The Xixia Writing System

Bachelor of Arts Honours Thesis

Alan Downes
89253388

Supervisor: Professor Daniel Kane

Macquarie University

November 26, 2008

Abstract

The Xixia were an independent state established during the Song dynasty within the borders of present-day China. They developed a script that rivals Chinese in its complexity. The methods that have been historically used to describe characters are presented. A new method called the recursive radical method is then developed to exactly describe the structure of a Xixia character based on its components. This is then further developed into a transliteration method for the characters. Dictionaries are built up based on this transliteration method. A categorical dictionary for the Xixia characters is also presented. The grammar of the Xixia language is briefly outlined before finishing with an overview of the Xixia research literature.

Contents

1 Introduction **3**

2 Character Description Methods **7**
2.1 Radical-Based Systems . . . 7
2.2 Four Corners Method . . . 8
2.3 Nishida Tatsuo's Structural Approach . . . 12
2.4 Recursive Index for Xixia . . . 13
2.5 Recursive Index for Chinese . . . 20
2.6 Transliteration Scheme . . . 20

3 Xixia Dictionaries **26**
3.1 Categorical Dictionary . . . 26
3.2 Transliteration Dictionary . . . 26
3.3 Suffix Dictionary . . . 27
3.4 Alternative Reading Orders . . . 27

4 Xixia Grammar **28**

5 Literature Overview **31**
5.1 Language Requirements . . . 31
5.2 Fonts . . . 31
5.3 Dictionaries . . . 32
5.4 Buddhist Texts . . . 34
5.5 Translations of Chinese Texts . . . 34
5.6 Collections of Original Texts . . . 34
5.7 Histories . . . 35

6 Conclusions **36**

A Translation Project - *Wenhai* Article **37**
A.1 The Simple Xixia Characters . . . 39
A.2 The Main Character Construction Methods are Composites of Simpler Characters and Mixtures of Sound and Meaning . . . 41
A.3 Initial-Final Characters at the Top and Bottom is a Special Way of Creating Xixia Characters . . . 48
A.4 Indirect Sound-Meaning Combinations Reflect the Influence of Chinese Character Construction Methods. Long Sound Combinations Record the Long Sounds of Sanskrit . . . 52
A.5 The Exchange of Parts is One Method of Creating Characters of Similar Meaning . . . 53
A.6 The Meaning and Radicals of Characters Similar in Form . . . 57
A.7 The Flexibility of Shapes When Forming Xixia Characters . . . 61
A.8 Notes . . . 63

B Xixia Recursive Radical Codes **64**
B.1 Ordered by Transliteration 64
B.2 Original Order 66
B.3 Ordered by Type 68
B.4 Ordered by Stroke Count 70
B.5 Ordered by Frequency 72

C Chinese Recursive Radical Codes **74**
C.1 Original Order 74
C.2 Ordered by Stroke Count 76

D Software Outline **78**

Bibliography **80**

Chapter 1

Introduction

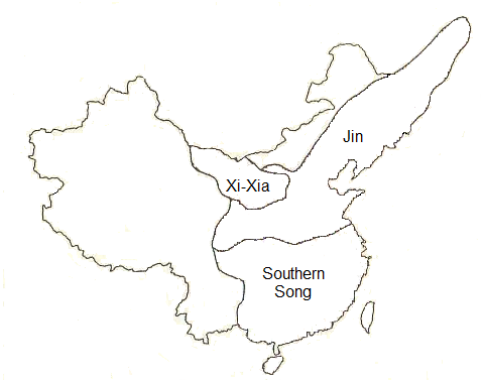
The Xixia or Tangut were also known as the Dangxiang (Chinese 党项). In their own language they called themselves Minia [26]. The earliest known reference to the Tangut name occurs in a Turkish monument of 735 AD [6, p.55]. The term is believed to be the Mongol word for the Xixia, coming from the term for the Chinese Tang 唐 dynasty appended with the Mongol plural suffix *-d* [19, pp.64–66]. They have been traced back to at least the Sui dynasty (581-618) [10, p.155]. A Chinese-Tangut army assisted the Chinese emperor during Huang Chao's rebellion of 875 AD, for which they were rewarded with military governorship of a province [10, p.163].

The Xixia state had been practically independent since as early as 1006 [10, p.177]. The ruler Li Yuanhao formally declared independence from the Song in 1038 [10, p.187]. There was much diplomatic maneuvering and warfare until the Song finally accepted this state of affairs in 1044 [10, pp.187-189]. The rest of the eleventh century and up to the end of the Northern Song in 1128 was characterised by sporadic warfare with the Song [10, pp.189-196]. With the expansion of the Jin there was no common border between the Xixia and the Southern Song - the rest of the twelfth century was relatively peaceful [10, p.197]. This period saw the growth of scholarly pursuits and Confucianism [10, pp.199-201]. The early thirteenth century saw a period of increasing instability until the destruction of the Xixia state at the hands of the Mongols in 1227 [10, p.206].

The maps below are intended to give a broad overview of the position of the Xixia state during the Northern and Southern Song dynasties. However, this was an unstable period within China's history and borders were constantly in flux (especially during the Northern Song period [10, p.155]). The position of present-day China is also shown.

Figure 1.1: Xixia During the Northern Song (adapted from [17, p.113])

Figure 1.2: Xixia During the Southern Song (adapted from [17, p.126])



The Xixia script was disseminated in 1036, under the reign of Li Yuanhao [12, p.118]. There are several conflicting accounts as to who actually created the script. One account says that it was created by Yeli Renrong (the emperor's chief cultural officer) [12, p.118]. It is possible that he may have been a Khitan, providing a potential link to the still partially undeciphered Khitan script [15, p.44], [53, p.2]. Indeed, there is addition of parts to and subtraction of parts from characters to form other characters, which is a feature of both the Khitan and Jurchen scripts [15, p.55]. However, it is also possible that he came from the Yeh-li clan, one of the clans comprising the Tangut state [10, p.160]. Kwanten [32] also takes issue with this link and believes it is more likely to have been created by Li Yuanhao's father, Li Deming, who was of a more scholarly disposition. The issue is still unresolved.

Despite the devastation with the destruction of the Xixia state by the Mongols in 1227, some of the Xixia people survived and escaped to form smaller communities [10, p.214]. The script was in use until the end of the Ming dynasty [10, p.214] and was forgotten during the Qing. In 1804 Zhang Shu discovered a bilingual Xixia - Chinese stone tablet in a temple in Liangzhou [42, p.16]. By comparison with the above tablet in analysis published in 1827, Liu Qingyuan was able to determine that some coins with inscriptions that were originally identified as Sanskrit were actually in the Xixia script [62, p.9]. The major breakthrough in the script's rediscovery came in 1908. As outlined in [42, pp.17-18], the Russian explorer P.K.Kozlov discovered a large amount of material in some collapsed temples in the ruins of the Xixia city Khara-Khoto (also known as the *Black Water City* or *Heishui Cheng* 黑水城). He returned a year later and found more material in the inner room of a pagoda. Some ceramic jars containing Buddhist texts in the Xixia script were discovered in Ningxia in 1917 [42, p.18].

The historical background is not the focus of this thesis, so this overview is necessarily brief. For further information on this area, consult the references listed in section 5.7.

Xixia characters are more complex than their Chinese equivalents. They contain many more strokes and the variability in the shapes making up the characters is smaller, which makes them harder to remember. The question remains as to how the Xixia literati remembered so many complex characters. Was it just a feat of rote learning (making it perhaps the most complex written language ever produced [38, p.4]), or is there some structure underlying the script that we have yet to uncover? The large amount of written material uncovered in the script provides some evidence that it was fairly easy to use [31, p.4].

Beyond the shape of the strokes themselves, there is little evidence of borrowing of characters from Chinese. Grinstead [15, p.56] gives the character for a door 𗥃 (M4794 - numbers prefixed

by an *M* refer to numbering in the Mojikyo font or Li Fanwen's dictionary [41]) as the only entire character directly borrowed from Chinese (namely the Chinese character 門 *men*). Whilst another can also be seen (the weigh character (M909) appears to be related to the Chinese character 兩 *liang*), such correspondences are extremely rare.

On the surface, it would appear that the Xixia script would be harder to decipher than the Khitan and Jurchen scripts of the same period, which are closer in form to Chinese. However, there is a lot more source material available for the Xixia - over the last hundred years the script has pretty much been deciphered.

The spoken language underlying the script has been lost and there is still not full agreement as to what this language was. Some scholars believe the script actually represents two languages - the language used by the common people and a ritual language [27].

Although a set of all-encompassing principles for the construction of Xixia characters still eludes scholars, there are tantalising hints that such a structure might exist. Some of these interesting relationships between characters are shown below:

swallow (M4540) and saliva (M4545) are related to lips (M4543).
gold (M152) is blue (M153) with an extra piece in the bottom right-hand corner.
fish (M3057) is water (M3058) with the bottom right-hand corner components reversed.
hand (M3485) is finger (M3484) with an extra stroke. There is a horizontally reversed version (M5298) that also means finger.

love (M4973) looks like two people (M1886) under a roof. marriage (M5051) is also similar.
celestial being (M4860) looks like a person (M1886) with wings.
fear (M4636) is thunderbolt (M4635) plus an extra stroke.
concave (M950) and collapse (M951) are almost the same.

die (M2003) is stuffy (M2002) plus an extra stroke.
between (M5747) is pair (M5757) plus an extra dividing stroke between the two components.
hat (M2275) is person (M1186) plus a brain (M124).
equal (M424) and equate (M1357) are symmetrical.

ride (M2407) is man (M2541) on a horse (M764).
onion (M2278) is chilli (M512) horizontally reversed (curly tails tend to be straightened when not on the right-hand side).
pine (M4391) is wind (M2302) with the tree radical on top.
fire (M4408) is cook (M439) with the tree radical on top.

A few characters look like they could be logograms:
plough (M1425) looks like the prongs on a plough.
weigh (M909) looks like a set of scales (potentially borrowed from Chinese).
door (M4794) looks like a door (potentially borrowed from Chinese).

In chapter two a variety of methods for describing characters (both Chinese and Xixia) are discussed, culminating in the development of a method for describing entire characters. A translit-

eration scheme is then developed based on this description method. In chapter three a categorical dictionary and various transliteration dictionaries are built on top of this transliteration scheme. These dictionaries are available in the dictionaries supplement to this thesis. In chapter four the grammar of the language represented by Xixia characters is outlined. Finally, in chapter five an overview is given of the Xixia research literature.

Chapter 2

Character Description Methods

This chapter examines methods used for describing and indexing characters. It starts with looking at radical-based systems in Chinese and Tangut before moving onto the four corners method. Nishida Tatsuo's structural approach to the Tangut characters is then outlined before development of a recursive radical index for the characters that can be used for efficient searching. This recursive radical scheme is then further developed into a transliteration scheme for the Xixia characters that can be used for searching or as a mnemonic for remembering the structure of individual characters.

2.1 Radical-Based Systems

Although the Chinese writing system looks very complicated to an outsider, it contains a number of components known as radicals which indicate meaning. For example, the basic insect character 虫 is contained in the specialised characters for various kinds of insects, such as in *grasshopper* below:

The Xixia writing system also contains a number of radical components, perhaps reflecting a Chinese influence on the original inventors. However, the influence of meaningful radicals within the Xixia writing system is far less obvious than in the case of Chinese. For convenience I have continued to use the term *radical* to describe character components in Xixia below, but these components might not actually have any underlying meaning.

There are a large number of characters for various types of grass. This reflects the strong interests of the Tangut people in stockbreeding [10, p.157]. These characters tend to have the common component or *grass radical* shown in red below (M2421):

The radical in the top-left hand corner of M1856 below appears to be associated with characters for shape and size:

In a similar manner to which the Chinese character for weighing (simplified 两 or traditional 兩) is derived from a pair of scales, the Xixia character meaning weigh (M909) also resembles a pair of scales. This has been carried through to other characters. The character for a small scale for weighing precious metals or medicine (M2097) is half of the original . Various characters meaning heavy also contain a portion of (M739, M1435,  M1461,  M3185 and  M5945).

The character for poison  (M8) is contained within the characters for scorpion (M2106) and snake (M80),  (M4030). The character for trees of the Simarubaceae family  (M4099) also contain the poison character. One member of this family, the *Ailanthus*, has been described as "an usurper in rather bad odor at home, which has come over to this land of liberty, under the garb of utility to make foul the air, with its pestilent breath ..." [60, p.24]

As Table 2.1 below shows there have been many attempts to set up radical systems for Xixia characters with wildly varying numbers of radicals.

Table 2.1: Xixia Radical Systems (adapted from [13])

Year	Author	Results
1914	Luo Fuchang	23 types of radicals
1915	Luo Fucheng	164 radicals
1919	Bernhardi und Zach	12 types of radicals
1960	Nevsky	12 types of radicals
1961	Nishida Tatsuo	348 character components, 13 character structures and 162 components with meaning
1964	Kychanov	8 types of strokes, 83 types of strokes, 650 sides, 20 combination patterns and 68 radicals
1966	Nishida Tatsuo	136 radicals with meanings
1968	Sofronov	68 radicals, 368 sides
1972	Grinstead [15]	107 sides
1983	Shi Jinbo, Bai Bin, Huang Zhenhua [63]	444 radicals
1986	Li Fanwen	385 radicals
1989	Huang Zhenhua, Nie Hongyin, Shi Jinbo	264 radicals
1997	Li Fanwen [41]	357 radicals

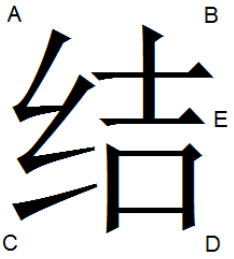
2.2 Four Corners Method

The four corners method is an attempt to describe characters via a single numerical code. The discussion of the application of this method to Chinese below relies on [1, pp.104–108].

As the name suggests, the four corners method takes its four numbers from the shapes at each of the four corners of a Chinese character. The corners are referenced in the order top left, top

right, bottom left and bottom right. In order to aid lookup, a fifth number can also be added. This is the shape directly above the fourth corner if it hasn't been already referenced elsewhere (otherwise it is set to zero). This results in the five digit code $ABCD_E$, as shown below:

Figure 2.1: Four Corners Method - Chinese Descriptor Positions



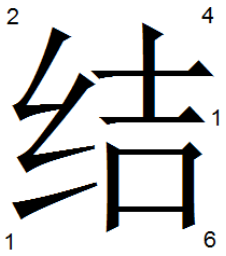
The mapping between shapes and digits in the Chinese four corners method is as follows:

Table 2.2: Chinese Four Corners Method [1, p.104])

Number	Shapes
0	亠
1	一
2	丨
3	丶
4	十
5	丰
6	口
7	冂
8	八　人
9	小

For our example in Figure 2.1 above, the upper left corner is a downward-sloping stroke so the first digit is *2*. The upper right corner has a cross shape, so the second digit is *4*. The lower left corner has an almost horizontal line, so the third digit is *1*. The bottom right corner has a square shape, so the fourth digit is *6*. Above this square is a horizontal line, so the fifth digit is *1*. This results in the code 2416_1 as shown below:

Figure 2.2: Chinese Four-Corners Code 2416_1



There are also a few more specialised rules:
1. If a shape at a corner has already been described by a preceding corner, this number is set to *0*. This reaches an extreme in the Chinese character for *one* 一. In this case there is only one shape (a horizontal line) and the four-corners code is *1000*.
2. If a corner is missing, as in the bottom-left corner of the character for *fly* 飞, this digit is set to *0* (four corners code of *1201*).
3. If the character encloses shapes in an interior, such as 田 and 開, then the third and fourth digits refer to the bottom of the interior rather than the bottom of the enclosure (codes are *6040* and *7744* respectively).

A major advantage of the four corners method is that once it has been mastered it enables one step to be cut out of dictionary lookup:

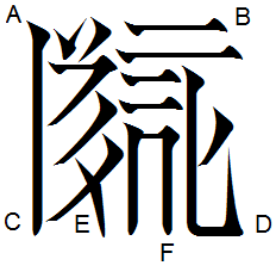
Using the radical approach, lookup of an unknown character in a Chinese dictionary consists of:
1. Lookup the radical number for the radical in the character of interest.
2. Count the number of strokes in the character and find the character in the list under the radical number found in step (1).
3. Go to the character's entry in the main dictionary, using the information found in step (2).

Under the four corners method, the steps consist of:
1. Determine the four corners number and look up the character in the list for this code. These lists are generally much shorter than those using the radical approach.
2. Go to the character's entry in the main dictionary.

Despite these advantages, the four corners method is nowhere near as popular as the radical approach and doesn't even appear in some dictionaries.

Li Fanwen has developed the four corners method for Xixia characters in his dictionary [41]. He develops a six-letter code, denoted here by *ABCDEF*, as a numerical descriptor for each character where each of the six digits describes the shape of the character in the position shown.

Figure 2.3: Four Corners Method - Xixia Descriptor Positions



The first four digits describe the shape of the top-left corner, top-right corner, bottom-left corner and bottom-right corner respectively. The fifth and sixth digits describe the shape of the interior of the bottom left and the interior of the bottom right.

