is no doubt that the school sector is the main beneficiary.

Direct beneficiaries: The ultimate beneficiaries will be the school teachers and students. The output database of the current project will be an invaluable reference for Chinese learning and teaching at different levels:

- a. The Chinese character database consists of various modules, including a genealogy of the characters from the archaic script forms to the modern script forms, the pronunciation of the characters, linkage to lexical entries, word formations, English look up, and most important of all, etymological accounts of the most important characters as well as a system of cross references from less popular characters to the main entries wherever applicable. Chinese language teachers of all levels can use the database as a quick reference to enrich their teaching materials.
- b. The Chinese pictographs including oracle bone scripts and bronze scripts are of great didactical values for a wide range of users. Besides being a channel for in-depth understanding of classical Chinese scholarship that is relevant to the upper forms, they can also arouse the interest of primary school children, as small children are generally more attracted to graphics (pictures) than to pure text. Since for each head character, the database will, as far as the situation allows it, keep multiple representative archaic/ancient script tokens, many of which are lively depictions of the respective objects or states of affairs related to the character. The use of these script tokens will effectively facilitate young children to memorize the gestalt of the characters while gaining a clear semantic understanding of them. The multi-function Chinese character database will embody approximately 2000 bronze scripts and 800 oracle bone scripts, which covers most of the characters in common use. For example, the ancient character for mountain(山), oracle bone script , bronze script ) is exactly a picture outlining the shape of a range of mountains with three apices. Such pictograph will help impress the character in junior student's memory. On the other hand, taking the character beginning(初) as an example, it is very common even for senior students to mistake the 衤 component for 礻. By referring to the oracle bone script  and bronze script , users will get to know that the original meaning of the character is to tailor a cloth in order to make clothes. Then the students will easily

recognize 衤 as the correct component for 初, given the fact that 衤 is the *lishu*-transformation of 衣, whereas 示 that of 示.

- c. For upper form students who are accustomed to web-surfing, the useful resources of the database will be a good starting point for independent learning. This applies especially to the more ambitious students. To facilitate them, we are planning to split the etymological accounts as much as possible into two sections, namely a simple account and a detailed account, in such a way that the self-learners can at least start with the simple accounts which will be aimed at the arousal of interest for getting into a deeper understanding by reading the detailed accounts.
- d. Given the “cognitive approach” of the database, as we will show later, this database, besides being beneficial to language teaching and learning, will also benefit the teaching and learning of other subjects such as Chinese history, Chinese culture, or Liberal Studies in the Hong Kong context. For instance, the character for piety (孝) is written as 𠄎 in bronze script which depicts a young child supporting an elderly person in walking, as such kind of action is deemed filial. Explaining this will help students understand the original connotation of the character as well as the culture in ancient times. This kind of content falls within the ambit of liberal studies. All this can be regarded as a “spin-off” from the basically language-related database. This additional benefit of the database is very much enhanced since we will have a full text search function that enable the users to search for key words that are of interests. The character adult man(夫) is written as 𠄎 in oracle bone script and 𠄎 in bronze script. The character consists of 大 and 一, the former meaning “man”, and the latter representing a hair clasp. The character reveals an ancient custom that a man at the age of twenty will have his hair held together showing that he has reached adulthood. As a man is allowed to get married only after he has reached manhood, 夫 also means husband. In ancient times, as an adult man would also have to render service as a labourer, so 夫 can be generally be used to denote any type of labourer, e.g. fisherman (漁夫) and farmer (農夫). The character wife(妻) (written as 妻 in oracle bone script) looks like a (man’s) hand getting hold of a woman’s hair. This is a vivid description of the practice of marriage by capture in ancient times; a man would go to another tribe and abduct a woman as his wife. For details, see section 5B.
- e. Lastly, we even think that the pictographic form of writing available in the database could assist students suffering from certain learning difficulties, e.g., dyslexia and dysgraphia. Given limited resources, we can not embark upon tailor-making a special front-end for this group of users at this stage, but we do think that the teachers of these special students might strengthen their teaching by integrating material from our database..

Indirect beneficiaries: Given the major enrichment in content, the strategically designed feature of English-Chinese lookup, as well as the adoption of the Unicode standard, the database will attract as much global users as local users. In our local community, university students as well as lay-people who are interested in the Chinese language and script will also find this database useful. This applies also to the vast number of users in Mainland China. Projecting this scenario into the future, our work will affect generations of users who are interested in learning and mastering the Chinese language and script.

4. Conceptual Framework

A. *The Multi-function Chinese Character Database: The Idea and Its Modules:*

Despite the fact that the proposed database is “built-upon” the Chinese character database of our first QEF funded project, it differs from its predecessor in one important aspect, namely the incorporation of a huge amount of information related to traditional Chinese scholarship, including script samples, phonology and etymology, systematically processed and managed on the web. A specially designed English-Chinese look up also opens up a new way of access to such information unprecedented by traditional scholars.

