Database of Selected Characters from Guodian and Mawangdui Manuscripts

Introduction by Matthias L. Richter, February 2006

The database is a small-scale example of a collection of data for the purpose of investigating and describing early Chinese manuscripts in detail. It is intended to demonstrate that such data can serve as a useful tool for students of these manuscripts. The database allows the user to sort and extract a variety of physical data of the manuscripts and thus enables him or her to look at comparable features (especially all instances of characters used to write a certain word) simultaneously, which is otherwise not possible. This allows to judge the degree of variation vis-à-vis standardisation on the levels of writing support, text layout, orthography and handwriting more easily and more objectively than by looking at the objects of comparison one by one.

For a study based on this collection of data, see my article "Towards a Profile of Graphic Variation: On the Distribution of Graphic Variants within the Mawangdui *Laozi* Manuscripts" in: *Asiatische Studien / Études Asiatiques* LIX.1, pp. 169–207. Further studies based on the same data will be published later this year and will be announced on this website.

The intention of presenting this collection of data on the internet is chiefly to offer an example as a stimulus and basis for discussion among students of early Chinese manuscripts about the question of how future databases should be designed. Everyone is welcome to use the data I have collected and adapt them according to their own judgement that may of course differ from my interpretation of details.

Quality of data:

This data collection is a mere model, also because the reliability of data needs to be improved in several aspects: The quality of photographic reproductions is clearly unsatisfactory in the case of the Mawangdui manuscripts, but even in the case of the Guodian manuscripts the user must be cautioned that the data refer to the published reproductions and not to the manuscripts themselves. Rather often characters seem to be incomplete at the edges. Whether the edges of the bamboo slips themselves are decayed or whether they were cut off in the photos can only be guessed at. Some examples prove that there are definitely problems about the photos, e.g. the recto and verso of both slip 27 of the manuscript *Yu cong* 4 語叢四 (GD18.27) and slip 40 of the *Zi yi* 緇衣 MS. (GD5.40) differ by 2 millimetres

each. Scientifically controlled high-quality digitisation of the manuscripts is of the utmost importance for future research in this field.

Scope of chosen data:

The data collected here comprise only a few manuscripts, since the data were all gathered in a laborious process of manually cutting out of characters, storing them, entering them into the database, measuring them, entering transcriptions and readings, references etc. The ratio of the great amount of labour input in relation to the benefit of the product can be drastically improved in the future by computer-aided methods of character analysis and semi-automatic data input. (Dr. Imre Galambos of the International Dunhuang Project at the British Library has already developed a software for semi-automatic extraction of individual characters from digital images of Chinese manuscripts and is preparing a database of Dunhuang characters.)
In the files pertaining to the Guodian manuscripts, the file "GD columns" comprises data of the entire corpus of bamboo slips, except for the 27 fragments that could not be attributed to certain manuscripts. Metadata for individual characters have been collected only for the manuscripts *Laozi* A–C, *Taiyi sheng shui* and *Yu cong* 4 (GD1–4+18) that were all written in the same style of script. The metadata of the individual characters are not yet complete.
Of the Mawangdui finds, only selected characters of two silk manuscripts containing a version of the *Laozi* and four other texts each (see the table under "manuscript codes" below) were collected. The choice of characters followed the needs of research. On the one hand, characters that are especially difficult to distinguish from each other were chosen e.g. 无/先, 天/夫. Another research interest was to determine the actual degree of the alleged orthographic arbitrariness, the most common form of which in the Mawangdui manuscripts is classifier variation. Therefore, characters sharing the same phonophoric and occurring with sufficient frequency were collected, e.g. the 青-series. As a third kind, characters were collected whose reading is certain but who are written differently (in abbreviated and full forms), e.g. 者 and 其/亓.