The mapping between shapes and numerical digits is described in the table below:

Table 2.3: Xixia Four Corners Method (adapted from [41, p.1 of section on Four Corner's Method])

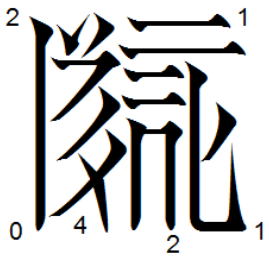
Number	Shapes
1	
2	
3	
4	
5	
6	
7	
8	
9	

It can be seen from the above that Li Fanwen differs from a traditional Chinese four corners allocation method in his use of digits *0* and *6*.

If a preceding number has already described a character component due to that component taking up a side of the character, the second number describing that component is set to zero. If the top side of a character is taken up with a horizontal line 一 as in 亘 , the first digit will be *1* and the second digit will be *0*. If the character has a vertical line on the left-hand side as in 豌, the first digit will be *2* and the third digit will be *0*. If there are no further components in the middle of the bottom of the character, or if the bottom of the character consists of a horizontal line, then the fifth and sixth digits are set to zero.

For our example in Figure 2.3 above, the upper left-hand corner consists of a vertical line, so the first digit is *2*. The upper right-hand corner consists of a horizontal line, so the second digit is *1*. The lower left-hand corner is also a vertical line, but as this line has already been described by the first digit, the third digit is set to *0*. The lower right-hand corner has a horizontal line that kicks up at the end, also represented by a *1*. On the inside of the bottom-left corner there is a cross, so the fifth digit is a *4*. The inside of the bottom-right corner contains a vertical line, so the sixth digit is a *2*. So the final four-corners code is *210142*, as shown in Figure 2.4 below.

Figure 2.4: Xixia Four Corners Code 210142

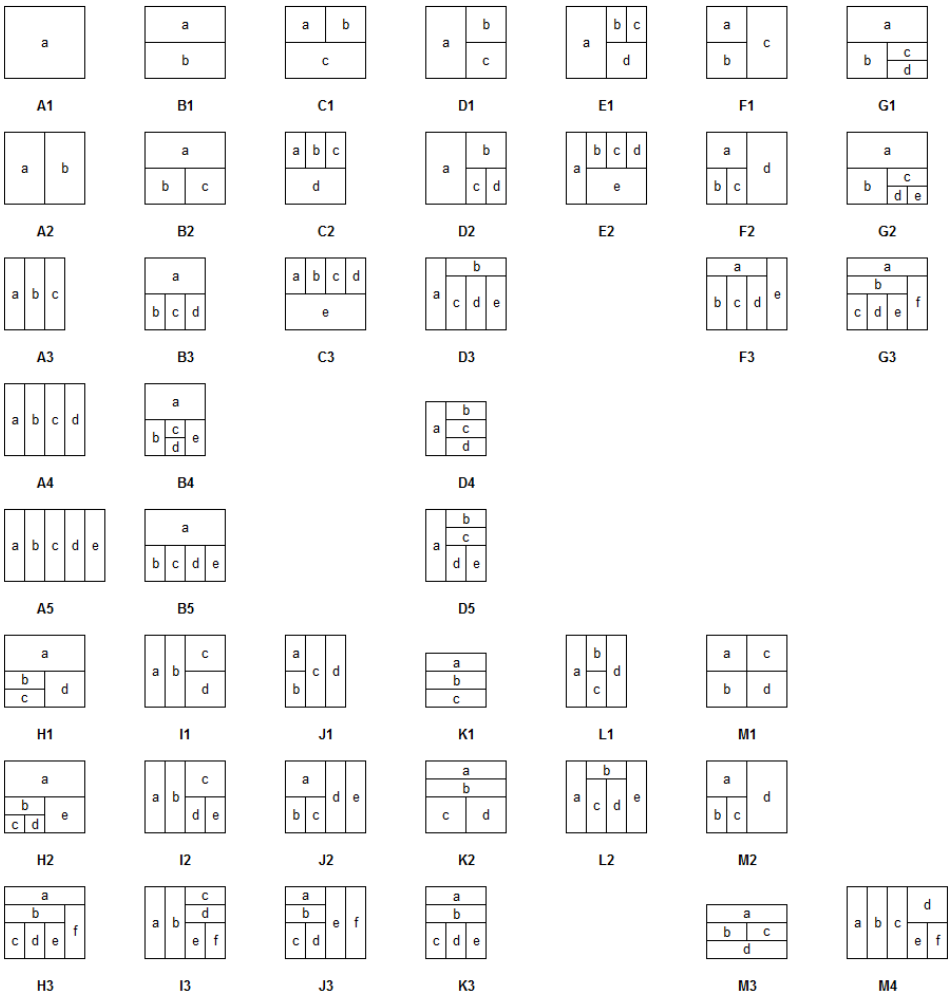


Li Fanwen indexes his dictionary using the four corners method, so the first entries in the dictionary all have 一 on the top of the character as this gives the four corners code *10xxxx*.

2.3 Nishida Tatsuo's Structural Approach

Xixia characters are quite recursive in nature. For example a combination that could be at the top level in 𗅋 (M2464) may be further into the character structure in a character like 𗥃 (M4998). Most of the characters consist of a structure containing rectangular pieces, a fact used by Nishida Tatsuo in [55, p.246]. Nishida classified the characters according to how they broke into rectangular pieces as shown below in Figure 2.5.

Figure 2.5: Nishida Tatsuo's Structural Approach [55, p.246]



Nishida gives a list of the character components contained at each position in the structure in his dictionary [55, pp.303–507].

Nishida's method is an improvement over radicals alone as it attempts to give some idea of the character structure. However, matching up a character with a particular structure requires the user to memorise the 319 radicals - the only reason that 豕 (M100) has structure B1 rather than B2 is that the bottom section under the horizontal stroke is a radical (no. 205).

2.4 Recursive Index for Xixia

The recursive index developed in this thesis takes the above system a step further by directly specifying the structure using a series of four simple rules along with a list of radicals:
1. Enclose a horizontal structure with brackets [...]
2. Enclose a vertical structure with braces {...}
3. Enclose a structure inside another structure with parentheses (...).
Note that this particular rule is not used in Xixia characters, as they don't have enclosures. One

reason given is that it allows demons to escape - they can't be trapped inside the Xixia character [15, p.57]. Enclosures do, however, play a role in applying the recursive index to Chinese characters (outlined in the next section).
4. Separate radical numbers with commas.

The radical numbers used in the index are listed in Appendix B, starting on page 64. The radicals are grouped into colour-coded types based on their shapes - whether they have a basic horizontal or vertical shape, whether the protrusions at the bottom of the character are straight ("straight legs"), whether there is curly tail on the bottom right-hand side ("curly tail"), whether there is a curly tail on the right-hand side that supports additional character components ("tail support") or whether there is a protrusion on the top right-hand side of the character ("head"). Any character that doesn't fit into one of these categories is treated as unclassified (left coloured black). Besides a listing in the original order, Appendix B also includes listings by type of radical, number of strokes and frequency of use to make it easy to find any particular radical.

As an example consider the character for milk 𗤓 (M53), analysed in Figure 2.6. The character as a whole has a vertical structure - it is not possible to subdivide the character using vertical divisions, horizontal divisions have to be used. As a result, using rule two above, this character's recursive code will start with a "{". There are three components in the vertical structure - a horizontal stroke (radical code 1), a horizontal stroke with three cross-strokes (radical code 3) and the rest of the character. The recursive code is now "{1,3". The bottom part of the character is a horizontal structure, so by rule 1 the next part of the code needs to start with a "[". The horizontal structure has two components - the left-hand one is radical code 40. The recursive code is now "{1,3,[40".

The structure in the bottom right-hand corner is another vertical structure, so this part of the code starts with another "{". The top part of this is radical code 11 and the second part is another horizontal line (radical code 1). The recursive code is now "{1,3,[40,{11,1". The third part of this vertical structure is another horizontal substructure, so another "[" needs to be added to the code. The horizontal substructure consists of three components with radical codes 14, 17 and 14. The recursive code is now "{1,3,[40,{11,1,[14,17,14". Finally, the braces and brackets that have been opened earlier need to be closed off to give the final recursive code of "{1,3,[40,{11,1,[14,17,14]}]}".

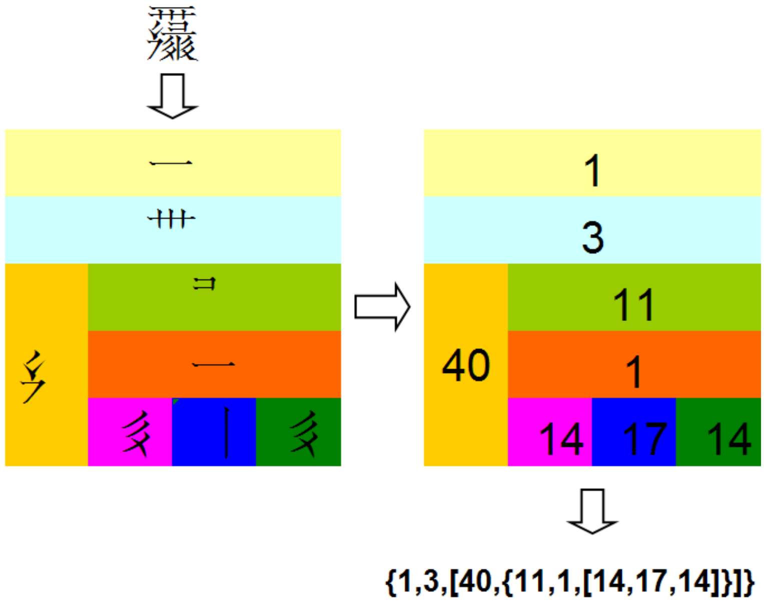
When using the software there are two methods of looking up characters. It is possible to directly specify the character by turning on the "Code Segment" checkbox and specifying the recursive code as {1,3,[40,{11,1,[14,17,14]}]}. This might look a bit daunting but once the radical list has been committed to memory it is actually quite easy and gives the desired character immediately.

The other option is to turn off the "Code Segment" checkbox and specify the components that make up the character in no particular order, together with their multiplicities - this method specifies the components making up the character but ignores the character's structure. The radical number comes first, followed by a comma and then the multiplicity. Different radicals are separated by a semi-colon. In the above example searching on 1,2;3,1;11,1;14,2;17,1;40,1 also returns only the desired character. However, this method is more likely to return additional characters (especially if the radical list is smaller) as the structure of the character is being ignored.

It is also possible to not specify the multiplicity and just specify radical numbers - if all the radicals in the list are present in a character it will be returned. Searching on 1;3;11;14;17;40 also only returns the desired character.

The recursive index can also be used to search for combinations of radicals. Consider, for example, trying to find the combination 𗥃 (M2464, meaning to milk an animal), coded as

Figure 2.6: The Recursive Index for Character M53 [illegible]



[14,17,14]. This could be at any level of recursion in the character. Searching on radicals is going to be very inefficient - 14 is one of the most common radicals in Xixia characters, appearing in 1591 characters (over a quarter of the total). There is also no guarantee it is going to be chosen as the radical for any given character. The four-corners method will also not be very useful - if this combination is stored in deeper levels of recursion the edges of it won't be on corners. However, using the recursive index all characters with this combination are picked up easily, shown in Table 2.4.

Table 2.4: Characters Containing the Combination [illegible]

Number	Character	Meaning
53	[illegible]	milk
373	[illegible]	mating
1265	[illegible]	butter, cheese
1338	[illegible]	love, like
1392	[illegible]	affection, love
1473	[illegible]	preposition
1532	[illegible]	extol, eulogise
2397	[illegible]	bed
2464	[illegible]	milk (a cow, etc.)
2465	[illegible]	thirdly
2963	[illegible]	like, love, desire
3065	[illegible]	milk
3342	[illegible]	surname
3883	[illegible]	power, might
3912	[illegible]	step, pace

Table 2.4: Characters Containing the Combination 衆

Number	Character	Meaning
4344	蒜	three
4614	嬸	nurse, breast-feed
4834	孈	milk
4973	蒜	love
4981	蠢	perhaps
4998	孈	surname
5518	龖	pudenda
5668	龖	sweet smell, flavour

The radical list has been constructed by starting at Mojikyo character no. 1 and working through to character no. 6000, adding components as I needed them. To keep the radical list manageably small I also avoided including most radicals that can be broken up into smaller pieces. A few of these have been kept for convenience, eg. radical no. 14  .

The recursive index provides a fast and precise method for character lookup. In the full list of 6000 characters there are only 72 characters which have a duplicate recursive code, and in each of these cases the code is only shared with one other character. Depending on the character being searched for, other methods (including a combination of four corners with radical selection) can sometimes still result in a list of one hundred or more characters that need to be examined. Using the recursive index it is usually only necessary to enter a portion of the character code to bring up a small list from which the desired character can be selected. The method is also convenient in terms of searching for what particular combinations of character components might mean.

The recursive index allows for characters of arbitrary recursive depth. This would allow the Xixia character system to be used for cryptography, encoding a recursive structure into a picture that would be meaningless without a key as to what the components mean. It could also be used a mnemonic device. It is well known that the human brain easily remembers items in a list by associating each item with a physical location, also known as the loci method [65, p.62]. Xixia characters could be used for remembering branching or recursive structures. Perhaps this is a clue as to how the Xixia characters were originally constructed. This is reminiscent of Wilkins' work on artificial languages [72].

There have been a few other methods developed to precisely describe the whole character in connection with Chinese. These have provided the inspiration for development of the recursive radical method.

HanGlyph [73] is a system that can be recursively used to build up complicated characters from a set of primitive strokes and a set of operators applied to these strokes [73, p.1004]. It allows for macros to be created that define character components - these can then be reused to create other characters [73, p.1005]. Using HanGlyph as a base would enable the Xixia radical list to be shrunk even further as it provides for intersections between components. However, there is a tradeoff between a smaller list of components and the effort needed to specify the character - specifying each stroke would result in a huge code for each character. Whilst good for character construction it would become unwieldy as a character lookup method. There doesn't appear to be a publicly-available implementation of HanGlyph [58, p.6].

CDL [3, 4] is an XML-based Chinese character description language. XML is like the familiar HTML format seen in web pages, but it is extended to allow for any data type. It describes

characters in terms of the positioning of components on a 128 x 128 grid [3, p.3]. CDL is better suited for font development as it gives precise information on positioning and size of character components. However, this information is not needed if all you want to do is search for a character. As a result of this size and positioning information it is also difficult to compare components in searching using CDL [58, p.5]. It has also been proposed to use CDL to assist with development of the Unicode standard for Chinese characters [3, p.6]. Perhaps a computer program could be written to convert the Xixia recursive codes developed here into CDL descriptors, which would be quite useful for font development.

As described in [58, pp.8–10], SCML is also an XML-based language that is built up in terms of strokes, anchors, layouts and locations. Anchors define the intersection points between the set of basic strokes. Layouts can be expressed as 'row', 'column' or 'free', similar to the '[' and '{' operators in the recursive radicals developed here. Locations are defined in terms of intersections between regions surrounding individual strokes. SCML also uses a fifth construct, called *components*, to describe collections of strokes that are connected to each other. SCML was designed to automatically check for duplicates in standards such as Unicode and to allow for automatic generation of stroke count and order [58, p.3]. SCML doesn't allow for size-relationships between characters, eg. □ and 囗 [58, p.4]. SCML also doesn't appear to be publicly available.

Nishida Tatsuo's structural approach [55, p.246] is the closest pre-existing work on Xixia to the recursive index developed here. Figure 2.7 maps Nishida's structures to recursive codes. Note that the radicals labelled *a* through to *f* might actually be a combination of the recursive radicals rather than a single radical.

Which parts are joined together as a "radical" is sometimes fairly arbitrary. I have tried to keep the number of radicals to a minimum. They could be even further reduced, as some are combinations of others, notably:

Table 2.5: Radicals That Could Be Removed

Radical	Part 1	Part 2	Part 3	Part 4
12	11	17		
13	6	7		
14	6	6	7	
15	8	9		
16	2	9		
20	9	6	17	17
22	1	6		
29	9	9	1	43
30	1	8	9	
31	9	9		
33	6	6	17	
38	22	22		
39	22	22	22	
40	41	22		
46	9	6	17	
48	1	43		
50	9	6		
53	6	6	7	
54	6	7	7	
60	9	9	1	17

Table 2.5: Radicals That Could Be Removed

Radical	Part 1	Part 2	Part 3	Part 4
63	6	6	62	
64	21	9		
67	5	6	6	7
75	17	17	17	
76	17	17	17	
78	17	17		
81	1	1	17	17
83	5	17		
84	5	17		
93	5	6	7	7
95	6	7		
97	17	17	1	1
112	6	7	9	
117	5	1	17	
118	5	6	7	
128	21	6	6	7
130	5	6	6	
137	9	1		
144	9	9	9	
154	6	7	9	9
157	1	17		
163	5	22	22	
173	9	17	9	

However it is convenient to keep these extra radicals around as it reduces the number of keystrokes required when searching. If a minimal number of radicals was used, the recursive code for the character 鬱 would change from the relatively compact [{22,7},{31,22,13},39] to the unwieldy [{1,6,7},{[9,9],1,6,6,7},{1,6,1,6,1,6}].