The full name of our new database is: “Multi-function Chinese Character Database: With Archaic Chinese Script Forms, Etymologies, English-Chinese Lookup, and Word-Formations Phonologically Disambiguated according to the Cantonese Dialect”, or in Chinese: 《漢語多功能字庫：古文字繫形、形義通解、英漢索引、及粵語審音配詞》. While the main title specifies the multi-functional nature of the database, the sub-title brings to the forefront its various major functions. Besides the existing feature of word-formations traceable back to our first QEF funded project, the three important modules have been successfully implemented in our second QEF funded project, but the huge amount of work involved still need to be fostered and brought to completion, and to be refined as far as possible through the undertaking of this current project. These modules are:

- a. **Archaic Chinese Script Forms:** For each character, given the availability of reliable resources, selected samples of graphical images of Oracle Bone Scripts (甲骨文), Bronze Scripts (金文), Seal Scripts (小篆), as well as Bamboo Scripts (竹書) and Silk Scripts (帛書) will be graphically incorporated, componentially categorized, and systematically marked-up with an SQL database. One main reason why archaic script forms instead of modern ones are considered lies in the fact that, as a result of the so-called “*lishu*-transformation” (隸變)², many cognitive components of the original script forms have been severely blurred, distorted, cross-matched, and confused, so that what remain in modern script forms are no longer suggestive enough for any serious attempt at an accurate etymological understanding of the Chinese script and its genealogy. The array of old script information mentioned above allows us as well as our users to overcome precisely the semantic barriers caused by *lishu*-transformation. This in turn will open up new horizons for teachers and learners in handling semantic issues related to Chinese language education.
- b. **Etymologies:** Based on graphical and textual evidences and after careful research and systematic researches on the scripts concerned, each characters with oracle bone scripts and bronze scripts will be given an explanatory account in respect of the character’s etymology and genealogy. To give our users a quick insight into the ongoing development or transformation of the script forms, various charts and tables reflecting the graphical and semantic changes over time will be generated for the users’ reference. This part of our work will be presented to the users in the form of:

² *Lishu*-transformation is the developmental process of the Chinese script from the archaic and ancient styles of writing into the classical, and for us nowadays more or less legible styles of calligraphic writing, namely, into *lishu* (隸書=漢隸, or verbally “clerical script”) and *kaishu* (楷書=今隸, or verbally “standard script”). *Lishu*-transformation involved the parallel tendencies of *li-convergence* (隸合) as well as *li-divergence* (隸分). While the former is the tendency for originally disparate components to be stereotyped into one same way of writing, the latter is but the opposite tendency for one originally identical component to be written differently when used in building up different characters.

Chinese texts up to the sixth century AD. The CHANT project first began under the tutelage of the late Professor [redacted] in collaboration with Professor [redacted] and Professor [redacted]. Since the retirement of Professors [redacted] and [redacted], Professor [redacted] has taken up the important duty of overseeing the progress and completion of this monumental work. CHANT has been supported since 1988 by a number of RGC grants, totaling more than 10 million subsidy in HK\$. The CHANT project has led to substantial publications, as well as tens of million words of texts released on the Internet (www.chant.org), covering extant texts on oracle bones (*jiagu*), bronze utensils (*jinwen*), silks and bamboos (*jianbo*), and those in a number of classical encyclopedic works (*leishu*). CHANT merits the homage of the local and international scholarly community, because by the provision of an electronic infrastructure for extant Chinese text CHANT paves the way for any future research work related to a systematic study of Chinese language and culture.

Among the many beneficiaries of the CHANT legacy, the current project team has benefited mostly, and this indeed through the personal commitment of Professor [redacted], who courteously supplied himself as one of the Co-Investigators of the current project. CHANT's support was not only moral, but substantial. When the PI was conceiving the possibility of the current project, Professor [redacted], upon knowing the idea behind our entire (long-term) plan, immediately agreed to grant us access to all the fonts (in the case of oracle bone scripts) and graphical images (in the case of bronze scripts, bamboo and silk scripts) they developed. Besides discussing with the PI about the various details of the project, Professor [redacted] has instructed the entire working team of CHANT to work closely with the PI's team so that within a short time, the fonts and graphics (together with inventories in Excel format) were successfully transmitted and converted by the PI's team into another format suitable for further processing.


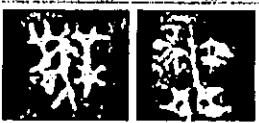



While the provision of raw script data by CHANT was an essential help, the work to be completed by the current project team has yet to be done anew. As its very name suggests, CHANT deals primarily with "ancient texts", but the current project aims at getting into the **internal structures** (i.e., the componential level) of the very characters themselves, as they were represented in their archaic and ancient script forms. We have focused on "**components**" (部件) and use them emphatically to supplement the more commonly known "**radicals**" (部首) because the latter were precisely defective results of the problematic process of "*lishu*-transformation" as described above. Regarding the study of Chinese archaic script components, it is obvious that countless work has already been done, both in history and in more recent times, but a systematic treatment of archaic script components with database strategies as conceived by the applying team is still wanting.

5. Innovation

All in all, this project considers itself as an innovative integration of traditional Chinese etymological scholarship with database management.