Structure of the database:

At present, the data exist in the form of "Excel" tables. Each line in these tables represents one physical object to which a number of metadata in the several columns of the same line relate. The general logic is that of a progression from the larger to the smaller physical object, i.e. from the level of manuscript to that of the individual columns of writing and then to the individual characters in a column. In the "columns" file, one line stands for

one column, in this case one bamboo slip. In the "character" files, the object of each line is one manuscript character. There are no extra entries for punctuation or other marks, as this would have lead to ambiguities of reference. As marks in the manuscripts are usually linked to the preceding character, they are registered in the database in the same entry together with this character as well. The data in the entries for the individual character are arranged in a progression from the technical data of the object (position and measurements) to the image of the object itself and its graphic analysis, down to the interpretation of its graphic form, i.e. the reading, and information about the word presumably represented by the character. At the end some references pertaining both to the graphic as well as to the linguistic levels are given.
The "Excel" files allow sorting of the data by three criteria in hierarchical order. Each column can be chosen as a sorting criterion. However, not all columns are useful for this purpose, especially those that only or mostly contain images. The most useful method of sorting will probably be by (1) pronunciation, (2) reading in publication [RP] and (3) character code.

Source of data:

All data – from the arrangement of manuscripts, numbering of columns down to the transcriptions, readings and images – are given according to the standard publications of the manuscripts.

For the Guodian manuscripts: ***Guodian Chu mu zhujian*. 1998. Jingmen shi bowuguan (ed.). Beijing: Wenwu.** 《郭店楚墓竹簡》 荊門市博物館編；北京：文物**.**

For the Mawangdui manuscripts MWD1 and MWD2: ***Mawangdui Han mu boshu*. Vol.1. 1980. Guojia wenwuju gu wenxian yanjiushi (ed.). Beijing: Wenwu.** 國家文物局古文獻研究室 編《馬王堆漢墓帛書》 北京：文物**.** (In a few cases, I have added images taken from colour photographs acquired from the Hunan Provincial Museum in Changsha.)

Following these publications does not imply any judgement of correctness but is intended to ensure unambiguous reference.

In certain fields, the volume of the entry exceeds what can be displayed in the chosen format of the table. In these cases, a double click on the field will fully display the entry.

In the following, explanations are given as to the names and contents of the various columns of the "Excel" tables. Explanations for the "columns" data precede those for the "characters" data.

Any comments are most welcome (e-mail: matthias.richter@colorado.edu).

"manuscript code"

Guodian manuscripts

ms. code	name / title		no. of slips
GD1	*Laozi* A	老子甲	39
GD2	*Laozi* A	老子乙	18
GD3	*Laozi* A	老子丙	14
GD4	*Tai yi sheng shui*	太一生水	14
GD5	*Ziyi*	緇衣	47
GD6	*Lu Mu gong wen Zisi*	魯穆公問子思	8
GD7	*Qiong da yi shi*	窮達以時	15
GD8	*Wu xing*	五行	50
GD9	*Tang Yu zhi dao*	唐虞之道	29
GD10	*Zhong xin zhi dao*	忠信之道	9
GD11	*Cheng zhi wen zhi*	成之聞之	40
GD12	*Zun de yi*	尊德義	39
GD13	*Xing zi ming chu*	性自命出	67
GD14	*Liu de*	六德	49
GD15	*Yu cong* 1	語叢一	112
GD16	*Yu cong* 2	語叢二	54
GD17	*Yu cong* 3	語叢三	72
GD18	*Yu cong* 4	語叢四	27
GD19	*Fragments*	(殘片)	27

Mawangdui manuscripts

ms. code	title of text	no. of cols.	columns
MWD1	no original titles		
	text 1 <*De* 德>	92	1–92
	text 2 <*Dao* 道>	77	93–169
	text 3 <*Wu xing* 五行>	182	170–351 (170–214 經, 215–351 說)
	text 4 <*Jiu zhu* 九主>	52	352-403
	text 5 <*Ming jun* 明君>	48	404–451
	text 6 <*De sheng* 德聖>	13	452–464
MWD2	original titles		
	text 1 *Jing fa* 經法	77	1–77b*
	text 2 *Shiliu jing* 十六經	65	78a–142b
	text 3 *Cheng* 稱	24½	143a–167a
	text 4 *Dao yuan* 道原	6½	168a-174a
	text 5 *De* 德	42½	175a–217a
	text 6 *Dao* 道	35	218a–252b

* While MWD1 is a silk scroll of 24 cm width, MWD2 is a folded silk cloth of 48 cm width, which is partly broken, partly cut along the fold, its fragments are thus divided into upper and lower halves, and the photographic reproductions follow this division; the upper halves of the column are here marked as "a", the lower halves as "b".