It is also worth noting how special cases that could be coded in several ways have been treated:

* The middle of 燚 is taken to be radical 33 氵 despite the fact that it overshadows the right-hand side. This gives the code [{22,13},33,{22,7}].

* The long diagonal stroke in the top-left hand side of 爨 is still treated as part of radical 28 ⺤ to give the code [{28,13},33,{22,7}]

* The top left-hand component of 厥 is treated as radical 35 厂. This gives a final code of [{35,7},30].

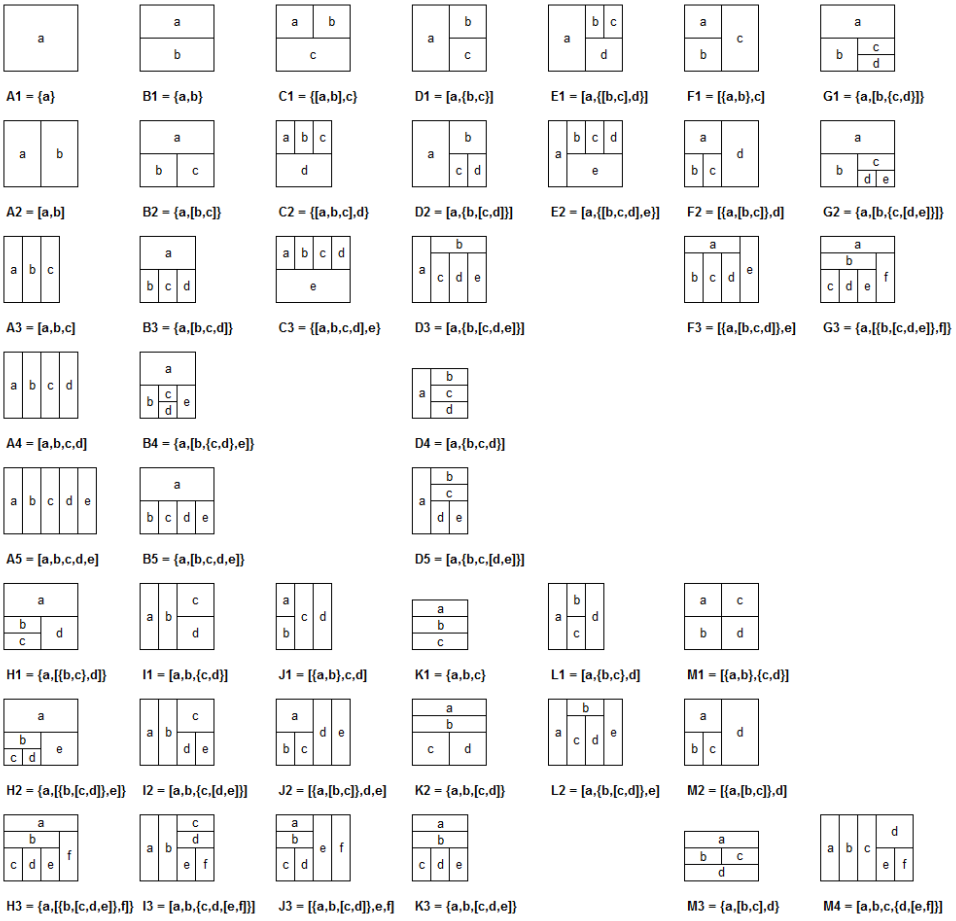
* The top of the middle section in 鱻 is treated as a separate radical despite the fact that it is connected to part of the bottom of the middle section. This gives the code [{22,7},{24,[41,104]},14].

* Where a long tail supports another radical, as in 𢍏, radical placement is encoded as though the long tail doesn't exist. The code for this example is [{1,1,65},14] - no account is taken of the fact that radical 14 彡 lies above the tail of radical 65 乚.

* Radical 12 冂 is treated as vertically above radical 13 犭 in character 𤴓. This gives the code {11,[10,{12,13}],{1,57}]}.

* Where there is a complicated arrangement above a tail, as in 齾, the section above the tail is treated as a block when considered in relation to the rest of the character. This results in the code [{50,101},[{31,37},{1,1,89}]] - this is the only time when two closing]'s will appear next to each other.

Figure 2.7: Nishida's Structures with Recursive Radicals



2.5 Recursive Index for Chinese

The recursive index has also been used to encode 7000 Chinese characters. The radical list in the Xinhua Character Dictionary [1] was expanded as new shapes were encountered. The final list of radicals is shown in Appendix C, starting on page 74. These are shown both in the original order and ordered by stroke count. The following differences with Xixia were found:

a) As Table 2.6 below shows, Xixia characters generally have more radicals than Chinese characters. 41% of Xixia characters have six or more radicals, but only 2.9% of Chinese characters are in this group. Although the radical count of Chinese characters drops off quickly, it also has a longer tail. Of the 7000 Chinese characters analysed, there are two with more than eleven radicals, namely 鬣 (13 radicals) and 齾 (15 radicals). There are no Xixia characters with more than eleven radicals.

Table 2.6: Comparison of Xixia and Chinese Radical Counts

Min. No. Radicals	Xixia Chars	Xixia Pct	Chin Chars	Chin Pct
1	6000	100.0%	7000	100.0%
2	5995	99.9%	6732	96.2%
3	5874	97.9%	4812	68.7%
4	5285	88.1%	2233	31.9%
5	4043	67.4%	781	11.2%
6	2464	41.1%	206	2.9%
7	1160	19.3%	66	0.9%
8	407	6.8%	23	0.3%
9	110	1.8%	12	0.2%
10	22	0.4%	5	0.1%
11	2	0.0%	3	0.0%
12	0	0.0%	2	0.0%
13	0	0.0%	2	0.0%
14	0	0.0%	1	0.0%
15	0	0.0%	1	0.0%
16	0	0.0%	0	0.0%

b) There is a lot more variation in the shapes used at each level in the structure in Chinese when compared to Xixia [15, p.53]. In encoding 7000 Chinese characters 420 different shapes were needed, compared to the 176 shapes needed to encode the 6000 Xixia characters. Obviously there is some degree of subjectivity in selection of the components, such as to what degree individual components are defined rather than being specified by their subcomponents. However, the difference is so big that it would still persist unless a significantly different approach was taken to radical selection. The Xixia writing system, which on the surface appears to be more complicated than Chinese, is actually simpler when examined in this manner.

2.6 Transliteration Scheme

There already exist transliteration schemes based on what is known about Xixia phonetics, such as that used in Gong [14]. The transliteration scheme developed in this thesis is designed to act as a mnemonic to aid in remembering the structure of individual characters - it has no relation to how the language behind the Xixia script may once have sounded.

The approach loosely follows that used by Vaccari in their Japanese dictionary [69], where characters are assigned letters in the Roman alphabet based on their shape. The vowels haven't been assigned shapes as they are used for a different purpose, as explained below. The following

gives a detailed rationale for the letters that were chosen - for an overall summary look at the table in Appendix B.1.

The letter *b* represents anything that has a basic L-shape with horizontal strokes emanating from the vertical of the L. It includes  ,  ,  , ,  and  .

The letter *c* represents anything that has a *c* shape, including rotations of *c*. It includes  , ,  ,  , ,  ,  ,  and  .

The letter *d* maps to radicals consisting of dots (the first letter of the word *dot* is *d*). It includes  , and  .

The letter *f* maps to anything that contains a top and has straight legs. This includes  ,  ,  ,  ,  , ,  ,  , ,  ,  ,  and  .

The letter *g* represents anything that has no top and a curly tail to the right-hand side (reflecting the shape of *G*). It includes  ,  ,  ,  , ,  ,  ,  ,  , , , , , , , , , and .

The letter *h* represents anything that has straight legs, no top and horizontal strokes that at least intersect with (and often cross) the vertical legs. It includes , , , , , , , , , (included as looks like a rotated *H*), , , and .

The letter *j* maps solely to the component . This looks like a dot with a curve underneath it.

The letter *k* maps solely to the component . The top part of this character (ignoring the dot on the left-hand side) looks like a *K*.

The letter *l* maps to radicals that look like two straight lines perpendicular to each other (reflecting the *L* shape). These include , , , and . It also maps to radicals that look like a single vertical stroke (reflecting the *l* shape). These include , , and .

The letter *m* represents radicals that contain a long horizontal stroke possibly intersected by shorter vertical strokes (if the horizontal stroke of was moved to the top you would have a character that looks like *m*). It includes , , , , , , , , and .

The letter *n* consists of components that have a basic lower-case *n* shape, with the interior of the *n* possibly filled by other strokes that don't intersect the vertical sides. It includes , , , , , , , , and .

The letter *p* is used by components consisting of a long vertical with something on top. These include and . It also incorporates a mixed bag of radicals that wouldn't fit anywhere else - radicals that are made up of multiple horizontal and vertical straight strokes along with characters that look like *E*. These include , , , , and .

The letter *q* is used by components that pretty much fill out the whole space and have a long tail on the bottom right-hand side (reflecting the shape of *Q*). It includes , and .

The letter *r* maps to radicals that have a head and a curly tail on the bottom right-hand side. It includes , , and .

The letter *s* is used by components that have a switch-back shape. It includes , and .

The letter *t* maps to components that have a long vertical crossed by shorter horizontals (some of these are ninety degree rotations of radicals mapped by the letter *m*). It includes  , ,  ,  ,  , ,  ,  , ,  ,  , , ,  , , , , and  .

The letter *v* maps to components where two legs form a basic *v* shape and a couple of cases where the dots at the top of the component could be viewed as a *v*. It includes , , , and .

The letter *w* represents components consisting of two or more vertical legs. It also contains radicals that could easily be further split into two separate components (the word "double" sounds like the name of the letter *w*). It includes  ,  , , , ,  , and .

The letter *x* is used for radicals that have a cross shape, possibly with additional strokes, but excluding those with a top matched to *n* such as . It includes ,  ,  , , , ,  , , , , , , and .

The letter *y* is used for the radicals and . This mapping is a bit more of a stretch of the imagination, but if you look at the first one you can almost see the top of a *y*.

Finally, the letter *z* maps to *Z*-shaped radicals such as , , , , and .

As an aid to remembering the structure of characters, it is possible to pronounce the words produced by this artificial transliteration language. The sound values of the letters are basically the same as English. However, the English letter x can sometimes be difficult to pronounce - this has been replaced by the sound *sh*. In order to clearly distinguish between the letters c and s, the sound of the letter c has also been replaced with a *ch* sound. The order of the letters in a transliteration is the same as the corresponding order of the radical numbers if the recursive radical method was used.

The transliteration scheme can also be used to look up characters rather than using the recursive radicals above. This has the advantage of being quicker to input and there is no longer the need to distinguish between similar radicals, such as and . It is not necessary to go any further than this transliteration scheme for searching as 89.4% of the transliterated words map to unique characters and in the worst case *xxx* maps to eleven characters. When the Xixia characters are implemented in Unicode the transliteration scheme will also be useful to develop a generic input scheme that can be used with any computer application.

Whilst the above is good for searching, it needs something more if it is going to act as a mnemonic for the exact structure of a character. Another device is needed to distinguish between the 21 different mappings to the letter *g* for example. This is where the vowels are used. The radicals mapped to a given letter are ranked from most to least frequent. An *a* is appended to the letter for the radical that is most frequent, an *e* to the second most frequent radical and so on as outlined in the following table:

Table 2.7: Transliteration Vowel Suffixes

Rank	Suffix
1	a
2	e
3	i
4	o
5	u
6	aa
7	ee
8	ii

Table 2.7: Transliteration Vowel Suffixes

Rank	Suffix
9	oo
10	uu
11	ae
12	ai
13	ao
14	au
15	ea
16	ei
17	eo
18	eu
19	ia
20	ie
21	io

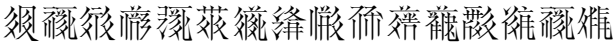
For example, if we consider radicals mapped to the letter *m*, the most frequent is 一, so this radical is pronounced *ma* (appending an *a* indicates top ranking in the radical class denoted by the letter *m*). Similarly, the second ranked radical is 艹, pronounced *me* and the third ranked radical is 巾, pronounced *mi*.

The one exception to this rule is for the *d* radicals - by frequency the first radical *da* would have two dots and the second *de* would have one dot, which is a bit confusing and hard to remember. This has been remapped so that *da* has one dot, *de* has two dots and *di* has three dots.

To make the transliterated words easier to remember and pronounce, each transliteration has been subdivided into groups of two syllables, separated by dashes. The final grouping has been set to three syllables where necessary to avoid having a syllable sitting on its own.

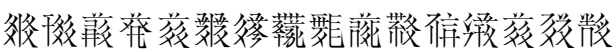
An extract from the Tangut version of the Lotus Sutra is transliterated below [48, pp.691–693]:

xene masa-waama-desa-xace xema-malaxe dema-wede-manasi cacaa-made-saxace mexa-laxa dema-nama-satace



xene masa-waama-desa-xace xema-malaxe dema-wede-manasi cacaa-made-saxace mexa-laxa xede-mala-xoce cicaa-xamaha wede-mahexe sawe-zawu dema-xuvena dema-desa-lata-laqe mame-faaxaxe jala-xada-magaa masa-waama-desa-xace xatiire

xaxa-wexe xeti decu-laxace woci-fati-wofa xela-decaxa tola-mamagii mahe-mama-mama-nace dema-wija-laxe cana-xela maxela dema-zawu-xase woci-fati-wofa mele-xura dema-tuuxa xetaa-xase sewaara



xolaxo texoxe dema-cahaxe dema-xixiga dema-xexe yasa-xila-yasaxi xeya-xase musa-xite-dema-mawace mame-maxabi dema-cana-xeco meza-wuxo sawe-dama-mahe wiya-texa dema-xexe saxa-desaxi vena-leda-daxa

janaa-lece xeda-mama-nare desa-xacase yala-xace dema-wija-laxe maxela dema-wija-laxe woci-fati-wofa cana-xela dema-leda-ledane xaxe-macana xatiixe dasa-xide-manaxe lewe-mama-xace teecu-

laxace teecu-laxace

[illegible]

deca-neda-mace woci-fati-wofa mele-xura jala-xada-magaa cacaa-made-saxace mexa-laxa xede-mala-xoce latea-xecana lati-dema-vixe yaxa-canace desa-xacase mala-desa-xasahee mala-desa-xasahee daxe-fode-macee lawi-cudane wile-weda-mama-saxi

[illegible]

mahii-xaxa xeca-mana lati-dema-vixe xeela-mazawu xema-lagao dema-lade-saxaqe cama-xewo-cifi mexe-laxe dema-desa-lataqe kadewee kada-xeda-mago wise-tama-mace wexeco peha-xeco wetuu-maha texeco

[illegible]

mane-dama-manaxe desa-xixeela cama-wada-mana janaa-lece manu-cuse manu-demacee cacaa-lama-saxira mahi-saxi-vacifi wema-desa-xace cawa-damaga mama-mana-mago dema-tuuxa lele-lema-neli yaxa-canace deza-wema-xase

[illegible]

cacaa-made-saxace mexa-laxa kadewee kada-xeda-mago desa-mexa-texaco xoxexe deca-naja-laxa xekace xema-manace deya-xigao zexa-satace jahe-qaxi mewexa sahu-caxe-damace deza-wema-xase yaxa-yaxa

[illegible]

xaxe-macana desa-xacaxa xela-macana xeda-mama-nare xema-lagao cacaa-made-saxace mexa-laxa nida-manace xela-yacagee kadewee kada-xeda-mago jala-xada-magaa xema-malaxe dema-wede-manasi masa-waama-desa-xace kada-xeda-mago

[illegible]

woci-fati-wofa mahi-yaxa deca-nama-mace xede-cana-xaga xeela dema-naxe-mage mame-xaxace dama-lame-zawu dama-lade-zawe yaxa-yaxa lele-wela-desaxi vena-leda-daxa desa-xalaxo maaci-faze-daxe cile-tija-laxe

[illegible]

lawi-cudane sahu-caxe-damace xema-manace dama-namago xeda-mama-nare mame-sewaa-zata wema-desa-xace dewa-yaxa cama-naxe manu-canace xatiixe desa-xixeela yama-naxaxa mahe-desaxa desa-xaca-xace

[illegible]

xema-mexeela xema-manace cacaa-lama-saxira deya-xigao dama-namago xole-pala xeela-mama-wace dema-xexe nima-mawa saxa-desaxi cile-tima-xala woci-faxe jahe-qaxi sahu-caxe-damace cama-lama-saxice

[illegible]

lawi-cudane mama-lafoqe cacaa-lama-saxira xeci-yacagee meza-wuxo cile-tide-malaga kadewee mama-mana-mago xoxa lade-saxa woci-faxe kada-xeda-mago xela-caxa caxa-xecace mahi-yaxa kamago

[illegible]

kadewee cixa-zala-tala-demacee dema-nama-mawace deza-wede-mavixe lawi-cudane janaa-lece mahii-xaxa dama-namago yala-xace yala-xace dema-sata-lasata xema-desa-xace wise-tama-mace cacaa-lama-saxira vami-wija-hece wica-mala-nace

[illegible]

xeela-xemala cahe-daxe xola-lepala yanixe dama-nade-saxi cama-naxe fele-lewe-caba caca-hema-manaqe xema-lagao maaci-laxe cude-zoxico lade-saxa-desaxa cahe-daxe maze-lanaxe xeca-sata

[illegible]

xejagaa xeya-mana-xaxa janaa-lece mahi-yaxa dema-cana-xeco cama-wada-mana xetaa-xase wexe-cace wexa-cico xesa-xixe zela-naxase dema-nama-satace yaxa-canace xede-zawe xeca-nabe xede-mala-xoce xaxe-macana

[illegible]

texoce maze-lanaxe lexu-zacifa dema-xixiga xede-saga cama-lama-saxice cana-xocata late-desaxa deza-wema-xase deca-xayaxi deza-wede-maga nesi deca-neda-malaxa sawe-dama-mahe cama-nadewee xaxa-masa-tace

[illegible]

lexu-zacifi wiya-texa dama-xela-caga dema-xexe mahi-xexe-case

Chapter 3

Xixia Dictionaries

This chapter gives the background to the dictionaries supplement to this thesis. The supplement contains a categorical dictionary and a variety of transliteration dictionaries. Further information on the existing published Xixia dictionaries can be found in the overview in 5.3.