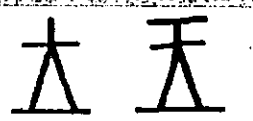


A. *Regarding computational strategy:*

All relevant texts, images, comments, annotations, etc., will be stored in a mySQL database in Unicode format. To facilitate the systematic treatment of the complex relationships between the

小篆	金文(部件)	甲骨(部件)	簡帛文字	傳世古文
	 白舞 無 大 人 止	 白舞 大 人		
	圖片：4/6	圖片：3/3		



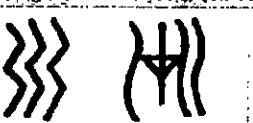
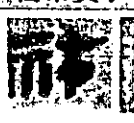

形義通解：

無、舞二字古時本無區別，甲骨文各字形皆象一人手執飾物(◎)翩翩起舞之狀，換言之，無的本義本即是舞。後來「無」被借作虛字，以表達「有無」的「無」，結果，舞蹈這個原來的意義反而要從原字的基礎上修改，另造成後來的「舞」字。從金文的字形看，舞其實是於無字中的「大」(人)字下面加上「舛」，即交叉相錯的二足(止)，以突出起舞時踟躕的舞姿。《左思·蜀都賦》中有「都人仕女，袿服靚粧；賈質增器，舛錯縱橫」之句，可視為舛字可解作「步履交錯」的側證。參見「無」，參見「舛」。

小篆	金文(部件)	甲骨(部件)	簡帛文字	傳世古文
	 白王 戌	 白王 戌		
	圖片：15/2173	圖片：4/4		

形義通解：

王字廣見於甲骨文及金文，漢許慎《說文》中「王」字的解釋為：「天下所歸往也。董仲舒曰：『古之造文者，三畫而連其中謂之王。三者，天、地、人也，而參通之者王也。』」然從甲骨及金文的字形察看，古「王」字實象斧鉞之鋒口向下形，斧鉞是軍權的象徵，因以稱王。「王」是把「戌」字轉九十度，是全字轉向以與原字區別的指事字。這種情況，於「土」字的結構亦可得見，參見「土」。

小篆	金文(部件)	甲骨(部件)	簡帛文字	傳世古文
	 白災 才 戈 其他：中	 白災 川 才 其他：一，火，戈，災		
	圖片：6/19	圖片：16/17		

形義通解：

「災」字異體繁多，小篆作「𤇓」，或省火作「𤇓」、「𤇓」或「𤇓」；又作「灾」、「𤇓」等，《說文》：「𤇓，天火曰𤇓。从火𤇓聲。灾，或从一火。𤇓，古文从才。災，籀文从𤇓。」

甲骨文有多種寫法：第一種表示水災，從川(古同「川」)或橫寫之𤇓，象洪水橫流，隸定後作「𤇓」，又作「𤇓」；後於「川」中加「才」作為聲符，形聲字例當後起。第二種表示火災，從一從火作「灾」，指室中失火釀禍，乃簡體中文之所本；又有從火從才聲作「𤇓」，與《說文》古文同。

第三種表示兵災。甲骨文兵災之字有二，除表示災害時可通用外，區分明確。第一為從戈從才聲，「戈」為兵器，有傷害之義，引伸為兵災；《說文》小篆「𤇓」即以「𤇓」下加「火」而成；又用為名詞，「𤇓方」即是方國名；晚期卜辭「𤇓」借用作「𤇓」，語氣詞，參見「𤇓」、「𤇓」。第二為從中從戈，除表示災害外，又用作動詞，表示在戰爭中傷害、毀壞；「中戈」表示傷害、打敗，是就戰爭的結果而言，表示敵人被傷害，以別於「征」、「伐」、「𤇓」等表示戰爭中的征伐行動。「𤇓」、「中戈」皆有傷害義，故亦有學者認為是一字的分化(參單周堯)。

以上六形，表示災害，則通用無別；可見甲骨文「災」字泛指各種自然或人為的禍害。商承祚云：「以其義言之，水災曰𤇓，兵災曰𤇓，火災曰𤇓。」晚期甲骨文則統一作從川從才。「𤇓」、「灾」、「𤇓」後世皆以《說文》籀文「災」代之，乃合川火而成災，統言水火乃自然災禍之由。

金文只見「𤇓」及從中從戈兩字。「𤇓」借作「𤇓」，而從中從戈一字上承甲骨文，保留戰爭中的傷害一義，又用作族氏名，然皆與災害無關，參見「𤇓」。

「災」古籍還借「𤇓」為之，《詩·生民》：「無𤇓無害」；亦作「𤇓」，《史記·秦始皇紀》：「𤇓害絕息」。

在《馬王堆老子乙本卷前古佚書》之中，災字曾寫作「𤇓」，即從示從才，其中「才」是聲符，而從「示」則又進一步紀錄了古人認為災禍乃由天所降之心跡。這一點，《左傳·莊公十一年》：「孤實不敬，天降之災」一語可資佐證。

小篆	金文 (部件)	甲骨文 (部件)	簡帛文字	傳世古文
	白義 白我 戈 羊	白義 白我 戈 羊		
	圖片: 4/109	圖片: 1/1		

形義通解:

「義」字的甲骨文和金文都從我從羊。「義」字的構形似簡易實頗繁複，因為「我」、「羊」兩個部件都不能從今日的字面意義去理解。首先，「我」原初並不指今日第一人稱的「我」，而是古代的一種兵器，至於「羊」，亦不直接指動物的羊，而是經引申而泛指美善之事或物。故《說文》訓「善」字時即有：「善，吉也。從誥從羊，此與義美同意」之語。至於「義」本字，《說文》的訓釋是：「義，己之威儀也。從我羊。」此中，「己」雖然未必合於原義，但「威儀」一訓卻是正確的。事實上，「義」最初的意義可能正是「威儀」、「儀仗」的「儀」，而「威儀」和「儀仗」又可能與「兵威」或「軍容」有關，故從「我」（一種兵器）。另一方面，古人可能認為藉著「義」這種威儀，美善乃得以維繫，故從「羊」。後世言「義軍」、「義師」、「起義」，大概亦由此而引申。

由於儒家的影響，義後來與仁並稱為「仁義」。不過，由於「義」中有「我」（兵器）的成份，故仁，義二者雖然都帶有道德意義的善，但其中「仁」偏於柔而「義」偏於剛，還是有跡可尋的。近世出土的《郭店楚簡·五行》中有關「仁義」的討論中有「（剛），義之方；（柔），（仁）之方」之語，基本上支持了這一點（此外，《周易·坤·六二·文言》「君子敬以直內，義以方外」一語亦可作側證）。不過，在文字發展的過程中，「義」字後來也被借作「意義」的「義」，並成為直到今天「義」字的另一個重要義項。至於「儀」、「議」二字，應該是從「義」的這兩個意義分別加注「人」與「言」而衍生的。

最後，有學者認為「義」字可能與《說文或體》的「彘」字有關，亦有認為與「悵」字前身的「彘」有關，姑存其說。

小篆	金文 (部件)	甲骨文 (部件)	簡帛文字	傳世古文
	白環 玉 白衰 目 衣 ◎ 其他: 衰, 手			
	圖片: 5/13			

形義通解:

「環」字於甲骨文中未見，金文的「環」最初只寫作「衰」，到後來才附「玉」於左側。「衰」的構形從衣從◎從目。目有時在衣之上，泛指一穿衣的人，◎（圓形小圈）應指此人懷於身上的圓形的器物。有時在「衣」之中，圓圈之上有「省」（從羊從目，即「祥」的異體），是指所懷者乃一祥瑞之物，而一些字形從手，是指以手把弄之（參見「弄」）。傳世的青銅器中，便有不少這一類的圓形的並刻有「衰」字的雜器。後來由於有用玉製成的衰，為了清楚表明所指的是玉塊，才有加注「玉」為義部的「環」字。

由於「衰」或「環」圓的形狀，後來便引申為環珮以外一般意義的「圓環」，和甚至「迴環」、「環繞」等意義，並由此而孳乳出許多以「衰」為邊旁而又帶有某一意義的「圓」的漢字：如注之以口而得「園」（指圓環，即對衰中的口再予強調），注之以走而得「還」（往而復返如圓之回到本位），注之以宀而得「寰」（指字內的方圓），注之以人而得「環」（指人之圓滑狡猾），注之以羽而得「翮」（指環繞飛翔），注之以水而得「環」（指回旋涌起的波流），注之以心而得「懷」（指心思糾纏繁繞），注之以髟而得「髻」（指「盤髮如環」，即梳成環狀的髮髻），注之以車而得「輦」（指「車裂」，即五馬分屍），注之以糸而得「纆」（指纆首之刑）等，其皆取義於吾人懷中玉環之「圓」，可謂同條牽屬。此外，我們又應注意，上面孳乳的各字中，「衰」都不是義部，而是一聲符，但作為聲符，「衰」卻帶有一籠統的「圓」的意義。而「衰」之在各字中的作用，其實都要帶出某一意義的「圓」。由是觀之，上引各字間的關係，實即「右文說」或「聲符載義」學說的又一範例。

Note that in the actual database environment, characters in blue are hyperlinked, all slide bars mean more script samples are available, and "mouse-over" on all script samples will reveal source of that sample (collection, bronze utensils, etc. with serial numbers), prompting our users to explore our database for further information.