"text"

Text numbers have not been assigned for the Guodian manuscript, for each of the eighteen groups of slips published under one title is, according to the apparent logic of the publication, one manuscript or one text. For the sake of unambiguous reference, even GD3–4 are treated as two separate items, although they most probably originally were one manuscript (i.e. a group of slips bound together) containing six texts or textual units (cf. William G. Boltz, "The Fourth-Century B.C. Guodiann Manuscripts from Chuu and the Composition of the *Laotzyy*", in: *Journal of the American Oriental Society* 119.4 [1999], p. 595). Likewise, parts or all of GD11–14 may have been one manuscript.

"context"

This category was used only for examining some cases in which the choice of character or word, respectively, clearly depended on the context, e.g. the ways of writing *hòu* {後} or *hòu* {后} and *yú* {於} or *yú* {于}.

"column code"

Manuscript code + number of column in the respective manuscript:
E.g. "GD1.2" is the 2nd column in the *Laozi* A manuscript. "A" and "B" have been added to the columns code in cases where there are two clearly separate pieces of a column, even if they have only one excavation number. In the case of GD2.11, the excavation number 203 probably refers to the larger "GD2.11A" only, whereas "GD2.11B" does not have an excavation number. The two pieces of slip 13 of *Laozi* C fit so well together that they need not be separately numbered, yet they have different excavation numbers (591 for GD3.13A and 197 for GD3.13B) and are thus treated as two objects, whereas other slips that are broken but fit together without doubt and have only one excavation number are listed under one column code.

"exc. no." (excavation number)

Excavation numbers as listed on pp.223–230 of *Guodian Chu mu zhujian* 1998.

"first?" / "last?"

Indicates the first and last of the slips published together as one text or one manuscript, or in the case of the character database the first and last characters of one column.

"frg?" (fragmented)

Indicates whether a bamboo slip (in the columns table) or a character (in the character tables) is fragmented or otherwise damaged.

➔ **"f"** (for "fragmented") stands instead of or after the metadata pertaining to such columns or characters. E.g., "f" is given instead of the height of a character that is damaged in a way so that its height cannot be measured, or it is given after the number of bindings of a fragment of a bamboo slip to indicate that the original entire slip must have had more bindings, also an "f" marks the length of the slip when it was originally longer.

"length"

Indicates the length of the slip in the sense of the physical object (not in the sense of actual writing space) as represented in its photographic reproduction, the lacking reliability of which has been mentioned above (see above, "Quality of data").

"upper marg." / "lower marg." (upper / lower margin)

"Upper margin" indicates the space between the top end of a slip and the top of the first character of a preceding mark. "Lower margin" likewise indicates the space between the bottom end of a slip and the end of the last character or mark. In the case of manuscripts where there is a ruled line or similar means of marking off the writing space, the space between this marking and the edge of the writing support is considered the margin.

"wd." (width of column)

Indicates the average width of the bamboo slip as represented in its photographic reproduction, the lacking reliability of which has been mentioned above (see, "Quality of data"). In the case of silk or paper manuscripts or of slips or boards with more than one line of writing on them, the space between an imaginary vertical line running through the centre of one column of characters and the same line in the next column is considered the width of line. If the columns are marked by ruled lines, the distance between them is measured.

"bind." (number of bindings)

This information pertains to manuscripts made up of bound slips. The number of binding strings can be ascertained from traces of the decayed strings on the writing support or sometimes markings in black ink of the position of the binding string or notches carved into the slips to better fix the string.

"space between bindings"

In cases of manuscripts with three binding strings, the measures of the space between the first and second and that between the second and third bindings are separated by a slash. E.g. "72/64" in the entry for GD16.20 means that while the first and second binding strings of slip 20 of the manuscript *Yu cong* 2 were 72 millimetres apart, the distance between the second and the third was only 64 millimetres.

"distance to first binding"

The distance between the top end of the slip and the position of the first binding helps to determine the exact position of the binding on the object. The position of the following bindings can be determined by adding the measures of "space between bindings".

"shape"

Describes at present only the main difference in the shape of the Guodian bamboo slips, namely that between even/level ends (e.g. GD2.1 ) and tapered ends (e.g. GD1.3 ).

"char." (number of characters in column)

➔ To determine the average number of characters per column of a certain group of slips, it is necessary to first sort the data by the criterion "frg?" to extract the fragmented slips and exclude them from the calculation.