3.1 Categorical Dictionary

There have been many attempts to build up dictionaries based on radicals, as outlined in Table 2.1. Although this can be done, it is hard to deduce meanings for many of the radicals. It doesn't appear that anyone has yet tried this approach from the bottom-up, classifying characters based on their meanings and then trying to find patterns in resultant character groupings. The categorical dictionary is an attempt to fill this gap in the literature. It is also useful for general lookup - it has advantages over a normal bilingual dictionary in that it is not necessary to guess at how something may have been described in the source language (English in this case) - starting at the broad grouping it is just a matter of browsing through a couple of pages until the desired character is found.

Not all characters are represented here - surnames and foreign transliterations are excluded. Some characters are also represented more than once where they naturally fit under several groupings. Characteristics that are on a sliding scale tend to also contain their opposites, eg. the *beauty* grouping also contains the characters meaning *ugly*. The definitions in both the categorical dictionary and the transliteration dictionaries below have been taken from the English definitions in Li Fanwen's dictionary [41].

3.2 Transliteration Dictionary

The transliteration dictionary orders the characters by the transliteration scheme developed above in section 2.6. Every character in the Mojikyo font is included in this dictionary, so it can be used to find the Mojikyo numbers and definitions of unfamiliar characters.

The dictionary is ordered by the first syllable of the transliteration, using the vowel mapping as shown on page 23. So for those syllables starting with the letter *g*, the order would be *ga, ge, gi, go, gu, gaa, gee, gii, goo, guu, gae, gai, gao, gau, gea, ...*. Within the listing for a particular first syllable the words are listed in alphabetical order, ignoring the dashes used to separate groups of syllables.

This dictionary can be treated solely as a convenient device for looking up characters. It is also useful in examining how the script might have worked if it was phonetic in a similar manner to Korean. This is a highly speculative hypothesis, but one worth examining as the principles

underlying the script are still unknown. If the script were phonetic a substitution would need to take place between the transliterated syllables shown here and the sounds of the original language to restore the script to its original usage.

Nishida specifically warns against going down this phonetic path [53, p.10]. Indeed, it is unlikely to be a pure phonetic system that ties in with the phonology known from other sources or the structure would have been uncovered long ago. However, given that scholars aren't even sure whether the script represents one language or two [27] it is important to keep an open mind. Even if the script is partly phonetic the transliteration scheme will be useful in uncovering the areas where it works. If it turns out there are no phonetic elements at all the effort has not been wasted as the transliteration is still useful as a mnemonic device for looking up and remembering the characters.

3.3 Suffix Dictionary

The suffix dictionary reverses the reading order to go from bottom right to top left. This has been done by reversing the order of all the syllables in the transliteration dictionary, but preserving the position of the dash separators. For example, the character for two  is transliterated as tola-mamagii, but is listed in the suffix dictionary under giimama-lato.

It is fairly well accepted that the Xixia language can be classified as a Tibeto-Burman language [14, p.602], [10, p.156], one of the two major subgroups of the Sino-Tibetan family (the other being Chinese) [67, p.6]. If this is the case, then it is likely that the language contains a system of prefixes and suffixes [37, pp.22-29]. As outlined on page 30, a system of verbal prefixes has already been uncovered at the level of whole characters. The transliteration and suffix dictionaries are useful in looking for prefixes and suffixes within a character if the script was in fact phonetic.

3.4 Alternative Reading Orders

The transliteration and suffix dictionaries above assume a reading order of top left to bottom right, reading vertically top to bottom before moving horizontally to the next section of the character. This is only one of a number of possible reading orders. The possibility of the Xixia people reading bottom right to top left has also been covered - this just means that the roles of the transliteration and suffix dictionaries above are swapped around. The transliteration dictionary becomes the suffix dictionary and the suffix dictionary becomes the transliteration dictionary.

The alternative transliteration dictionary assumes a reading order of top right to bottom left. This has been constructed by reversing the order of the horizontal components in the recursive radical code. A new transliteration has then been constructed based on this modified code.

Finally, the alternative suffix dictionary is constructed from the alternative transliteration dictionary in the same manner as the main suffix dictionary is constructed from the main transliteration dictionary. If the reading order was actually bottom left to top right, this can be achieved just by reversing the roles of the alternative transliteration and alternative suffix dictionaries.

By providing all four possible reading orders, it is possible to use these dictionaries to start from any corner in looking up a character. This is useful as quite often in the original literature a particular corner is almost illegible. It is also easy to find characters which are similar in a particular corner.

A hypothesis that the reading order is horizontal first rather than vertical is much harder to sustain, as it is difficult to split up most characters into horizontal bands.

Chapter 4

Xixia Grammar

This chapter presents a brief summary of the features of Xixia grammar. For a much more detailed discussion of this area (in English), consult [14] and [55, pp.557–587]. There are also extensive investigations of grammar (in Russian) in the books by Sofronov [64] and Kepping [21].

In English, the basic word order for sentences is Subject–Verb–Object. In the sentence "I see John", the subject is *I*, the verb is *see* and the object is *John*. In Xixia (like Tibetan [9, p.11], Japanese and Korean [7, p.48]), the basic word order is Subject–Object–Verb [14, p.613]. The above simple sentence would be transformed into "I John see".

There are both formal and informal forms for first person (English: I) and second person (English: you) pronouns. The third person pronoun (English: he, she, it) is the same regardless of gender or whether the object referred to is animate. There is also a reflexive pronoun (English: oneself) - these pronouns are shown in the table below:

Table 4.1: Xixia Pronouns (adapted from [14, p.607])

Type	Informal		Formal	
First Person	2098	[illegible]	261	[illegible]
Second Person	3926	[illegible]	4028	[illegible]
Third Person	388	[illegible]		
Reflexive	1245	[illegible]		

The Xixia language has a case structure, indicated by particles [14, p.606]. The genitive, dative and accusative cases are all represented by one character. There are many different characters used to represent locative case.

Table 4.2: Xixia Case Structure (adapted from [14, pp.606–607])

Case	Number	Character
Genitive, dative and accusative	1139	𗍫
Instrumental	5880	𘝞
Locative - In (the garden)	2983	𗗚
Locative - In (the book, the water)	5993	𘜶
Locative - In (the water)	1477	𗾞
Locative - In (the sky, the heart)	5856	𘟙
Locative - On	89	𗦳
Locative - Under	5399	𘜘
Allative - To	5447	𘝦
Comitative - With	4950	𗥃
Comparative	1473	𗵘
Ergative marker	5604, 5113	𗤶𗼻

For those whose formal grammar is a bit rusty, definitions of the above terms are outlined below [50]:

Table 4.3: Grammatical Case Terminology

Case	Definition
Genitive	Indicates possession - In English this is indicated by 's, eg. John's book
Accusative	Used for the object a verb acts on directly. Eg. In "I see John", John would be in accusative case.
Dative	Used for the object that is indirectly affected by an action. eg. In "I gave a bone to the dog", the direct object is the bone (which would be in accusative case) and the object indirectly affected by the action is the dog (which would be in dative case).
Instrumental	Indicates use of an instrument or a tool to perform an action. Eg. In "I cut the grass with a lawnmower", lawnmower would be in instrumental case.
Locative	Indicates location, Eg. In "I live in Australia", Australia would be in locative case.
Allative	Indicates movement towards a location. Eg. In "I swam to the island", island would be in allative case.
Comitative	Indicates companionship - Best indicated by English word "with"
Comparative	Used to relate one entity to another, often indicated by the ending *er* in English
Ergative Marker	Indicates the subject of transitive verbs (verbs which take one or more objects). Eg. In the phrase "I see John": The verb *see* is transitive as it takes the object *John*. The subject of this verb is *I* (which could be indicated by an ergative marker in Xixia).

Xixia also has a series of verbal prefixes [14, p.608]. The directional markers also indicate perfective

aspect (that the action indicated by the verb has been or will be completed - often shown in English by an *ed* suffix). The optative markers indicate a wish - they are used to show verbal mood [50, p.251].

Table 4.4: Xixia Verbal Prefixes [14, p.608]

Direction	Directional Marker		Optative Marker	
upward	5981		3989	
downward	1452		3846	
here, inside	1326		2219	
there, outside	2590		2536	
towards the speaker	804		4841	
away from the speaker	4342		4841	
(direction not found)	795		3707	

As shown in the categorical dictionary, there are a number of different characters used to represent numbers in the Xixia script. The ones used in page numbering in *W*enhai [63] (and so likely to be in relatively common use) are the following:

Table 4.5: Xixia Numbers

Mojikyo No.	Character	Meaning
100		One
4027		Two
4344		Three
2205		Four
36		Five
3200		Six
332		Seven
4602		Eight
3113		Nine
1040		Ten

Chapter 5

Literature Overview

This section gives a brief outline of the research literature. For a detailed index of almost everything ever published on the Xixia, refer to the 135 page listing in [43].

5.1 Language Requirements

The main language requirement for working with Xixia material is Chinese. The majority of papers and books are written in Chinese and the construction of the Xixia script has many parallels with the Chinese script. The second major support language is Russian, as the majority of primary material is stored in St Petersburg following Kozlov's expeditions [29]. Some material has been written in English [15, 30, 11], but not much recently. Knowledge of Japanese is useful in working through the works of Nishida Tatsuo [54, 55]. Some material has been published in French and German (including Guillaume's recent book [18]), but these languages play only a minor supporting role.

Knowledge of classical Chinese is also useful in delving into the history of the Xixia people. It is only since the early 20th century that the Chinese have used the modern spoken language in their written works [59, p.314]. References to the Xixia in historical documents of the period use the classical language and some of the material found has been translations of classical Chinese works into the Xixia language.

5.2 Fonts

It is only in recent years that fonts have been available for working with the Xixia script on a computer. Many of the earlier books have been painstakingly written by hand. Going back 25 years even the Chinese explanations in these books were hand-written (such as [63]).

The Mojikyo font[1] is a milestone in that it was the first font available for working with the script. It uses the same numbering system as Li Fanwen's dictionary [41], i.e. in four corners order. However, it has a few limitations. The character set is spread over two fonts. For characters up to no. 5280, font "Mojikyo M202" should be used. For characters from no. 5281 onwards, font "Mojikyo M203" should be used. This can cause difficulties when working with multiple characters as the font name needs to change depending upon the character used. The Mojikyo font also requires a licence to use in publications, but the software given with this licence is only guaranteed to work on Japanese Windows.

[1] http://www.mojikyo.org

Some of the characters in Mojikyo appear to be duplicates, eg. character numbers 3488 and 3489 are both 𗗙. This is due to characters with more than one reading being given separate numbers for each reading, and follows the treatment in Li FanWen's dictionary [41, 71].

Ningxia People's Press has recently published a software input system based on the four-corners method [20]. There is also a font developed by Zhongyi Electronics Ltd[2].

There is currently a proposal to incorporate the Xixia script into Unicode, following work by Cook [8]. This should make it much easier to work with the script provided that fonts are developed that can render the characters in the newly allocated positions.

5.3 Dictionaries

There are a number of dictionaries available for working with the Xixia script. Several dictionaries were amongst the source material recovered from Khara-Khoto. It is quite rare to have such good source material when trying to decipher a dead language. Other dictionaries have been constructed by scholars in the hundred years since these initial discoveries.

The *Pearl in the Palm* is a bilingual Xixia - Chinese dictionary recovered from Khara-Khoto [30, p.2]. It is important to note that this should be related back to twelfth century north-western Chinese rather than modern Mandarin [30, p.4]. It has the following information in four columns (from left to right) [30, p.3]:
1. Xixia character with the same sound as the Chinese character in column 2
2. Chinese character
3. Xixia translation of Chinese character in column 2
4. Chinese character with the same sound as the Xixia character in column 3

The document is broken up into a series of basic categories as follows [54, pp.189–223]:

Heaven
I. Heaven itself - Different descriptors of heaven
II. Heavenly bodies - Sun, planets, moon, etc.
III. Phenomena of heaven - Weather, time expressions

Earth
I. Earth itself - Descriptors of earth
II. Earthly bodies - Geographical terms, eg. rivers, lakes
III. Phenomena and products of earth - Disasters, minerals, plants, animals, insects

Man
I. Man himself - Types of men
II. Physical aspects of man - Body parts
III. Affairs of man - Relationships, instruments, buildings, tools, household goods, clothing, government positions, food, moral stories, Buddhist terms

The most accessible version of the *Pearl in the Palm* for the English language speaker is the English translation in Nishida's work [54, pp.186–223]. Kwanten gives a detailed analysis of the work in English [30]. There is also an extensive analysis of the work in Chinese by Li Fanwen [40].

Other works that give information on Xixia phonetics include *Homophones* [39] and *Rhyme Tables of Five Sound Categories* [44].

[2] http://www.china-e.com.cn/en/products/XiXia.htm

The *Sea of Characters*, or *Wenhai*, is an original Xixia dictionary written only in Xixia characters. It shows how characters are built up from simpler characters. *Wenhai* was originally in three parts, namely the level tone, the rising tone and miscellaneous characters. Currently access is only available to the level tone and miscellaneous parts. The rising tone section was also originally recovered, but was lost sometime between 1937 and 1956. This was a period of political turmoil in the old Soviet Union during which time Tangut studies were not undertaken [70, p.521]. There exist both Chinese [63] and Russian [22] translations of *Wenhai*. For a full explanation of how the *Wenhai* works see the translation of an article by Shi Jinbo in Appendix A.

Probably the most comprehensive of the early dictionaries following rediscovery of the script was that produced by Nevsky [46, pp.111-1216]. His handwritten manuscript was published posthumously and shows the state of his work when he died in 1938. Although it is not in a final polished form it still gives information to varying degrees on 6000 Xixia characters [46, p.111]. For a given Xixia character it gives information on Chinese characters with the same sound, its tone, its phonetic grouping (labial, dental, etc.) and its meaning in Chinese, Russian, English and sometimes Tibetan. It also gives example sentences in Xixia and Chinese, with occasional translations into Russian, English, Tibetan or Sanskrit and corresponding sounds in other languages in the Tibeto-Burman family [46, p.112].

Nishida Tatsuo's *Small Xixia Character Dictionary* [55, pp.303–507] is a bilingual Tangut-Japanese dictionary. It is ordered by a set of 319 radicals. The numbering system has inherent meaning - character no. 1-082 refers to the second character with radical no. 1 and 8 strokes in the character apart from the radical. The structure tag for each character is shown, as outlined in Figure 2.5 on page 13. The components in each position within this structure are also shown - if this was computerised it would give a relatively powerful approach to searching for Xixia characters, but it is hard to use this for searching in its printed form. It contains an index at the back of radicals by stroke count.

Grinstead's dictionary [15] is the most accessible dictionary for the English speaker without foreign language skills. Grinstead sets up his own "Telecode" index, indexed by the shape of the component in the bottom right-hand corner [15, p.39]. As most other radical-type indexing methods tend to use the top-left corner this index complements the other dictionaries. It corresponds to the indexing used in the main suffix dictionary in the dictionaries supplement to this thesis. Grinstead gives a mapping from *Wenhai* numbers to his Telecode index [15, p.152]. He also gives an English-Tangut word list and notes the richness of Tangut synonyms for a given English word [15, p.199].

The most famous modern dictionary for the Xixia script would be Li Fanwen's *Tangut-Chinese dictionary* [41]. This dictionary introduces the four corners method explained above - the characters are ordered by this method. It also includes radical and phonetic indices. The dictionary entries include extensive quotations and translations from primary source documents. The dictionary could have done with a native English speaking editor - most of the mistakes in the English definitions are just spelling errors but occasionally a word is unrecognisable. The preface is well worth reading, outlining the trials and tribulations of Li Fanwen's efforts to compile a dictionary over more than twenty years. A new (2008) edition of this dictionary has just been published.