B. Regarding traditional Chinese scholarship:

As far as semantics is concerned, Chinese scripts might involve thematic issues of the most different topics exemplified as follows : **Natural sciences** (astronomic events such as planetary orbits [e.g. 農·歲·參·商], solar/lunar eclipses, calendar [e.g. 春·秋·季·時], zoology [e.g. 虎·狐·為·兕], botany [e.g. 來·黍·禾·果], meteorology [e.g. 申·云·易·籊]); **history** (official laudation, imperial/royal matrimonies, royal appointment of offices, warring activities [e.g. 旅·旋·師·征], punishment and massacres [e.g. 服·孚·臧·蔡], legal disputes [e.g.: 訊·訟·法(灋)·律]); **social systems and anthropology** (slavery [e.g. 奚·奴·妾·僕], gender relationship [e.g. 取·帚·妻·媵], coat of arms of clans [e.g. 萬·戊·亞·單], educational curricula [e.g. 孝·改·事·典], worships and rites [e.g. 祝·且(祖)·羞·豐], mythologies [e.g. 鬼·神·夔·鷹], divinations [e.g. 問·卜·貞·占], burials [e.g. 葬·棄·死·喪]); **economics** (household activities [e.g. 盥·鄉(饗)·既·飲], trading [e.g. 買·賈·易·賞], use of currencies [e.g. 貝·朋·玉·珌], hunting [e.g. 禽(擒)·獸(狩)·射·弔], livestock farming [e.g. 牛·羊·豕·豕], agriculture [e.g. 封·藉·農·喬(穡)]); **technologies** (jade manufactory [e.g. 璋(璋)·環·璽(璽)·璧], arts and crafts [e.g. 文(紋)·朕·巨(矩)·工], textiles and clothing [e.g. 初·裘·絲·絕], utensils [e.g. 彝·鼎·孟·簋], weaponry [e.g. 戈·我·弓·矢], metallurgy [e.g. 鑄·錫·銘·冶]), and of course, **philosophy** (abstract concepts [e.g. 無·幾·苟(敬)·中], moral precepts [e.g. 仁·義·善·德], emotions [e.g. 懼·蔑·畏·憂], cognitive models [e.g. 疑·哲·思·念], metaphysical fictions [e.g. 道·天·神·帝]), etc. In short, our Chinese character database with archaic Chinese script information embedded and etymologically explained is like Chinese culture in a nutshell. In this way, the database could also lend support to liberal studies and Chinese culture subject of the school sector.

6. Extent of Teachers' and Principals' Involvement in the Project

While the project itself especially the construction of the database will be carried out by the RCHC at the Chinese University, Chinese language teachers from secondary schools will be invited to take part as beta users before the final release of our work to the public. This is necessary because as targeted users of our database, the strengths and weaknesses of our work from their perspective are useful information that might help improve our work. Our plan is that:

- a. After some initial phase of preparation we will first invite teachers/principals from the TABS group of secondary schools³ to come for a briefing and initiation session for the use of our database. We believe that the TABS schools belong to some of the best schools in Hong Kong and their feedback on our database will provide us with valuable indications on didactic issues we should address from the users' point of view.
- b. Special accounts will be issued to the teachers of the TABS schools so that they can, over a period of time, start to use our database in their actual teaching preparation, and report to us on all kinds of real-life scenarios and teaching problems in using the database. Besides bug reports, we also welcome suggestions of possible addition of functions, presentation strategies, what cross references or hyper-links to be established, etc.

³ TABS is the acronym for "Towards a Better School Movement (學校達善運動)". It comprises currently 25 schools with an Exco now chaired by [redacted] and vice-chaired by [redacted] who was previous chairman of TABS, will on behalf of TABS continue to offer us full support. For details of TABS, see: <http://www.tabs.edu.hk/>.

- c. At a later stage of the project, workshops open to teachers of all middle schools of Hong Kong will be organized. Alongside our own presentations, dedicated teachers from the TABS schools will be invited to speak on how the database can be used to improve teaching. Related lectures will be organized so that Chinese language teachers of our targeted schools will be invited to participate, with elaborated sessions of Q & A.
- d. While a bilingual documentation/user manual of the database should have been drafted before official release of the database, such documentation will be revised and fine-tuned after taking the opinions of the school sector into consideration. Furthermore, some promotion pamphlets/flyers focusing on the various useful features of the database will be sent to the school sector.
- e. After the completion of the project, teachers of partner schools will be encouraged to keep in touch with us so that we can further benefit from each other on issues related to contents and usage of the database. We might consider setting up some kind of web-forum so that teachers of the school sector might, even upon completion of the project, use the forum as a channel for their own interaction with reference to the use of our database.

7. Implementation Plan with Time-line

Owing to the huge workload and the data complexity of the database the project will have to cover a span of three years to safeguard quality. However, our work will be partially released for beta testing in the middle of the second year:

Phase/Period	Work
Phase 1: Preparation (4 months)	<ul style="list-style-type: none"> ● Recruiting staff ● Reviewing existing work modules to be built-upon ● Revising and optimizing electronic work bench ● Briefing and training new staff ● Continuously screening reference material for use
Phase 2: Intensive work (16 months)	<ul style="list-style-type: none"> ● Categorization of archaic/ancient script components ● Writing of etymological accounts ● English-Chinese look-up ● Input of reference pagination information ● Review web-page design and service features ● Preparation for Beta-release of database to targeted schools and to invited experts ● Organization of topic-related public talks for school teachers
Phase 3: Peer Review / Adjustments (10 months)	<ul style="list-style-type: none"> ● Briefing and initiation session for teachers/principals of partnership schools (tentatively TABS) ● Continuation of remaining work, especially etymological accounts ● Collection of opinions from TABS as well as from external experts ● Analysis of feedback ● Organization of topic-related public talks for school teachers ● Internal review for improvement
Phase 4: Refinement / Documentation (6 months)	<ul style="list-style-type: none"> ● Refining database structure ● Refining database front-end ● Finishing remaining part of work ● Bilingual documentation (users' manual) ● Invited talks by external experts ● Enabling public access. Dispatch of promotion package ● Publication of related articles in academic and educational platforms