"Cui Renyi column no."

Indicates the different order and thus designation of the bamboo slips of GD1 through GD4 in: Cui Renyi崔仁義, *Jingmen Guodian Chu jian* Laozi *yanjiu*荊門郭店楚簡〈老子〉研究, Beijing: Kexue, 1998. Cui treats GD3+4 as one manuscript, beginning with GD4, and calls it *Laozi* A. Cui's *Laozi* C consists of the same slips as *Laozi* A of the standard edition, but in a different order of slips. The same is true of the *Laozi* B in both editions.

"character code"

Column code + number of character in the respective column:
e.g. "MWD1.094.07" is the seventh character in the 94th column in the manuscript MWD1 (i.e. the second column of its 'text 2', *Dao*); GD2.

➔ Users of the "Excel" Guodian characters file are advised not to delete the seemingly useless first column titled **"no."**. It contains the consecutive number of characters in this table and is necessary to restore their original order after having sorted them by other criteria. For the character codes in

the MWD files, a different number format was chosen. In these files, sorting by character code will restore the original order.

"height"

Indicates the height of a character.

"space"

Indicates the space after a character.

➔ To determine the average space between characters, it is necessary to first sort the data by the criterion "last" to exclude characters followed by a lower margin from the calculation.

"marks":

Indicates number and type of marks in a column. E.g. the entry "3d" for GD1.1 means that there are three black square marks o this slip. Especially the distinction between types b and c is not clear. In the character data, the marks are registered under the entry of the preceding character. The punctuation system of the Guodian manuscripts is not yet fully understood. The following descriptions are tentative.

	Description	examples
a	short double dash, often called "ditto mark", functions usually either as a repetition mark (*chongwen hao* 重文號) or a ligature mark (*hewen hao* 合文號)	GD1.12.15 (重文號)  GD4.12.13 (合文號) 
b	single short stroke, often occurs in the same function as the short double strokes, but usually seems to mark caesurae between clauses	GD1.6.16 
c	"bold", thick short stroke, marks caesurae	GD2.13.20 
d	small square, marks caesurae or ends of textual units	GD1.1.8 
e	broad stroke across the entire width of the slip, apparently marks end of a major textual unit, does not occur in GD1–4 and GD18	GD6.8.7  GD14.26.7 

f	so-called "tadpole mark" (*kedou hao* 蝌蚪號), apparently marks end of a major textual unit		GD1.32.22  GD1.39.9 
	internal abbreviation mark (long double dash [occasionally single dash], integrated into the character)	GD1.6.11  GD1.7.30  GD1.22.1 	GD18.8.8  GD2.2.3  GD1.1.14 

"freq. in ms." (frequency of word in manuscript)

Indicates how many characters that are supposed to write a certain word are actually present in the manuscript, regardless of what different characters that may be. Instances of the same word that are conjectured for missing parts of the manuscript are not included in the count.

"reproduction"

The images of the characters are scans with a resolution of 150 dpi. Always the full width of the columns has been cut out. In the many cases in which characters seem to be incomplete at one or both sides, this may be either due to decay of the original slips or the edges were cut off in the published reproductions (see above, "Quality of data"). The images include marks following the character. If the distance between character and mark is too large for the image to fit into the field. The mark is placed beside the character as a separate image (e.g. GD1.18.8).

"comp." (constituent components of the character)

This field does not aim to reflect the structure of the character adequately, but rather to register the constituent components separately in order to make them searchable. No judgement is implied about what must be considered the immediate components of a character and how it can be further analysed on the second or third levels (such as 箱 = 竹+相 [=木+目]). "Xp" indicates that only part of component "X" functions as a component of the character. For example the word *xiàng* {象} appears as a combination of the upper part of 象 with the 肉-component below (GD2.12.14 ), the components are thus given as "象p, 肉". This is not the same as cases of suspected abbreviation (case of 則, here it is not clear, to which degree the

character is abbreviated, thus the components that might be represented by the abbreviated form of writing are given regardless of how much of them is actually present in the character.
Generally, a slash (/) is used to indicate alternatives relating to the same item, whereas different items are divided by comma. Alternative forms are given to improve searchability. In the "structure" field, the entry "之/屮, 日, 又" means that the character  is composed of three components. The top one can be transcribed either as 之 or alternatively as 屮. Despite the preference for 之 according to my own transcription principles, both shall be searchable in the database for the purpose of comparing forms. Different components are divided by a comma.