The *Tangut-Russian-English-Chinese Dictionary* has also been recently published [35]. This dictionary is indexed by radicals, with the radicals themselves being ordered by their basic shape (horizontal radicals are classified as shape A, vertical radicals as shape B, etc.). This dictionary has the advantage of providing definitions for Xixia words (combinations of characters) in addition to the individual characters themselves. It is very neat and tidy, being the first dictionary to take advantage of the Mojikyo font. Having the four languages in one dictionary means it is also of use to researchers trying to learn one of the component languages. This dictionary is only available

to Xixia researchers directly from the publisher.

5.4 Buddhist Texts

More than 80% of the material recovered from the ancient capital Khara-Khoto has been Buddhist texts [42, p.14]. This reflects the fact that Xixia was a Buddhist state - the Buddhist Tripitaka had been translated into the Xixia script before 1200 AD [10, pp.196, 204]. Trying to understand this material presents a double difficulty - the Xixia language is difficult to read and Buddhist texts are difficult in their own right. However, in recent years researchers have been making inroads in this area.

Nishida Tatsuo's work on the *Lotus Sutra* [56] has been published in an attractive book featuring a colour reproduction of the original manuscript along with a Japanese translation. This book is only available by donation from Soka Gakkai to institutions - access to this book was obtained by getting the National Library of Australia to request a copy. Lin Yingchin has also recently published an extensive analysis of the Tangut version of *Chanting the Names of Manjusri* [49].

5.5 Translations of Chinese Texts

A number of translations of Chinese works into Xixia have also been uncovered. These include the Chinese military treatise *The Art of War* (Sunzi Bingfa) [47, 23]. There are also translations into Xixia of the *Confucian Analects* and *Mencius* [76, p.584].

Two works that are no longer available in their original Chinese editions have also been recovered [51, p.332], namely the reference book *Forest of Categories* (Lei Lin) [52, 24] and the *Classic of Filial Piety* (Xiao Jing) [15, pp.277–376].

There are also Xixia translations of the Chinese medical works *Chinese Materia Medica* (Ben Cao) and *A Thousand Golden Remedies* (Qian Jin Fang) [42, p.643].

Biographies based on Chinese works include *Moral Biographies* and *New Collection of Compassionate and Filial Biographies* [76, p.585].

5.6 Collections of Original Texts

Xixia legal works give a good insight into the workings of their society. Research on the legal documents uncovered are contained in [34, 75, 5]. In particular, [34] reproduces the text of the *New and Reformed Law Codes of the Prosperous Kingdom of Heaven.*

Textbooks for recognising characters have also been uncovered, namely the *Thousand Character Classic* and *New Collected Analysis of Characters for the Classroom* [42, p.620].

There exists a Xixia text for the treatment of scabies, which was the most common disease of the Xixia people [42, p.645]. There are also various medical references in the *Wenhai*, *Homophones* and *Pearl in the Palm* [42, pp.645-647]. Xixia astronomical works include the *Almanac of the Nine Celestial Bodies* [42, p.635] and *Methods for Annual Observations of the Heavens* [76, p.587].

As discussed in [76, p.585], there is an historical work titled *Twelve Countries* that shows similarities to both *National Language (Guo Yu)* and *Strategies of the Warring States (Zhan Guo Ce)*. It is not clear whether this is a native Xixia work or a translation from Chinese.

There is also a Xixia encyclopedia called the *Sea of Sacred Information (Shengli Yihai)* [36], published in 1182 [76, p.587]. It includes entries on astronomy, geography, agriculture, terms of address, palace etiquette, registration of births and ethics [42, p.12].

Tangut odes and ritual songs have also been uncovered. The odes include [28, p.120]:
a) The Ode on Ritual Verses
b) The Great Ode
c) The Ode on Monthly Pleasures
d) The Ode of Sayings
e) The Ode on Wisdom
The odes appear to be older than the ritual songs, as they lack Buddhist terminology and also provide support for the idea that the Tanguts had a ritual language [28, p.121].

5.7 Histories

The most extensive published history in English is the sixty page chapter by Dunnell [10]. Dunnell's PhD thesis gives a more detailed view of this material [12]. Dunnell has also published a book on Buddhism in the Xixia state [11]. A broad history in Russian is Kychanov's book [33].

The recent Xixia history (in Chinese) by Li Fanwen [42] at 725 pages is probably the most comprehensive, including many illustrations of artefacts. However, it is fairly heavy-going for the non-native Chinese speaker, containing frequent excursions into classical Chinese. For a basic introduction to Xixia history, Shi Jinbo's *Xixia Culture* is easier to understand [62]. For the serious historian there is also a nine-volume collection of references in historical Chinese documents to the Xixia [16] - this is impossible to understand without a strong grounding in classical Chinese. Two very recent books on Xixia history are those of Wu Tianchi [74] and Niu Dasheng [57].

More popular books include *Xixia* [2], written in traditional characters and containing many colour illustrations. There is also the recently published *Kingdom Extinguished* by Tang Rongyao [66].

Chapter 6

Conclusions

The main outcomes of this thesis are:

a) The development of a method to transliterate Xixia characters and analyse the structure within individual characters. The method is also convenient to use in searching for particular characters and acts as a mnemonic to help in remembering them. It may also lead to a convenient method for character input when the Xixia characters become part of the Unicode standard.

b) An overview of the Xixia script and research literature for the English speaker. The Xixia research literature is fairly inaccessible for those without Chinese language skills - hopefully this thesis will assist in opening up the language to others.

As can be seen there are tantalising hints that a structure exists in the Xixia script, but despite one hundred years of analysis an all-encompassing structure has yet to be uncovered. It may be that the Xixia had a method of thinking about the world that we have yet to uncover, so that their script was actually fairly easy to use. Or it may be that the script really is as complicated as it appears, taking the Xixia scribes years to master. The next stage will be to take the techniques developed here together with tools from mathematics and computational linguistics in order to get insight into this question.

Appendix A

Translation Project - *Wenhai* Article

The Structural Features of Xixia Characters that can be Inferred from the *Sea of Characters (Wenhai)*

Author: Shi Jinbo 史金波

1. The simple Xixia characters
2. The main character construction methods are composites of simpler characters and mixtures of sounds and meaning.
3. Initial-final characters at the top and bottom is a special way of creating Xixia characters.
4. Indirect sound-meaning combinations reflect the influence of Chinese character construction methods. Long sound combinations record the long sounds of Sanskrit.
5. The exchange of parts is one method of creating characters of similar meaning.
6. The meaning and radicals of characters similar in form.
7. The flexibility of shapes when forming Xixia characters.

Below every character in the Xixia dictionary *Wenhai* is an explanation of how that character was composed. This is of conclusive significance for research on the construction of Xixia characters. In the past, Chinese and foreign scholars produced much research on the construction of Xixia characters, which was quite an achievement [note 1]. However, as they either hadn't seen the *Wenhai*, or it was difficult to get access to it and they hadn't placed sufficient importance on the construction of these character shapes, the results of their research had certain limitations. Over the last two years the authors have conducted the first steps in research and classification of the composition of more than three thousand characters contained in the character strips (the level tone and miscellaneous character parts). As the analysis was refined it can be said that we gradually came into contact with the principles and rules of Xixia character construction. In 1978 the author wrote "Summary of the Structure of Xixia Characters" [note 2], which introduced research on the composition of Xixia characters. It explained the significance of the research and looked at past research both within China and abroad. It also emphasised the general principles of Xixia character construction and discussed the relationship between Xixia and Chinese characters. This work will not systematically cover this material again - my only plan is to reflect several characteristics of Xixia character construction contained in the *Wenhai* and discuss examples.

Every character in *Wenhai* has a four-character explanation (over two lines) of how it was formed [note 3]. Often the first character on each line depicts either part of the character or a whole character to which parts are added. The first character on the first line often comprises

that part of the character that is written first. The first character on the second line often forms that part of the character which is written last. The second character on each line indicate what part of the first character is utilised. These have been termed the indicative character. There are twelve indicative characters, which have meanings of left, right, top, bottom, middle, surround, whole, foot, subtract, interior, remove and retain. Sometimes numbers are also used. For those *Wenhai* entries composed of three characters there is only one indicative character. For characters with four components no indicative character is shown.

The examples below are shown in the format used by the *Wenhai* and give the general idea as to how characters are composed:

12.162		36.121		43.232	
 [yu]		 (skirt)		 (defeat)	
 [yu]	 (sound)	 (surround)	 (waist)	 (lose)	 (army)
 (left)	 (left)	 (whole)	 (whole)	 (left)	 (surround)

In the above parentheses indicate a character's meaning and brackets indicate a character's sound. For example, the largest character in the first example is a character used for transliteration (pronounced yu[1]). The character is formed from the left-hand side of a homophonic character and the left-hand side of a character meaning "sound". The other examples are similar. The Xixia character for "skirt" is composed from the whole character for "surround" and the whole character for "waist". The Xixia character for "defeat" is made up from the left part of the character for "lose" and over half of the character for "army".

Wenhai has seven different ways of choosing the components from two characters that are matched together to create a new one - left, right, whole, top, surround, middle and bottom. There are 39 methods of this type. There are five methods of removing part of a character to create a new one. Method no. 45 is in a class of its own. There are two methods that use the foot of a character. Of the methods that use three component characters, there are twelve methods that use one indicative character and one method that doesn't have any indicative characters. There is only one method that uses four component characters and it doesn't have an indicative character. There are four specialised methods (nos 62-65). In total there are 65 different methods, as outlined in the table below. In this table the component characters are labelled (1) through to (4).

1. (1) left + (2) right	2. (1) left + (2) left
3. (1) left + (2) whole	4. (1) left + (2) middle
5. (1) left + (2) bottom	6. (1) left + (2) surround
7. (1) right + (2) right	8. (1) right + (2) left
9. (1) right + (2) whole	10. (1) right + (2) middle
11. (1) right + (2) bottom	12. (1) top + (2) right
13. (1) top + (2) left	14. (1) top + (2) whole
15. (1) top + (2) middle	16. (1) top + (2) bottom
17. (1) top + (2) surround	18. (1) top + (2) top
19. (1) whole + (2) right	20. (1) whole + (2) left
21. (1) whole + (2) whole	22. (1) whole + (2) middle

[1]In this translation modern Chinese pinyin is used to represent readings. To be fully correct the pronunciation of twelfth century North-Western Chinese should be used, but this is a major project in its own right - see Li Fanwen's work [40] for further details.

23. (1) whole + (2) bottom	24. (1) whole + (2) top
25. (1) surround + (2) right	26. (1) surround + (2) left
27. (1) surround + (2) entire	28. (1) surround + (2) middle
29. (1) surround + (2) bottom	30. (1) middle + (2) right
31. (1) middle + (2) left	32. (1) middle + (2) whole
33. (1) middle + (2) middle	34. (1) middle + (2) bottom
35. (1) bottom + (2) right	36. (1) bottom + (2) left
37. (1) bottom + (2) whole	38. (1) bottom + (2) middle
39. (1) bottom + (2) bottom	40. Subtract left of (1)
41. Subtract (2) from (1)	42. Subtract top of (1)
43. Remove top of (1)	44. Remove left of (2)
45. Retain right of (1)	46. (1) foot (2) whole
47. (1) foot (2) interior	48. (1) right + (2) + (3)
49. (1) left + (2) + (3)	50. (1) whole + (2) + (3)
51. (1) surround + (2) + (3)	52. (1) middle + (2) + (3)
53. (1) top + (2) + (3)	54. (1) + (2) + (3) right
55. (1) + (2) + (3) left	56. (1) + (2) + (3) whole
57. (1) + (2) + (3) middle	58. (1) + (2) + (3) bottom
59. (1) + (2) + (3) top	60. (1) + (2) + (3)
61. (1) + (2) + (3) + (4)	62. 2 x (1) + (2) middle
63. (1) swap components	64. people link three worlds[2]
65. similar to (1)	

The first six types above amount to over 40% of the characters in *Wenhai*, but the last 26 types only account for around 10%. What follows is an exposition of how the Xixia characters are composed.

A.1 The Simple Xixia Characters

It is well-known that the Xixia characters are complex. Ancient Chinese dictionaries recorded the Xixia characters as being "overly complicated, somewhat resembling the seal characters" [note 4]. It can be seen that they are more complicated than Chinese characters. However, from the more than 6,000 Xixia characters it is possible to distinguish some of them as being relatively simple. This simplicity extends to both the individual strokes and the overall shape of the character. They are not formed from other characters, but form the foundation for them. We call these characters the "simple characters". These simple characters can't be further subdivided into smaller units, regardless of whether analysed from the point of view of phonetics or semantics. If these characters are subdivided into smaller components, it is impossible to see the relationship with the original characters in either sound or meaning.

Wenhai's explanation of the formation of the simple characters doesn't give a direct analysis of them and doesn't seem to recognise that this class of characters exists. However, if we carefully analyse the explanations given for these characters, we can see from many different angles that the simple characters indeed exist. The authors of *Wenhai* didn't specifically state that these characters aren't formed from others, but there isn't a better explanation of this class. So we use two imperfect methods. One method is to look at the more complicated characters that incorporate the simple characters and other characters where the simple characters take a leading role in the mixture. For example:

[2]This explanation comes from the Xixia character for Buddha 丱 being composed of the character for person 夂 and three horizontal strokes linked by a vertical stroke. [61]

5.123  from  and 
10.141  from  and 
37.162  from  and 
65.112  from  and 
84.152  from  and 
Miscellaneous 9.113  from  and 

The above character formation methods are hard to explain. It should be pointed out that some characters are simply constructed but aren't simple characters. *Wenhai* has also adopted this method. A more unusual method is to remove part of a more complex character (such as the top or a side) to create a new one. For example:

1. Remove the left part
59.243  (insect, pronounced *dang*) is formed by  (remove the left part of )
92.111  (five, pronounced *ze*) is formed by  (remove the left part of )

2. Remove the top
85.232  (eight, pronounced *ye*) is formed by  (remove the top of )
Miscellaneous 7.242  (waist, pronounced *zhou ni*) is formed by  (remove the top of )
Miscellaneous 15.221  (immortal, pronounced *cu ni*) is formed by (remove the top of)
16.222  (thin, pronounced *mi*) is formed by  (take the right part of )
The last example points out that this category can also contain characters formed from the right part of another character.

3. Remove a side
27.119  (ten, pronounced *she*) is formed by  (remove  from)
21.171 (old, pronounced *chen*) is formed by (remove from)
Miscellaneous 6.171 (skin, pronounced *ze ni*)) is formed by (remove from)
Miscellaneous 15.172 (one, pronounced *jing ni*) is formed by (remove from)

The above methods all form simple characters from more complex ones. Examining this not entirely satisfactory character formation method it seems that the Xixia people weren't able to break up these simple characters either. If we try and analyse them we are unable to explain the meaning of the component parts. The authors of *Wenhai* have no way to comprehensively explain how these characters were derived.

Where the Xixia simple characters have come from is a question worthy of further study. The Chinese simple characters have mainly come from logograms and self-explanatory characters. These two classes of characters are not obvious in the Xixia characters. This is probably another reason why the *Wenhai* authors have difficulty explaining where the simple characters have come from. However, the formation of some of the simple characters possibly have some relation with logograms. For example, (insect) is like an insect with many feet, (person) is like a person walking and (grain) resembles the shape of a seedling.