8. Expected Deliverables and Outcomes

- a. Multi-function Chinese Character Database: With Archaic Chinese Script Forms, Etymologies, English-Chinese Lookup, and Word-Formations Phonologically Disambiguated according to the Cantonese Dialect.
- b. Besides the above user-oriented web page, a corresponding web-driven front-end for administration will be in place. This guarantees effective maintenance, debugging, and updating of the database in the long run.
- c. Full documentation (users' guide) of the database in both English and Chinese.
- d. Related sub-databases as by-products, e.g. reference pagination, set phrases indices, etc.
- e. Promotion package on database usage to be sent to all secondary schools in HK.
- f. Initiation Workshops, test accounts of database, and lectures related to archaic Chinese scripts, etymologies, and their relevance for contemporary Chinese language education, etc. will be organized for Chinese language teachers of the school sector.
- g. Series of anthologies:
 - (i) Series A: Academic articles related to Chinese language and scripts.
 - (ii) Series B: Selected records from database with script samples and etymologies.

9.

<u>Budget</u>	<u>Quan.</u>	<u>Item</u>	<u>Cost</u>			<u>Total</u>	
(a) Staff Cost							
			<i>initial</i>				
			<i>Salary/m</i>	<i>1st yr.</i>	<i>2nd yr.</i>	<i>3rd yr.</i>	
1	(i)	Research Assistant with PhD/M. phil. Med Ins (*)+ MPF=\$1000/month.	\$20,010	\$240,120	\$240,120	\$240,120	\$720,360
				\$12,000	\$12,000	\$12,000	\$36,000
1	(ii)	Research Assistant with B.A. Med Ins (*)+ 5% MPF	\$14,450	\$173,400	\$173,400	\$173,400	\$520,200
				\$8,670	\$8,670	\$8,670	\$26,010
		Sub-total		\$434,190	\$434,190	\$434,190	\$1,302,570
(b) Equipment (necessary for start-up)							
			<i>Unit cost</i>	<i>1st yr.</i>	<i>2nd yr.</i>	<i>3rd yr.</i>	
x		Main server cost + annual maintenance will be absorbed by Research Centre/CUHK					\$0
1		Off-site backup Server running on RedHat Linux	\$13,000	\$13,000			\$13,000
1		Printer & Scanner		\$8,000			\$8,000
2		Notebook computers	\$8,500	\$17,000			\$17,000
2		Desktop PCs for new staff/	\$6,000	\$12,000			\$12,000
		Sub-total		\$50,000			\$50,000
(c) Services							
				<i>1st yr.</i>	<i>2nd yr.</i>	<i>3rd yr.</i>	
1		Computation, Web page Design & Construction		\$40,000	\$50,000	\$40,000	\$130,000
			<i>Unit cost</i>				
1		Visiting Scholars (Travel Stipend+Per Diem)	\$17,600		\$17,600	\$17,600	\$35,200
		Consultation Fees for visiting scholars	\$12,000		\$12,000	\$12,000	\$24,000
2		Student helpers @\$50/hr.		\$12,000	\$12,000	\$12,000	\$36,000
1		University Overhead		\$13,682	\$13,682	\$13,684	\$41,048
		Sub-total		\$65,682	\$105,282	\$95,284	\$266,248

(c) General expenses	1st yr.	2nd yr.	3rd yr.	
Such as consumables, books, software, and research materials, workshops and lectures organization/honoraria, network rights for ancient scripts payable to Chin. Publishers, local transportation, mailing, etc.	\$92,200	\$109,500	\$109,500	\$311,200
<i>Sub-total</i>	\$92,200	\$109,500	\$109,500	\$311,200
Contingency	\$6,274	\$6,274	\$6,334	\$18,882

Grand Total	<i>Triennium breakdown</i>	1st yr.	2nd yr.	3rd yr.	
		\$648,346	\$655,246	\$645,308	\$1,948,900

* Expenses for medical insurance will be sought through other means

Justifications:

- (i) Research assistant of PhD/M.phil. rank: Owing to the vast amount of qualitative work, which forms the core of the working procedures and is the key for a long-term impact of the project, a basic crew with two research assistants is a minimum requirement. Besides the need to cover the actual work load involved, it is also necessary to have at least two RAs (preferably of slightly different backgrounds) so that they can readily discuss difficult and controversial issues among themselves. Of the two RAs, one has to be hired at a higher rank to guarantee quality of our work.
- (ii) Research assistant with B.A.: In order to pave the way for a sustained development of the project, it is necessary to have a second RA who has experience with archaic Chinese scripts and preferably with humanities computing. Besides creating new blood for the future of the project, this post is also necessary in view of the tremendous amount of qualitative-analytical work to be done. It also acts as a buffer to guard against possible manpower fluctuation during the funding period.
- (iii) Visiting scholars: In order to benefit from the expertise of a wider scholarly circle of classical Chinese etymology, one renowned scholar working in related fields will be invited in the second and third year of our project as visiting scholar to stay short-term to discuss with our team, comment on our database, give lectures to schools teachers and participate in workshops.
- (iv) Student helpers will cover some repetitive but labor intensive work.
- (v) Equipment: Note that the RCHC receives no regular budget monies from the CUHK. The server cost has so far been covered either through the Academic Equipment Grants (shared with the Department of Philosophy), or through donation from third parties (e.g. HP). Owing to some major upgrading recently, we consider our hardware as still fit for the span of the project. However, the RCHC has no budget to cover desktop computers for new staff to be recruited for the project.
- (vi) Programming, webpage design and construction: Given the high complexity of the database and the continuous need of programming revision, module addition, data structure modification, web-layout optimization, etc., and the need of system overhaul in the course of interaction with the school sector, an in-house programmer would be the best solution, as was the case in the earlier phase of this project. But given the resignation of our previous programmer, enough money has to be set aside for outsourcing of programming work, or if the situation requires it, for the recruitment of a programmer on part-time basis, in which case, a reshuffling of budget items might be necessary.