"DT" (direct transcription)

The direct transcription aims to "translate" as faithfully as possible the forms of the original character into the equivalent forms of modern standard characters. Ideally, all structurally relevant graphic elements of the original character and their positions in relation to each other should be represented by the direct transcriptions. There are several factors that make direct transcription a highly complicated task: Most early Chinese manuscripts display a high degree of graphic variability. In several cases different graphic elements in the early script merged into one element in modern script and not less frequently an identical early graph developed into several different forms in modern script. The identification of graphic elements and their modern equivalents as well as determining the relative positions of the elements depend on subjective judgement and make unambiguous principles for a direct transcription virtually impossible. Nevertheless, developing such principles remains an important methodological desideratum.

"AMC" (analogous modern character)

In many cases, the more faithful the direct transcription of a character is the less it explains how it is to be understood in terms of modern writing and eventually read, because a direct transcription often results in a historically discontinuous character (for this term, see Takashima 2000), i.e. a character not attested in any traditional lexicon. The "analogous modern character" gives the historically continuous character with the closest resemblance to the direct transcription and thus helps to bridge the gap between transcription and reading.

"TP" / "RP" (transcription / reading in publication)

In these fields the transcriptions and readings in the standard publication are registered. An entry "才" under TP and "在" under RP means that the editors assume the original character, which is transcribed a graphic equivalent of modern 才, was used in the manuscript to write the word *zài* {在}. In the majority of Chinese manuscript transcriptions the conventional notation for

this is "才(在)". Many published transcriptions do not strictly distinguish between transcription and reading. In what the editors seem to consider obvious cases, the transcription is omitted and the reading is given straight away, e.g. when the characters 丌 and 亓 that are both frequently used to write the word *qí* {其} are directly rendered "其" without referring to the actual form of the character in the manuscript. A similar case is GD18.3.8, which is rendered "參" without mentioning that it is really written as 厽 in the manuscript. In doubtful cases, only a transcription is given and a reading is only mentioned in the notes: E.g., while for GD1.25.5 both transcription (𡈼) and reading (持) are given, GD1.37.27 is rendered "𡈼", and only in the notes a tentative reading *zhí* 殖 is suggested. In a few especially problematic cases, only a tracing of the character is given instead of a transcription and/or reading, e.g. GD18.11.15  , for which Li Ling (2002: 45–46) suggests the reading *shì* 世.

"AR" (alternative readings)

In this field, readings suggested by other scholars are mentioned. The absence of an entry does, of course, not imply that other readings have not been suggested.

"pronunciation"

In this field, the modern Mandarin Chinese pronunciation of the reading (i.e. the word that the manuscript character presumedly stands for) is given, basically according to the *pinyin* 拼音 standard. For technical reasons, the tones of the syllables had to be represented by numbers (e.g. yi1 instead of *yī*, yi2 for *yí*, yi3 for *yǐ* and yi4 for *yì*). As the category of pronunciation deals with words rather than characters, the reading comes first in this field. To improve searchability of the reading in relation to the transcription, the pronunciation of the character given as transcription is added after a comma.
➔ In fields dealing primarily with the word written by the character in question, the information relating to the word precedes that about the character. In fields dealing primarily with the character and only secondarily with its reading, the order is reversed. Information relating to these different levels is separated by a comma. Information relating to the same object in the sense of alternatives, are separated by a slash "/". The user of the database should bear this principle in mind, because for technical reasons it is not always possible to specify the character or word to which the information given in a certain field relates.
E.g., the entry "ru2, ru3/nü3" indicates that the character  is read as standing for the word *rú* {如} and is transcribed as 女, which can be read as either *rǔ* or *nǚ*. The entry for GD1.8.23  is "rong2, song4", indicating that the character is here read as *róng* {容}, although it is written with a

character structured like that used to write the word *sòng* {頌} in modern orthography.

"GSR" (series number in *Grammata Serica Recensa*)

The same order of information as in the field for the modern pronunciation is used also here (e.g. "1187a (容), 1190d (頌)"), because this category deals primarily with the word. Even though Karlgren's reconstructions of Old Chinese do not represent the state of the art, the GSR is an easily accessible reference, and the phonetic series as such are, apart from a few exceptions, still reliable.