The simple characters can be classified into two types. One type consists of commonly used words. The other type consists of borrowings from foreign languages, place names, names of people and transliterations from Buddhist texts. This classification can be seen in the following table:

1. Characters Reflecting Meaning

	small		person
	half		one
	above		old
	waist		sacred
	belt		grain
	toil		ladder
	insect		threaten
	horse		skin
	grass		hand
	thin		understand
	no		true
	seed		alone
	tree		female
	iron		

2. Characters Reflecting Sound

	chi		dou
	xia		wei
	he		ju
	le		guo

A.2 The Main Character Construction Methods are Composites of Simpler Characters and Mixtures of Sound and Meaning

The vast majority of Xixia characters are merged characters. Of these the majority are composites of simpler characters or mixtures of components representing sound and meaning. The so-called combination methods use two (more rarely three or four) Xixia characters. Either parts or the whole character from each of these components is used to create a new character. The meaning of the new character is derived from the meanings of the component parts. This method is used much more often in creating Xixia characters than it is used in Chinese. This method was published in *Wenhai* and influenced the early Xixia researchers. Luo Fuchang first adopted this method in 1914 to analyse the construction of some Xixia characters [note 5]. Currently through the thorough explanations of *Wenhai* we are able to recognise the composite characters in a deeper and more systematic way. Through an analysis of the treatment of composite characters in *Wenhai*, we can get a basic idea of how certain characters were formed and some of the unique methods of thought of the Xixia people. In order for readers to easily understand, many examples are listed below:

Table A.2: Characters Formed From Two Component Characters

	choke	=		narrow	+		throat
	inform	=		not	+		clothes
	stop	=		not	+		do
	wide	=		wide	+		wide
	immortal	=		mountain	+		person
	wolf	=		tooth	+		beast
	power	=		step	+		strength
	beg	=		not	+		sufficient
	pots	=		contain	+		mud
	deep water	=		water	+		black
	doubt	=		fear	+		have
	leisure	=		work	+		none
	stupid	=		heart	+		heavy
	envoy	=		speech	+		tutor
	snore	=		air	+		nose
	cangue	=		tree	+		hole
	ancestors	=		head	+		white
	punishment	=		repent	+		crime
	loyal	=		correct	+		virtue
	onion	=		grass	+		tired
	louse	=		small	+		ghost
	curtail	=		length	+		reduce
	benefit	=		seed	+		cross
	scar	=		ugly	+		mark
	sink	=		water	+		solid
	boots	=		leg	+		cover
	marriage	=		woman	+		marry
	shake	=		wine	+		plunder
	autumn	=		rice	+		see
	pottery	=		mud	+		burn
	fear	=		self	+		afraid
	maiden name	=		woman	+		surname
	reed	=		water	+		grass
	lonely	=		mother	+		none
	twelfth moon	=		cold	+		season
	fall	=		step	+		miss
	brass	=		copper	+		make
	choose	=		inspect	+		choose
	proofread	=		justice	+		see
	filthy	=		evil	+		moisten

Table A.2: Characters Formed From Two Component Characters

[illegible]	about 1/6th of acre	=	[illegible]	land	+	[illegible]	tiny
[illegible]	hair rolled in bun	=	[illegible]	head	+	[illegible]	blend
[illegible]	grandson	=	[illegible]	man	+	[illegible]	son
[illegible]	fast	=	[illegible]	exceed	+	[illegible]	horse
[illegible]	rain	=	[illegible]	holy	+	[illegible]	benificence
[illegible]	gather	=	[illegible]	not	+	[illegible]	none
[illegible]	sharp	=	[illegible]	iron	+	[illegible]	skillful
[illegible]	hungry	=	[illegible]	mouth	+	[illegible]	empty

Table A.3: Characters Formed From Three Component Characters

[illegible]	rotten	=	[illegible]	meat	+	[illegible]	ferry	+	[illegible]	not
[illegible]	crawl	=	[illegible]	knee	+	[illegible]	hand	+	[illegible]	go
[illegible]	fluctuate	=	[illegible]	white	+	[illegible]	to	+	[illegible]	black
[illegible]	announce plans	=	[illegible]	not	+	[illegible]	plan	+	[illegible]	treasure
[illegible]	anxiety	=	[illegible]	heart	+	[illegible]	no	+	[illegible]	rest
[illegible]	ring	=	[illegible]	loop	+	[illegible]	angle	+	[illegible]	not
[illegible]	miss	=	[illegible]	heart	+	[illegible]	read aloud	+	[illegible]	as if
[illegible]	stubborn	=	[illegible]	speak	+	[illegible]	not	+	[illegible]	listen
[illegible]	warm period	=	[illegible]	spring	+	[illegible]	autumn	+	[illegible]	source
[illegible]	extinguish	=	[illegible]	from	+	[illegible]	fire	+	[illegible]	not
[illegible]	bite	=	[illegible]	mouth	+	[illegible]	enter	+	[illegible]	relinquish
[illegible]	seed plough	=	[illegible]	agriculture	+	[illegible]	plough	+	[illegible]	seed
[illegible]	die young	=	[illegible]	life	+	[illegible]	not	+	[illegible]	complete
[illegible]	flame	=	[illegible]	end	+	[illegible]	hot	+	[illegible]	fire
[illegible]	warm	=	[illegible]	not	+	[illegible]	hot	+	[illegible]	cold
[illegible]	call of a calf	=	[illegible]	calf	+	[illegible]	of	+	[illegible]	call

The above table shows characters formed by combining three different characters. It is not hard to see that the component characters are like a phrase, combining the meanings of the individual components to form the meaning of the new character. Without the explanations in *Wenhai* it would be fairly difficult to correctly explain the formation of these characters. This type of character construction method is of much interest to us. Not only do they show the formation of characters that are relatively hard to explain, but also, through an analysis of the formation of character shapes, can help us to understand and verify the meanings of these characters.

The knowledge gained by an analysis of the shapes of the great quantity of combination characters makes it easier to analyse the combination characters outside of *Wenhai*. These are mainly rising tone characters. This method is tried on several examples below:

[illegible]	chop	=	[illegible]	move	+	[illegible]	knife

	Chinese clothes	=		clothes	+		Chinese
	scared	=		heart	+		empty
	pen	=		character	+		make
	lamb	=		small	+		sheep

According to these it can be seen that the combination method is an extremely important method of construction. This is not to say that *Wenhai*'s use of combination methods to explain the construction of characters is always reasonable. Sometimes it is possible that the authors of *Wenhai* haven't completely understood the original meaning of the constructed character and have produced some farfetched analyses. Therefore, every character's explanation has to be analysed - it is not possible to take them at face value.

For the characters formed from two components, the relationship between them is relatively complex. Xixia grammar is able to reflect almost all of the relationships in word formation. For example:

Juxtaposed Type: See 14.121, 63.211 and 83.162

Attributive Type: There are a lot of characters of this type, see 9.111, 12.152, 15.212, 15.251, 20.133, 25.233, 29.232, 56.241, 64.231, 72.221, 74.242, 77.212, 14.211 miscellaneous, 17.252 miscellaneous, etc. Among these there are nouns modifying nouns and adjectives modifying nouns (there are both adjectives in front and adjectives behind). There are also adjectives or adverbs that modify verbs.

Verb-Object Type: See 30.133, 36.222, 66.171, 74.111, 74.141, 76.121, 86.132, 4.132 miscellaneous, etc. According to Xixia grammar, the verb-object type should have the object in front and the verb behind. However, in the construction of characters it seems that there are also instances of the verb in front and the object behind, eg. 63.113.

Subject-Predicate Type: See 32.162, 76.152 and 19.221 miscellaneous. It should be pointed out here that for some of the cases where the noun is in front and the adjective behind the attributive and subject-predicate constructions are difficult to separate. For example in 56.241, the Xixia characters for *head* and *white* are combined to form *ancestral honour*. This can be understood to mean *head white* or *white head*. This is because in the absence of knowledge of the linguistic environment of the time it is not possible to prove what the top and bottom component characters are in the construction of the character. For the same reason, there are examples where it is difficult to distinguish between verb-object and subject-predicate types. For example, in 66.171, the characters for *long* and *reduce* are combined to form *curtail*. This can be understood to mean *long reduce* or *reduce long*.

Supplementary Type: This type is relatively uncommon. See 79.241.

The relationship between the components of characters formed by combining by meaning more than two sub-characters are mainly of the subject-predicate and attributive types.

The characters that combine sound and meaning are formed by using parts or the whole of two component characters to form a new character. One component reflects the sound and one component reflects the meaning. i.e. The sound of the new character is the same or similar to the sound of one of the components and the meaning of the new character has a relationship to

the meaning of the other component. These types of characters are similar to the Chinese shape-sound characters, but in these the side that relates to the character's shape is often derived from logograms. Xixia characters lack logograms, so this type can't be called shape-sound characters. They are called sound-meaning combination characters instead.

The number of sound-meaning combination characters is rather large - this is another important way of constructing Xixia characters. Some people think that the Xixia shape-sound characters (i.e. sound-meaning combination characters) are very rare. They think that they are incomplete, immature or even that they are fairly pure meaning-only characters (examples of these types of characters are rare worldwide). These opinions neglect the large number of characters that are composed from sounds. From the many homophonic characters in the sub-categories of the Xixia dictionaries Homophones or *Wenhai* it can be seen that homophonic characters have a common shape - this shape is often the notation for the common sound. Some examples from the first type in Homophones (the bilabials) are:

Among the first subcategory of homophonic sounds *mi*, the following characters have similar shape: [illegible] and [illegible], [illegible] and [illegible].
Among the third subcategory of homophonic sounds *ming*, the following characters have similar shape: [illegible] and [illegible].
Among the fourth category of homophonic sounds *lao*, the following characters have similar shape: [illegible] and [illegible], [illegible] and [illegible].
Among the fifth subcategory of homophonic sounds *ming*, the following characters have similar shape: [illegible] and [illegible].

Also, within the first rhyme of *Wenhai*:

In the first subcategory of homophonic sounds *bu*, characters with similar shapes are [illegible] and [illegible], [illegible] and [illegible].
In the second subcategory of homophonic sounds *pu*, characters with similar shapes are [illegible], [illegible] and [illegible].
In the third subcategory of homophonic sounds *mou*, characters with similar shapes are [illegible] and [illegible].
In the fourth subcategory of homophonic sounds *dou*, [illegible] and [illegible] have a similar shape, [illegible] and [illegible] have a similar shape and [illegible] and [illegible] have a similar shape.

Careful translation and research of character construction methods enables a deeper understanding of the system and large quantity of sound-meaning combination characters. Several examples are listed below:

6.212	[illegible]	surname	*su*	[illegible]	surname		+	[illegible]		*su*
18.111	[illegible]	overgrown with weeds	*yi*	[illegible]	wood		+	[illegible]		*yi*
22.152	[illegible]	name of an insect	*da*	[illegible]		*da*	+	[illegible]	insect	
32.112	[illegible]	foreign transcription	*huo*	[illegible]	sound		+	[illegible]		*he*
35.212	[illegible]	eastern wren	*shi*	[illegible]		*shi*	+	[illegible]	birds	

36.272	[illegible]	grass name	*le*	[illegible]		*le*	+	[illegible]	grass name	
41.131	[illegible]	detest	*wu*	[illegible]		*wu*	+	[illegible]	loathe	
45.121	[illegible]	multitudinous, numerous	*yi jie*	[illegible]		*yi jie*	+	[illegible]	many	
55.151	[illegible]	tribute	*chang*	[illegible]	clothes		+	[illegible]		*wen*
63.262	[illegible]	insect name	*lu*	[illegible]			+	[illegible]	bee	
64.272	[illegible]	sprinkle; scatter	*chui*	[illegible]		*chui*	+	[illegible]	end	
80.151	[illegible]	name of a tree	*mu*	[illegible]	tree		+	[illegible]		*mu*
87.161	[illegible]	?	*yi ze*	[illegible]	griddle		+	[illegible]		*yi ze*
91.261	[illegible]	hurry	*chi*	[illegible]		*chi*	+	[illegible]	hurry	
6.261 Misc.	[illegible]	comment, talk	*zheng ni*	[illegible]	go, parade	*zheng ni*	+	[illegible]	discuss	
13.252 Misc.	[illegible]	cry		[illegible]	cover, surname		+	[illegible]	cry	

For the characters not in the *Wenhai* that has already been published, mainly the rising tone characters, there are also many formed from sound-meaning combinations. Despite not being explained by the Xixia people, we are still able to analyse them based on this construction principle. For example:

[illegible]	give	*yong*	[illegible]		*yong*	+	[illegible]	travel	
[illegible]	owl	*ge*	[illegible]		*ge*	+	[illegible]	ghost	
[illegible]	paste	*lu*	[illegible]	juice		+	[illegible]	store up, amass	*lu*
[illegible]	to ferry	*yu*	[illegible]	to ferry		+	[illegible]		*yu*

The sound markers in the Chinese shape-sound characters are relatively fixed, in contrast to the Xixia sound-meaning combination characters. The sound markers are mostly formed by leaving out a shape from another character, moreover when a character is used to construct others the part that is omitted is not fixed. As a result, both the sound and meaning markers are unstable. This point can be seen in the above examples. The Xixia sound-meaning combination characters in *Wenhai* are extremely prominent in some homophone subclasses. If they could be lined up together two characteristics could be seen: they are numerous and the sound markers are not fixed.

13.141-13.213 forms a homophonic group - a total of twelve characters have the sound *wei*: 1. [illegible] comprehend, 2. [illegible] dawn, 3. [illegible] conspire with, 4. [illegible] birds, 5. [illegible] six, 6. [illegible] fire, 7. [illegible] wooden materials in a room, 8. [illegible] fine, 9. [illegible] six, 10. [illegible] fourth heavenly stem, 11. [illegible] dragon, 12. [illegible] dragon-eye tree.

36.251-36.273 forms a homophonic group - a total of eight characters have the sound *le*: 1.  wind, 2.  pine tree, 3.  prestige, 4.  amber, 5.  sluggish, 6.  round bone, 7.  grass name, 8.  midday

73.171-73.221 forms a homophonic group - a total of five characters have the sound *ni*: 1. toil, 2.  drown, 3.  contract a disease, 4. trample, 5. to mate

81.161-81.232 forms a homophonic group - a total of ten characters have the sound *?*: 1.  firefly; glowworm, 2.  all; many; various, 3. world, 4.  star, 5.  melt, 6.  hate, 7. gadfly, 8.  ?, 9.  nod, 10. sigh

Among the sound-meaning combination characters there is a peculiar characteristic - some of the characters that express sound can also be used to express meaning. That is, there is also a relationship between the new character's meaning and the meaning of the character that expresses sound. This type can be called simultaneous sound-meaning combination characters.

9.261			zu		iron			pot	zu
9.262		dye	jue		blue-green	jue			se
12.231		taste	yu		taste			see	yu
14.143		bitter	ke		loathe	ke		taste	
15.161		basis	zhi		basis	zhi		guide	
15.251		wolf	shuo		tooth	chui		wild animal	
22.231		midnight (tomorrow)	na		midnight	na		pass	
23.143		healthy	ge			ge		power	shi
24.111		withered grass	yan		grass			thin	yan
29.142		stove	jue		iron			roast, bake	jue
29.232		deep water	hei		water			black	hei
71.161		reasonable	mo		see			wear	mo
71.211		spear	mo		iron			penetrate	mo
4.171 Misc.		boat			wood			to ferry	
4.272 Misc.		tweezers	ze ni		iron			catch	ze ni
15.272 Misc.		unit of area (qing)	ze ni		soil			gather	ze ni

From the above examples it can be seen that the meaning of the sub-character expressing sound together with the sub-character expressing meaning jointly determine the meaning of the newly formed character. It is possible that in the Xixia language, due to the stretching of the meaning of words when constructing characters, beautiful pot, blue/green, inspect, be satisfied, root, tooth, etc. have the meanings of small medicine pot, dye, taste, bitter, root and wolf. When these characters were constructed consideration was given to the original meanings, which also

explains the differences between them. It can be said that they are original. It is not hard to see that the authors of *Wenhai* clearly understood both the original and extended meanings of this type of character, the component sub-characters and the relationships between them. They correctly explained their formation.

There are also some characters formed from three component characters, among which two components express meaning and one component expresses sound. The two components that express meaning form the meaning of the new character, the other component character has the same sound as the new character. These should also be classified as a type of sound-meaning combination character. For example:

20.141		mourn	xi			xi		eye			cry		
61.251		first month	zhang			nian			zhang		get up		
68.231		horn	bing		head			horn				bing	
70.142		flower	zun		wonderful			beautiful	zun			guo	
84.141		mountain	luo		stone			Sanskrit				luo	
90.161		tremble	liu			liu		fear			cold		
2.132 Misc.		careful examination	geng		shepherd			see			announce		geng

(the last example is formed from four characters)

In total, the combination of meaning and sound-meaning combination methods form between 80-90% of the characters. These are the two main methods and should be given sufficient attention.

A.3 Initial-Final Characters at the Top and Bottom is a Special Way of Creating Xixia Characters

There are some Xixia characters that have a very peculiar construction method. The shape of these characters is formed by combining the parts or the whole of two component characters. The pronunciation is formed by joining the initial of the first character and the final of the second character. These kinds of characters in *Wenhai* are formed from two characters, the same as the phonetic notation for characters formed from the initial-final at the top and bottom. As a result, this method is called the initial-final top-bottom combination method. There are just over twenty of these characters in *Wenhai*:

14.161		transliteration	U [3]		zhu		U
15.121		sing	suan		gu		yi
20.211		transliteration	deng		dou		nan

[3]The code U is used for an entry that is unknown in the original (represented in the original by a question mark).

21.132		transliteration	min		ming		lin
21.271		transliteration	xun		xu		xun
21.272		Chinese surname	jin		ji		yin
21.273		Chinese surname	qin		qi		yin
24.212		transliteration	zhuo		dou		e
29.162		transliteration	jia		qi		jia
29.262		separate	chi-shuo		chi		shuo
32.113		Chinese surname	luan		lun		duan
35.131		transliteration	U		de		U
61.222		transliteration	U		wu		lu
63.131		transliteration	U		U		zhu
81.262		special instructions	Final of mou-ling		mou		ling
86.131		transliteration	U		U		U
91.112		transliteration	mao		ming		zhuang
9.212 Misc.		transliteration	wu		nong		wu
9.221 Misc.		transliteration			nong		U
9.222 Misc.		transliteration	a		a		e
19.261 Misc.		transliteration	U		nong		e
20.142 Misc.		transliteration	ya		nong		ha
93.121		transliteration	U		qi		niang

From the above examples it can be seen that these kinds of characters are mainly used in personal names, the phonetic transcriptions of geographic names or Buddhist texts, i.e. they are used to transcribe the sounds of foreign languages. Using the initial-final method to construct characters is used for both specialist phonetic transcription and to connect foreign words and sounds. In the original syllables of the Xixia language there wasn't a similar sound, it was only possible to use the initial-final method to spell the correct or a similar sound to the syllable. In *Wenhai* and *Homophones* they are all independent characters and don't have a homophonic character proof. Also, as many of these sounds use the *biao* sound and don't have a concrete meaning, if using combination of meaning and sound-meaning combination construction methods difficulties are met when adopting the parts for meaning. So we need to go down a new path and create another method of constructing characters, i.e. use the initial-final characters to form a new character.