- (vii) Reference Books: In view of the vast amount of important literature in the field, and the need for day to day reference, our in-house reference library has to be enriched to guarantee the academic quality of our work.
- (viii) Network rights: Although we have received gratis many graphical images from CHANT, there are still some collections of texts/scripts that we need to cover in our database. Monies have therefore to be earmarked to negotiate with individual publishers in China for the right to use certain texts or images on the network.
- (ix) Organization of workshops/lecture series: This creates a forum of discussion to promote the general interest for the understanding of Chinese language and script. Besides academic issues, issues of didactics will also be focused. Contributions of the workshops and lecture series can be compiled as anthologies to accompany the final database so that the effort of the project will be more “tangible”.
- (x) University Overhead includes the most basic services provided by the university to safeguard the smooth functioning of our project.

10. Evaluation Parameters and Method

- a. The project team, with its multi-disciplinary composition of renowned scholars (3 department chairmen, 2 research centre directors) will guarantee the quality of the project in the first place.
- b. To facilitate the long term work of the project, a web-driven management platform has been developed which supports the addition and editing of records, debugging and back-tracking of our work so that we can trace the work of individual workers, on particular modules, in particular time spans, and with particular data types, etc. In other words, we are in possession of a self-designed computational workbench that allows us to evaluate every single aspect of our work so that we can quickly learn from our own mistakes safeguarding thus the future quality of our work.
- c. Teachers from the TABS schools will be our most treasured evaluators, as they can comment on our work from the perspective of our end beneficiaries. We will build up a web-driven platform for all feedbacks of teachers to be channeled to the project team.
- d. External experts (from HK, Taiwan, and/or mainland China) will be invited to evaluate our database at different stages of its development.
- e. We will keep access logs of our database and perform data analysis regularly to gain quantitative indicators for the merits or otherwise of our work.

11. Sustainability of the Outcomes of the Project

As suggested by the subtitle of this project, what we are planning to work on could be compared to the building up of an infrastructure that should have long lasting influence on basic Chinese education. Given the proven popularity and indispensability of the project’s predecessors (see section 12), we have grounds to cautiously expect the same success for this current project, which has so much more powerful features to offer while retaining the original modules.

The sustained success of our earlier work also indicates that after completion of the project, the output database can run on its own while gaining influence with time. We on our part, with a well designed management/maintenance platform in place, will take up the task to modify and update the database wherever possible. Given the coverage and

prospective quality of the database, we expect it to benefit not only today's teachers and clientele, but many more generations of them. In short, we think it is no exaggeration to claim that eventually the entire education sector will benefit from our work.

12. Dissemination / Promotion

Since our current project is related to our earlier projects and that the Multi-function Chinese Character Database is a further development built-upon our earlier work, we have to look back on our track record to assess the future impact of our work:

- A. **Quantity:** Two constituent web pages produced by the Q&T project have experienced unforeseen popularity from all circles of users. The electronic web edition of "Lin Yutang's Chinese-English Dictionary of Modern Usage", the oldest product of the project, has been searched more than **36.1 million** times to date. The other webpage "Chinese Character Database: With word- formations phonologically disambiguated according to the Cantonese Dialect", which was the youngest of the constituent projects, has been searched more than **73.5 million** times to date, becoming our most popular output so far.⁴
- B. **Quality:** The QEF-supported B&T project was distinguished as the **most outstanding research project** representative of the Arts Faculty, CUHK, for 2001-2002, and a feature appeared in the 2001-2002 CUHK Annual Report. (p. 37) [see: <http://humanum.arts.cuhk.edu.hk/QEF5073/image/cuhk-an-rep-feature1.jpg>]
- C. Over the past year we were much encouraged by some new development. Firstly, the Lin Yutang Dictionary web page has been nominated three times (2006, 2007, 2008) by the Hong Kong Education Bureau for the **Meritorious Website Contest**. In 2009, it was the Chinese Character Database (direct predecessor of the applying project) which was nominated for the same contest. Secondly, after repeated attempts to contact us, **Google** has once signed a preliminary agreement with the CU for the licensing of some information modules of our previously developed Chinese Character Database, which is an obvious indication of the value of our work.⁵
- D. **Extent of Clientele:** Besides being linked by many schools, access logs of our server have shown that the tremendous volume of accesses to our web pages is evenly distributed between the local and global clienteles. For more information see our access analysis online.⁶ With such a widespread clientele established, we plan to retain the major IP address of the previous work in disseminating the new database. In this way the future dissemination and impact of the new database can be guaranteed.