"HDZ" (*Hanyu da zidian* 漢語大字典)

This field gives references to characters listed in the *Hanyu da zidian* that are relevant either for the manuscript character (its direct transcription and modern character analogy) or for the modern standard way of writing the word presumably represented by the manuscript character. (References are made to page number and number of character on this page, e.g. 2580.4 indicates that 矣 is the fourth character on page 2580.) The field "HDZ", which primarily deals with script, not language, has the reverse order of information as compared with the preceding two fields: e.g. "4362.3 (頌), 933.17 (容)".

"references"

Occasionally, references to mostly palaeographic works are given – as a mere convenience and without any claim to completeness.

Bibliographical References:

A/W 2000	Allan, Sarah; Crispin Williams (ed.). 2000. *The Guodian* Laozi*: Proceedings of the International Conference, Dartmouth College, May 1998*. (Early China Special Monograph Series 5). Berkeley: The Society for the Study of Early China.
Boltz 1990	Boltz, William G. 1990. "Three Footnotes on the *ting* 鼎 'Tripod'." *JAOS* 110.1: 1–8.
Boltz 2005	Boltz, William G. 2005. "Reading the Early *Laotzyy*." *Asiatische Studien / Études Asiatiques* LIX.1: 209–232.
Cui	Cui Renyi. 1998. *Jingmen Guodian Chu jian* Laozi *yanjiu*. Beijing: Kexue. 崔仁義 著《荊門郭店楚簡〈老子〉研究》 北京：科學.
CWB	Li Shoukui. 2003. *Chu wenzi bian*. Shanghai: Huadong shifan daxue. 李守奎 編著《楚文字編》上海：華東師範大學.
GCZ	*Guodian Chu mu zhujian*. 1998. Jingmen shi bowuguan (ed.). Beijing: Wenwu.《郭店楚墓竹簡》荊門市博物館 編；北京：文物.

GDLW	*Guodian Chu jian guoji xueshu yantaohui lunwen ji*. 2000. Wuhan daxue Zhongguo wenhua yanjiu yuan (ed.). (Renwen luncong). Wuhan: Hubei renmin chubanshe.《郭店楚簡國際學術研討會論文集》武漢大學中國文化研究院 編（人文論叢）武漢：湖北人民.
GZB	Chen Wei. 2003. *Guodian zhushu bieshi*. (Xin chu jian bo yanjiu congshu). Wuhan: Hubei jiaoyu. 陳偉《郭店竹書別釋》（新出簡帛研究叢書）武漢：湖北教育.
GSR	Karlgren, Bernhard. 1957. "Grammata Serica Recensa." *Bulletin of the Museum of Far Eastern Antiquities* 29: 1–332.
GWGL	*Gu wenzi gulin*. 1999–. Shanghai: Shanghai jiaoyu.《古文字詁林》上海：上海教育.
He	He Linyi. 1998. *Zhanguo guwen zidian: Zhanguo wenzi shengxi*. 2 vols. Beijing: Zhonghua. 何琳儀 著 《戰國古文字典：戰國文字聲系》北京：中華.
Li Ling	Li Ling. 2002. *Guodian Chu jian jiaoduji: zengding ben*. Beijing: Beijing daxue. 李零《郭店楚簡校讀記：增訂本》 北京：北京大學.
Liao	Liao Mingchun. 2003. *Guodian Chu jian Laozi jiaoshi*. Beijing: Qinghua daxue chubanshe. 廖名春 著 《郭店楚簡老子校釋》北京：清華大學出版社.
Liao 2003b	Liao Mingchun. 2003. *Chutu jian bo congkao.* (Xin chu jian bo yanjiu congshu). Wuhan: Hubei jiaoyu. 廖名春 著《出土簡帛叢考》（新出簡帛研究叢書）武漢：湖北教育.
Qiu	annotations marked "裘按" in ⇨GCZ
Takashima	Takashima Ken'ichi. 2000. "Towards a More Rigorous Methodology of Deciphering Oracle-Bone Inscriptions." *T'oung Pao* 86: 363–399.
XZG	Xu Zaiguo. 2002. *Liding guwen shuzheng*. Hefei: Anhui daxue. 徐在國 著《隸定古文疏證》合肥：安徽大學.