In *Homophones* the concise explanation for these types of characters is not like the other characters which have a clear and concise meaning, but rather underneath the character is a list of the initial and final characters, reflecting its sound and shape. The authors of *Homophones* and *Wenhai* placed great importance on the initial-final method and have a special analysis of it. In *Homophones* the arrangement for this kind of character is already comparatively fixed - they are all placed in the last part of the section for the sounds of a particular type. The types are as follows: 1. Bilabials (12 characters), 2. Labiodentals (2 characters), 3. Palatals (19 characters), 4. Velars

(there aren't enough characters of this type - although there are 29, none are of the initial-final form), 5. Open Sounds (6 characters), 6. Dental-Affricates (6 characters), 7. Palato-Alveolars (5 characters), 8. Glottals (7 characters) and 9. Dentals (it so happens that the last part of the independent characters section is damaged, so it is not possible to read several characters). These compose roughly 1% of the Xixia characters. Below the initial-final characters in *Homophones* and *Wenhai* are shown:

Type	Homophones	Wenhai	Main	Sound	Initial	Sound	Final	Sound	Remarks
Bilabials	8.211	21.132		min		ming		lin	
	8.243	65.132				lu		bei	WH 
	9.116					shu		e	
	9.122					ming		ye	
	9.123					pi		ye	
	9.124					pi		yang	
	9.125					pi		yan	
	9.126					ye		ming	
	9.127					ming		bin	
	9.128					pi		ye	
	9.131			mian		ming		ye	
	9.132					ming		e	
	9.118	91.112		mao		ming			H
	9.162	81.262				mou		ling	H
Labiodentals	11.242	80.211				labiodental		bu	WH
	11.252	61.222				wu		lu	
Palatals	18.263	20.211		deng		dou		nan	
	18.264	33.252				dou		U	WH
	18.277	63.151				neng		dou	WH
	19.223					di		wu	
	19.231	30.241				ye		U	WH
	19.232	86.131				yiding		U	
	19.234					ni		yin	
	19.235					U		he	
	19.236					mi		U	
	19.237					di		U	
	19.238			ni		ning		ye	
	19.241			die		ning		ye	
	19.242					nu		dan	
	19.243					tie		ding yi	
	19.244					U		dou	
	19.245					di		ye	
	19.246					di		ye	
	19.264			?		ti		ye	PP

Type	Homophones	Wenhai	Main	Sound	Initial	Sound	Final	Sound	Remarks
	19.265		[illegible]	tui	[illegible]	tu	[illegible]	lei	
	19.228	24.212	[illegible]	zhuo	[illegible]	dou	[illegible]	e	H [illegible]
	28.231		[illegible]		[illegible]	ke	[illegible]	U	
Open Sounds	28.232		[illegible]	jiao	[illegible]	qi	[illegible]	U	
	28.233		[illegible]		[illegible]	wu	[illegible]	gan	
	28.234		[illegible]		[illegible]	jie	[illegible]	ye	
	28.235		[illegible]		[illegible]	yi	[illegible]	yin	
	28.236		[illegible]		[illegible]	yi	[illegible]	ye	
	25.238		[illegible]		[illegible]	gu	[illegible]	yi	H
	28.225		[illegible]		[illegible]	qi	[illegible]	jia	H [illegible]
	28.224		[illegible]		[illegible]	qi	[illegible]	niang	H [illegible]
Dental-Affricates	34.164	58.172	[illegible]	jiang	[illegible]	zi	[illegible]	U	WH [illegible]
	34.217	21.271	[illegible]	xun	[illegible]	xu	[illegible]	U	
	34.234		[illegible]	xian	[illegible]	si	[illegible]	yan	
	34.237		[illegible]	zi	[illegible]	zuo	[illegible]	chi	
	34.238		[illegible]		[illegible]	zei	[illegible]	luo	
	34.248		[illegible]	jie	[illegible]	ji	[illegible]	ye	
	34.223	21.272	[illegible]	jin	[illegible]	ji	[illegible]	yin	H [illegible]
	33.231	21.273	[illegible]	qin	[illegible]	qi	[illegible]	yin	H [illegible]
Palato-Alveolars	41.152		[illegible]		[illegible]	zhi	[illegible]	shan	
	41.153		[illegible]		[illegible]	ni zu yi	[illegible]	yan	
	41.154		[illegible]		[illegible]	zhi	[illegible]	jiao	
	41.155		[illegible]		[illegible]	ze ni	[illegible]	gan	
	41.156		[illegible]		[illegible]		[illegible]	ling	
	41.142	29.262	[illegible]		[illegible]	chi	[illegible]	shuo	H [illegible]
Glottals	46.124		[illegible]	xuan	[illegible]	xu	[illegible]	juan	
	46.125		[illegible]	xiao	[illegible]	xu	[illegible]		
	46.126		[illegible]		[illegible]	xi	[illegible]	ju	
	46.127		[illegible]		[illegible]	xi	[illegible]	ye	
	46.128		[illegible]		[illegible]	wei	[illegible]	juan	
	46.131		[illegible]	xie	[illegible]	xu	[illegible]	ye	
	46.137		[illegible]		[illegible]	xu	[illegible]	ye	
	45.277	9.212 M	[illegible]	wu	[illegible]	nong	[illegible]	wu	H [illegible]
	46.111	9.221 M	[illegible]	?	[illegible]	nong	[illegible]	U	H [illegible]
	46.113	9.222 M	[illegible]	a	[illegible]	a	[illegible]	e	H [illegible]
	46.117	19.261 M	[illegible]		[illegible]	yi	[illegible]	e (light)	H [illegible]
	46.115	20.142 M	[illegible]	ya	[illegible]	yi	[illegible]	ha	H [illegible]

Type	Homophones	Wenhai	Main	Sound	Initial	Sound	Final	Sound	Remarks
Dentals	55.265	32.113		luan		lun		duan	H
	55.235	63.131						zhu	H

Abbreviations: M = Miscellaneous, WH = *Wenhai*, H = *Homophones* annotation character, PP = *Pearl in the Palm*

The initial-final method has a special significance. It shows that in the twelfth century the Xixia area had already adopted a method of spelling to create characters. Although there aren't many of these characters, they are mainly phonetic characters and the sound components aren't fixed, they take a step towards characters that reflect spelling. This point is hardly praiseworthy. After Sun Yan from Wei (in the Three Kingdoms period) started to use the initial-final method, this became a traditional phonetic method in the study of Chinese rhymes. However, the use of the initial-final method to jointly analyse sounds and character shapes, namely using a method resembling spelling to construct characters, originated with the Xixia characters. There are a much lower proportion of Xixia sound-meaning combination characters than the related shape-sound characters of Chinese and the make-up of characters reflecting sounds are not as developed as those of Chinese. However, this was an early attempt at the creation of characters that reflect spelling.

A.4 Indirect Sound-Meaning Combinations Reflect the Influence of Chinese Character Construction Methods. Long Sound Combinations Record the Long Sounds of Sanskrit

There is also a small set of Xixia characters where part of the shape is formed from meaning but the other part doesn't reflect either meaning or a direct representation of sound. This part reflects the sound of the character after it has been translated into Chinese. We call these indirect sound-meaning combination characters, for example:

21.142		surname	jin		gold		surname
21.211		place name	shen		tree		god
21.242		place name	feng		auxiliary word		divide
21.252		person's name	lun		sound		wheel
22.252		transliteration	nan		place		south
31.222		transliteration	han		Han (surname)		surname
55.121		transliteration	niu		surname		cow

There are not many characters of this type and for the most part they reflect borrowed words from Chinese, the names of Chinese people or Chinese placenames. This is a specialised method of creating characters from borrowed Chinese words. It clearly shows that the Dangxiang minority group of the Hexi region had close contact over a long period of time with other ethnic groups (especially Chinese). This is one aspect of the enormous influence they had on the Xixia language

and writing system. From the *Wenhai* and other material in the Xixia script the large quantity of borrowed words from Chinese can be seen. Among these are commonly used personal names, place names and basic vocabulary. As these Chinese borrowings were used with high frequency and over a broad range they had already become part of Xixia vocabulary. The Xixia people created independent characters for these words and also acknowledged their existence in their rhyme books. They had already formed an inseparable part of the Xixia language. The indirect sound-meaning combination characters show that Chinese not only had a big influence on the Xixia language, but also had an influence on the construction of Xixia characters.

Long sound combinations are another construction method that has been influenced by a foreign language. A large number of Buddhist works were translated into Xixia. The incantantions in the Buddhist scriptures had Sanskrit long vowels - these weren't present in the Xixia rhymes. The Xixia people created a few Xixia characters to record these long vowels fairly accurately. The shape of one part of these characters was similar to the Sanskrit letters that recorded these sounds. The shape of the other part was formed from part of the Xixia character meaning *long*, showing that this character was read as a long vowel. For example:

Wenhai No.	Char.	Sound	Sound Char.	Sound	Long Char.	Meaning
19.262 Misc.	[illegible]	a (long shout)	[illegible]	a	[illegible]	long
19.271 Misc.	[illegible]	yi (long)	[illegible]	yi	[illegible]	long
19.272 Misc.	[illegible]	wu	[illegible]	wu	[illegible]	long
20.111 Misc.	[illegible]	he	[illegible]	a	[illegible]	long

These characters are all classified as glottals, are in the *Wenhai* Miscellaneous Types and are all independent characters. They represent sounds from the Buddhist scriptures. It can be seen that translation of Buddhist works has also had some influence on the construction of Xixia characters.

A.5 The Exchange of Parts is One Method of Creating Characters of Similar Meaning

In researching the construction of Xixia characters, one phenomenon often comes to our attention. This is that the left and right components of a Xixia character and the left and right components of another character happen to be opposite to each other. The meanings of these two characters often have a close relationship, even to the point that they often form a word or a word grouping when connected together. These types of characters are formed by swapping the left and right components, so this construction method is called the exchange method. There are many characters of this type, but when their construction is explained in *Wenhai* an exchange of parts isn't referred to. *Wenhai* still uses the normal combination method to analyse these specially constructed characters. Under these circumstances, it was quite difficult to reach our conclusion. The explanation of only one character's shape suitably reflected the principle of this construction method:

51.141 [illegible] (beast) is constructed from:
[illegible] beast
[illegible] character
[illegible] prefix
[illegible] exchange

The rough meaning of the character can be obtained from an exchange of the two sides of the first explanatory character. Although this is only one example, it is possible to see that the *Wenhai* authors noticed this special construction method. It is unfortunate that the authors didn't use this method to analyse these characters. In order to understand the characteristics of this type of character, we might as well take a list of examples from the Xixia literature as follows:

1. Left and right parts swap

Char 1	Meaning	Char 2	Meaning
	finger		toe
	break		break
	enemy		bandit
	light, ray, bright		bask in the sun
	luxuriant		flourishing
	slander		slander
	male		son
	give birth		multitude
	discrepancy		distinguish
	wave		pass
	dog		bark
	release		dispatch
	person		person
	break		drop
	bite		gnaw
	discuss		discuss
	soil		earth/female principle
	firm, fast		?
	deliver		match
	evil		to speak
	small		small
	looks		attitude
	swallow (bird)		little bird
	indignant		indignant
	mucus, tears		saliva
	steal		steal
	rows		rows
	boil		boil
	dirt		greasy
	poor		to the extreme
	breast		house
	regret		change
	revolve		wind
	slow, gentle		slow, gentle

Char 1	Meaning	Char 2	Meaning
𥘅	strong	𣏌	enemy

2. Middle stays fixed, left and right components swap

Char 1	Meaning	Char 2	Meaning	Full Meaning
𧾷	reason	𧿹	reason	
𧹞	beast	𩙿	beast	
𧼨		𧽧		pour into
𧺆	place	𨈆	break	

3. Top stays fixed, left and right components of the bottom swap

Char 1	Meaning	Char 2	Meaning	Full Meaning
𦓐	plant root used in medicine	𣲘	to hide	
𧛓		𣲖		coarse woollen shirt
𦾔		𦉫		please come

4. One side stays fixed, left and right components of the other side swap

Char 1	Meaning	Char 2	Meaning
𤇾	claw	𩠐	one-sided
𣲑	water	𣳾	black
𤉣	contract	𤊓	wrinkle
𠯿	exclude	𠕋	submit
𤋮	white stain	𤉡	land, soil

5. Top and bottom parts swap

Char 1	Meaning	Char 2	Meaning
𦎫	burn, cook	𠔃	roast, bake
𦊓	branch	𠕁	leaf

Where two characters constructed using the exchange method above have similar meaning they

often form two-syllable words. The explanations in *Wenhai* and *Homophones* are often mutually annotated to form mutually instructive characters. This reflects the tendency for two-syllable words to be more developed in the Xixia language. This point is very close to Chinese. However, in Chinese there are very few characters constructed by swapping the left and right components. There are some other Xixia characters that have swapped left and right components but generally don't have related meanings - they aren't formed by the exchange method. For example,  (do not) and  (reside).

There are several types of relationships between the Xixia characters formed by the exchange method. Sometimes the two characters can each be used independently. In other cases only one of the two characters can be independently used, the other character is not free to create words with further characters - it is only used in this two-character combination. There are also cases where one character is a word that is an intrinsic part of the Xixia language and the other character is a word borrowed from Chinese. Finally, there are cases where the two characters can't be used independently - they only make sense when joined together. These are two-syllable simple words or onomatopoeia.

The characters formed by the exchange method make it more convenient to understand the meaning of other characters. Consider the case where the meaning of a character is not clear but the corresponding character can be found with the left and right components swapped and this character has a clear meaning. After confirming the relationship between the characters and other conditions, it can generally be said that the unclear character's meaning is either similar or related.

There are some characters where the left and right components are the same. Some of these characters have the meaning of *double*. For example:

Character	Meaning
	gather
	lip
	pair, both
	thorn
	distribute
	run quickly
	lucky
	surname
	glare with anger
	hills
	auxiliary word

Some characters also have equal left and right components, but a vertical line is placed in the middle. These characters often have meanings related to *middle* or *to thread*. The middle vertical line has an indicative flavour. For example:

Character	Meaning
	weigh, estimate

Character	Meaning
	middle, among
	string of coins
	string
	furrow
	side, border
	to stand
	between, among
	fine horse
	milk an animal
	vast, wide
	similar, resemble
	surname

A.6 The Meaning and Radicals of Characters Similar in Form

There are many Xixia characters of similar shape, researching these helps improve the ability to recognise Xixia characters. Xixia characters have complicated shapes and many characters are similar but not quite the same. The differences between some characters are tiny and it is easy to make mistakes in reading or writing them. It is even more important to notice this in some documents where the printing is not clear. Overall the Xixia characters with similar shapes can be divided into three classes:

1. There is a relationship between the meanings of the characters. Both of the components in meaning combination characters or the meaning character in sound-meaning combinations can have this sort of relationship with the new character. It is also possible for the indirect shifted combinations to have this sort of relationship.

2. The sound of the characters is either the same or similar. Many of these sorts of characters have come from the sound characters of sound-meaning combinations. Initial-final combination characters can also form this sort of relationship.

3. Although the shapes are similar there is no obvious relationship between them in either sound or meaning. The formation of these types of characters probably has no connection with differences in either sound or meaning. There are also some shapes formed from identical characters.

Especially worthy of notice are characters of similar shape that have a relationship between their meanings; they often join with each other to form juxtaposed word groups. These are among the more commonly used Xixia words and are worthy of our notice. The differences are illustrated in the examples below:

1. Relationship between the meanings

Char 1	Meaning	Char 2	Meaning	Char 3	Meaning
[illegible]	recover, duplicate	[illegible]	cover		
[illegible]	fetters	[illegible]	fetters, manacles		
[illegible]	dispatch	[illegible]	dispatch		
[illegible]	call	[illegible]	call		
[illegible]	leave behind	[illegible]	lose		
[illegible]	hear	[illegible]	know		
[illegible]	grass	[illegible]	grass		
[illegible]	just, fair	[illegible]	mother's brother		
[illegible]	complain	[illegible]	hate		
[illegible]	platform	[illegible]	west		
[illegible]	first	[illegible]	begin		
[illegible]	ask	[illegible]	handsome		
[illegible]	consume, disappear	[illegible]	exhaust		
[illegible]	big wooden tray	[illegible]	measure word for lamps		
[illegible]	hungry	[illegible]	belly, abdomen, stomach		
[illegible]	sharpen, grind	[illegible]	revolve		
[illegible]	bird name	[illegible]	birds		
[illegible]	time of day	[illegible]	day		
[illegible]	complain	[illegible]	have	[illegible]	complain
[illegible]	scold, abuse	[illegible]	?	[illegible]	scold, abuse

2. Characters with the same sound

Char 1	Meaning	Char 2	Meaning	Char 3	Meaning	Sound
[illegible]	wave	[illegible]	truth	[illegible]	pig	yu
[illegible]	surname	[illegible]	tree name			niu
[illegible]	axe	[illegible]	dregs, spoil			chi
[illegible]	walk	[illegible]	sweep			shu

3. No relationship in either sound or meaning

Char 1	Meaning	Char 2	Meaning
[illegible]	this	[illegible]	tall
[illegible]	according to	[illegible]	small
[illegible]	divide, distribute	[illegible]	fragrant flower
[illegible]	marriage	[illegible]	treasure, cherish
[illegible]	hat	[illegible]	gather
[illegible]	degenerate, fall	[illegible]	fall ill

From the above it can be seen that characters with similar shapes might have a relationship between their meanings, a relationship between their sounds or no relationship between either sounds or meanings.