Finally, it might be noteworthy that a great motivation behind this project is the attempt to render cognitive information and cultural contents immanent in the Chinese language and script more accessible to contemporary users by the use of new technology and media. We hope that, in this way, besides furnishing an infrastructure in support of basic Chinese education in HK, the competitiveness of Chinese as an academic and intellectual

⁴ Other technically much more sophisticated and for the users more demanding web pages (e.g. the "character frequency database", and to some extent our RGC project on XML) were less popular, but they contribute significantly to the accumulation of our technical experience.

⁵ Unfortunately the CU Licensing office eventually failed to reach an agreement with the Google team because of some trivial and hairsplitting terms.

⁶ See <http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/access-analysis>.

language can also be enhanced to meet the challenge of the global dominance of English.⁷ With the valuable experience we have accumulated, with the dedicated and complementary research team we have assembled, and with a well-designed working platform now in place, it is a golden chance for Hong Kong to undertake such a project so that we can make a difference by throwing new light on Chinese language education and all Chinese related disciplines of learning. By supporting this project, a great number of users might benefit, be they teachers or students, local or global, for now or for generations to come. If necessary, the PI is delighted to give a presentation of the project upon request of the QEF.

⁷ See Tze-wan Kwan, "The Over-dominance of English in Global Education - Is an Alternative Scenario Thinkable", paper presented at the "Ninth East-West Philosophers' Conference". Now published in *Educations & Their Purposes: A Philosophical Dialogue Among Cultures*, edited by Roger T. Ames and Peter D. Herschok (Honolulu: University of Hawai'i Press and East-West Philosophers Conference, 2007), pp. 54-71. On the suggestion of Professor Th. Wm. de Bary, formerly Provost of Columbia University, and with the consent of Professor Ames, convener of the Honolulu conference, the above paper was slightly modified and incorporated as a chapter in de Bary's most recent book *Confucian Tradition and Global Education: Essays in Honor of Tang Junyi* (New York: Columbia University Press, 2007), pp. 75-99. The paper there bears a slightly different title: "The Overdominance of English in Global Education - A Glocal Response". Besides incorporating this article, Professor de Bary also wrote another article in reply to issues raised in the PI's paper. For other papers published by the PI on language theory, on Chinese language, and on language policy, see: <http://epimetheus.arts.cuhk.edu.hk/lang/>

Asset Usage Plan

Category (in alphabetical order)	Item / Description	No. of Units	Total Cost	Proposed Plan for Deployment <i>(Note)</i>
audio and video equipment				
book & VCD	Books and Research Materials	To be confirmed in the future	\$35,000	Will be deployed to the Research Centre for Humanities Computing(RCHC) to facilitate research work, and will be kept in RCHC for future references upon project completion
computer hardware	A. Off-site backup Server running on RedHat Linux	1	\$13,000	Will be deployed to RCHC to carry out research work, and will be kept in RCHC for long term sustainable development of the same project upon project completion.
	B. Notebook computer	2	\$17,000	
	C. Desktop PCs for new staff	2	\$12,000	
computer software	software	To be confirmed	\$10,000	Will be deployed to RCHC to facilitate research work, and will be kept in RCHC for long term sustainable development of the same project upon project completion.
musical instrument	---			
office equipment	Printer & Scanner	2	\$8000	Will be deployed to RCHC, and will be kept in RCHC for long term sustainable development of the same project upon project completion.
office furniture	---			
sports equipment	---			
Others	---			

Note: for use by school / organization / in other projects (please provide details of the department / centre to which the asset will be deployed and the planned usage of the asset in activities upon project completion).

Report Submission Schedule

遞交報告時間表

I / My school / My organization commit(s) to submit proper reports in strict accordance with the following schedule:

本人/本校/本機構承諾準時按以下日期遞交合規格的報告：

Project Management 計劃管理		Financial Management 財政管理	
Type of Report and covering period 報告類別及涵蓋時間	Report due day 報告到期日	Type of Report and covering period 報告類別及涵蓋時間	Report due day 報告到期日
Progress Report 計劃進度報告 1/1/2011 - 30/6/2011	31/7/2011	Interim Financial Report 中期財政報告 1/1/2011 - 30/6/2011	31/7/2011
Progress Report 計劃進度報告 1/7/2011 - 31/12/2011	31/1/2012	Interim Financial Report 中期財政報告 1/7/2011 - 31/12/2011	31/1/2012
Progress Report 計劃進度報告 1/1/2012 - 30/6/2012	31/7/2012	Interim Financial Report 中期財政報告 1/1/2012 - 30/6/2012	31/7/2012
Final Report 計劃總結報告 1/7/2012 - 31/12/2012	31/1/2013	Final Financial Report 財政總結報告 1/7/2012 - 31/12/2012	31/1/2013
Progress Report 計劃進度報告 1/1/2013 - 30/6/2013	31/7/2013	Interim Financial Report 中期財政報告 1/1/2013 - 30/6/2013	31/7/2013
Final Report 計劃總結報告 1/7/2013 - 31/12/2013	31/3/2014	Final Financial Report 財政總結報告 1/1/2011 - 31/12/2013	31/3/2014