The removal of a shape or a sound in the construction of Xixia characters is very common. There are no fixed rules for forming a character by removing a portion from another and there are no other signs as to which part represents sound or meaning. The Chinese shape-sound characters generally have the shape portion at the top and the sound portion at the bottom, or the shape at the left and the sound at the right. Xixia characters don't have such a prominent feature. When constructing Xixia characters, the same shapes at the same positions can have a fixed meaning, arising from the shapes having the same role. It is also possible for them to have differing meanings, reflecting a completely different role. So sorting out which radicals have a fixed meaning is relatively difficult. For example, when forming characters from 𗍫 (meaning to have or possess, pronounced *liu ni*), there are at least the following types, dependent on which part is used and its meaning:

65.171 𗍬 (surname, pronounced *?*) is formed from 𗍫 (have, pronounced *liu ni*) and 𗍫 (surname), with 𗂧 expressing sound.
89.142 𗍭 (filled) is formed from 𗍫 (have, pronounced *liu ni*) and 𗍮 (filled), with 𗂧 expressing meaning.
30.133 𗍯 (doubt) is formed from 𗍰 (fear) and 𗍫 (have, pronounced *liu ni*), with 𗃵 expressing meaning.
50.162 𗍱 (surpass) is formed from 𗍫 (have, pronounced *liu ni*) and 𗍲 (a hundred million), with 𗁅 expressing meaning.
92.261 𗍳 (not) is formed from 𗍫 (have, pronounced *liu ni*) and 𗍴 (overflow) with 𗍵 expressing meaning.

The formation of the above examples all start with the same 𗍫 character. After removing parts of the character they form several different character shapes. Some of these express sound, some of them express meaning and some of them can express either sound or meaning. So when analysing the construction of characters it is not possible to restrict one type of shape to represent one function. It is important to take a broader view and think about which characters a shape can be removed from. For example:

𗏁 , 𗏂 and 𗏃 could all be related to 𗏃 (female, pronounced *ming*).
𗍵, 𗐯 and 𗐰 could all be related to 𗐱 (stone, pronounced *lu*).
𗍶 , 𗐲 and 𗐳 could all be related to 𗐳 (hand, pronounced *?*).
𗐴 , 𗐵 and 𗐶 could all be related to 𗐶 (clothing, pronounced *wei*).
𗐷 , 𗐸 and 𗐹 could all be related to 𗐹 (heart or mind, pronounced *ni*).

There are quite a lot of Xixia characters that have the same shape in the same position. Among these there are some characters that have the same meaning - these shapes seem to have the function of a radical. However, there are not many of this type of shape in the Xixia characters. Carefully analysing the source of this type of character's identical shapes and its function in the construction of the character, there are myriad kinds and they are very complicated. They are not as strong as the representative power of Chinese radicals. For example, there are 320 characters with the 𗍺 shape. Among these there are roughly 200 characters that have a relationship to *trees* or *wood*. The remaining 120 characters don't have any relationship to these meanings. These remaining characters have a variety of meanings - actions, surnames, place names, livestock and

auxiliary words. However, the shape has a relatively focussed meaning - it is still possible to say that this shape is a wood/tree radical. These kinds of shapes include the water radical  , the sound radical  , the earth radical  , etc. If the remaining shapes that occur relatively often were to be analysed it would be discovered that they aren't really representative of anything. Even if an attempt is made to abstract a stronger meaning, this is very difficult if not impossible. For example:  ,  , ,  , ,  ,  ,  ,  ,  ,  and  .

In order to advance the explanation of Xixia characters containing identical shapes, quite different phenomenon are needed. Examine the following examples:

 can come from  (stone, pronounced *lu*) or (have, pronounced *liu ni* - see previous page).  (stone roller, pronounced *te*) (is formed from) the addition of  (stone) and  (garden roller).

 can come from (see, pronounced *ling*) or  (yes; to be, pronounced *wu*). (dream, pronounced *ming*) is formed from the addition of  (night) and  (see). (not, pronunciation unknown) is formed from the addition of (no) and (yes; to be).

 can come from  (no, pronounced *ming*) or  (repeat, pronounced *niang*).  (grand, pronounced *ma*) is formed from the addition of  (no) and  (fine, slender). (prohibit, pronounced *chi*) is formed from the addition of (repeat, duplicate) and  (cut short, disappear).

Through research on the *Wenhai* we can see a related, very significant phenomenon. Two Xixia characters can be different if they have differing shapes and positions. This is an even stronger explanation of new character formation and the removed shape can be of more than one type. For example:

45.131	 (fast)	+	(sound)	⟶	(disease)
45.132	(sound)	+	(fast)	⟶	(fast)
29.232	(water)	+	(black)	⟶	(deep water)
5.262	(water)	+	(black)	⟶	(mud)
69.261	(bird)	+	(black)	⟶	(crow)
65.271	(bird)	+	(black)	⟶	(vulture)
89.261	(true)	+	(heavy)	⟶	(heavy, cumbersome)
89.132	(heavy)	+	(true)	⟶	(heavy)
10.211	(true)	+	(heavy)	⟶	(dusky)

From the above analysis we can see fairly clearly that using the idea of radicals from Chinese to explain the formation of Xixia characters is often quite difficult. As a result, when analysing the formation of Xixia characters the rules need to be used flexibly to give a definitive analysis of a character. It is quite acceptable to use shapes to act as radicals for the purpose of easily working with the character set. This is a separate issue from analysing the construction principles of the characters.

It is also worthwhile to examine the analysis of Xixia people towards those shapes that appear to have meaning - it seems as if they also had the idea of radicals (note 6). This is not a universal principle and there are a lot of limitations. It is limited to several types - person, grass, earth,

negatives, etc. These shapes also only have this meaning in some of the characters in which they appear - they have a different meaning in other characters.

A.7 The Flexibility of Shapes When Forming Xixia Characters

When using combination methods to form Xixia characters, there is another phenomenon that should be noticed. This is that a given shape can change in different characters. Researching these special characteristics has immediate application to a clear-cut construction method for the characters, understanding the meaning a bit better and standardising the calligraphy of certain characters. The changes in shape when Xixia characters are constructed can be subdivided into three types.

The first type is changes that occur when the shapes are in different positions. For example, if the left of the character for *earth* is used in character construction, it can be placed on the left, on the right or in the middle of the new character and its shape won't change. If it forms the top of the new character, although the strokes don't change, in order to suit the changed position the shape flattens. For example:

1. The common shape

No.	Char 1	Meaning	Char 2	Meaning	Char 3	Meaning
66.242		sprout		land, soil		penetrate
14.251 Misc.		to plant		land, soil		to plant
22.252		place name (nan)		land, soil		south
20.232		tomb, hill, mountain		tall		land, soil
28.151		land (ma)		(ma)		land, soil
91.122		mud		water		land, soil

2. Change in the shape 

No.	Char 1	Meaning	Char 2	Meaning	Char 3	Meaning
10.221		valley (gu)		land, soil		narrow
75.161		bury		land, soil		below
83.262		drought		land, soil		fire
86.121		hell		land, soil		prison
86.271		shake, shock		land, soil		move, change

The second type of shape change can't be simply regarded as arranging the shapes as one pleases, but rather has a specific meaning. The meaning of these types of characters are often related to *underground*. As a result, when characters are constructed the saved part of the character for *earth* is placed on the top of the new character, prominently indicating the *underground* meaning. It is important to understand this point. Another example is that using the left portion of the Xixia character for *water* to create a new character can result in two different shapes:

1. The common shape [illegible]

No.	Char 1	Meaning	Char 2	Meaning	Char 3	Meaning
77.261	[illegible]	reed	[illegible]	water	[illegible]	grass
91.122	[illegible]	mud	[illegible]	water	[illegible]	land, soil

2. The first changed shape [illegible]

No.	Char 1	Meaning	Char 2	Meaning	Char 3	Meaning
5.262	[illegible]	mud	[illegible]	water	[illegible]	black

3. The second changed shape [illegible]

No.	Char 1	Meaning	Char 2	Meaning	Char 3	Meaning
39.261	[illegible]	bent, curved	[illegible]	bent	[illegible]	place name

Apart from these there are also some changes that can occur when a shape is on the left or right of a character. For example:

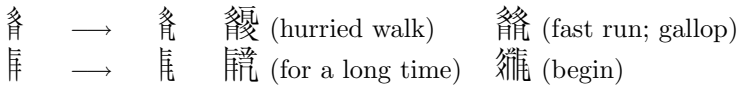
On the left:
[illegible] [illegible] [illegible]
On the right:
[illegible] [illegible] [illegible]

The second type is where the strokes change. This change is often done for aesthetic reasons and to make the character easier to write. It is desirable to smooth out the shape of some of the simple characters that have a lower number of strokes. As a result, the ends of some strokes are turned into hooks. However, when they are used to form new characters with an increased number of strokes the hooks are often straightened. Some shapes also have a hook at the end of the tail when placed on the right. When placed on the left, however, the aesthetics of the hook are destroyed, so the hook is straightened. For example:

1. The shape of a simple character changes when placed in a combination character:

[illegible] (belt)	⟶	[illegible]	[illegible] (strip)	[illegible] (tie)	
[illegible] (hard work)	⟶	[illegible]	[illegible] (blame)	[illegible] (bitter)	[illegible] (prison)
[illegible] (tall)	⟶	[illegible]	[illegible] (storied building)	[illegible] (block, obstruct)	[illegible] (tomb, hill)
[illegible] (disobey)	⟶	[illegible]	[illegible] (absurd; fantastic)	[illegible] (to deceive)	[illegible] (entice, tempt)
[illegible] (half)	⟶	[illegible]	[illegible] (nothing)	[illegible] (chop)	[illegible] (to rob)

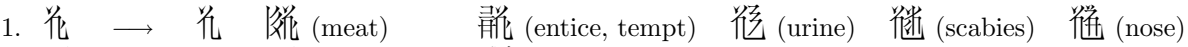
2. The shape changes when placed on the left or the right:

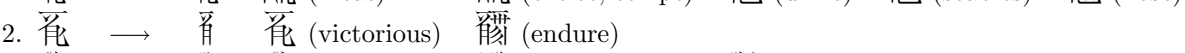


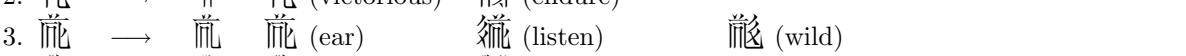
3. The shape moves to the left, but the tail is stretched back to the right:

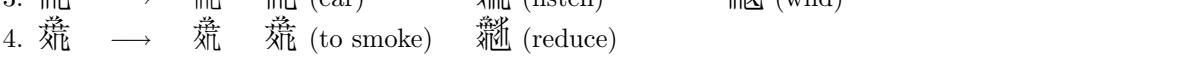
Character	Meaning
	sound
	transliteration character
	ghost
	demon
	have, possess
	to exchange
	life
	trample
	trample

The third type of shape change is dropping a stroke when a shape moves from the right to the left. This is done in order to avoid confusion with other strokes on the right. There is also often a straightening of a hook in conjunction with this change. For example:

1. 

2. 

3. 

4. 

A.8 Notes

1. See: Bernhard A. und E.von Zach *Einige Bemerkungen* Über *Si-hia Schrift und Sprache* Ostasiatische zeitschrift 7 1909; Luo Fuchang *Overview of Xixia Books*, Dong Shan Social Education Press, 1914, pp.3-19; Nishida Tatsuo *Research on the Xixia Language*, Zauho Kankokai, 1966, Section 2, pp.225–252; Sofronov *Xixia Grammar*, Moscow, 1968, pp.43–68.
2. Shi Jinbo, "An Overview of the Structure of Xixia Characters" in *Minority Language and Literature Collection* No.1, 1980.
3. Three characters are only given in the explanation of 7.112 *to climb*, i.e. "knee hand walk".
4. See Li Tao, *Jizi Zhitong ? Changpian*, scroll 29, p.16, 1881, Zhejiang Publishing House block-printed edition; Zeng Gong, *Longping Ji*, scroll 20, p.5, 1708 block-printed edition; Shen Kuo, *Dream Pool Essays*, scroll 25, p.3, 1978, Historical Relics Press, photocopy of the Dong Shan Academy 1305 block-printed edition.
5. See Luo Fuchang *Overview of Xixia Books*, p.5.
6. See *Wenhai* 21.171, Misc. 18.211, 27.113, 60.212, 85.122, Misc. 2.171, Misc. 15.172, Misc. 6.171 and 23.162.

Appendix B

Xixia Recursive Radical Codes

B.1 Ordered by Transliteration

ba	120		fa	104		gaa	89		hu	70		loo	157	
be	55		fe	63		gee	21		haa	98		ma	1	
bi	79		fi	102		gii	87		hee	88		me	3	
bo	85		fo	18		goo	57		hii	105		mi	44	
bu	110		fu	23		guu	91		hoo	74		mo	4	
baa	139		faa	27		gae	58		huu	26		mu	32	
ca	11		fee	92		gai	77		hae	108		maa	119	
ce	15		fii	94		gao	136		hai	150		mee	122	
ci	41		foo	152		gau	128		hao	147		mii	162	
co	30		fuu	146		gea	99		hau	148		moo	161	
cu	158		fae	171		gei	71		ja	50		muu	137	
caa	73		fai	115		geo	145		ka	72		na	5	
cee	90		fao	132		geu	155		la	17		ne	67	
cii	123		fau	153		gia	159		le	6		ni	20	
coo	125		fea	176		gie	166		li	8		no	81	
cuu	135		ga	25		gio	175		lo	69		nu	84	
cae	156		ge	64		ha	52		lu	36		naa	163	
da	9		gi	101		he	62		laa	65		nee	93	
de	31		go	16		hi	61		lee	133		nii	118	
di	144		gu	2		ho	49		lii	151		noo	83	

nuu	117		si	39		tia	168		xu	112	
nae	130		ta	45		tie	169		xaa	96	
pa	33		te	48		va	34		xee	68	
pe	97		ti	51		ve	56		xii	53	
pi	126		to	43		vi	47		xoo	95	
po	12		tu	29		vo	107		xuu	154	
pu	172		taa	60		vu	141		xae	100	
paa	106		tee	10		wa	37		xai	127	
pee	140		tii	121		we	78		xao	109	
pii	142		too	114		wi	40		xau	116	
qa	35		tuu	46		wo	24		xea	138	
qe	19		tae	143		wu	75		ya	28	
qi	160		tai	167		waa	76		ye	131	
ra	80		tao	59		wee	103		za	66	
re	86		tau	129		wii	173		ze	42	
ri	82		tea	124		xa	13		zi	174	
ro	149		tei	164		xe	14		zo	134	
sa	22		teo	170		xi	7		zu	111	
se	38		teu	113		xo	54		zaa	165	

Classifications * horizontal * vertical * straight legs * curly tail * tail support * head * unclassified

B.2 Original Order

1	ma		21	gee		41	ci		61	hi		81	no	
2	gu		22	sa		42	ze		62	he		82	ri	
3	me		23	fu		43	to		63	fe		83	noo	
4	mo		24	wo		44	mi		64	ge		84	nu	
5	na		25	ga		45	ta		65	laa		85	bo	
6	le		26	huu		46	tuu		66	za		86	re	
7	xi		27	faa		47	vi		67	ne		87	gii	
8	li		28	ya		48	te		68	xee		88	hee	
9	da		29	tu		49	ho		69	lo		89	gaa	
10	tee		30	co		50	ja		70	hu		90	cee	
11	ca		31	de		51	ti		71	gei		91	guu	
12	po		32	mu		52	ha		72	ka		92	fee	
13	xa		33	pa		53	xii		73	caa		93	nee	
14	xe		34	va		54	xo		74	hoo		94	fii	
15	ce		35	qa		55	be		75	wu		95	xoo	
16	go		36	lu		56	ve		76	waa		96	xaa	
17	la		37	wa		57	goo		77	gai		97	pe	
18	fo		38	se		58	gae		78	we		98	haa	
19	qe		39	si		59	tao		79	bi		99	gea	
20	ni		40	wi		60	taa		80	ra		100	xae	

101	gi		120	ba		139	baa		158	cu	
102	fi		121	tii		140	pee		159	gia	
103	wee		122	mee		141	vu		160	qi	
104	fa		123	cii		142	pii		161	moo	
105	hii		124	tea		143	tae		162	mii	
106	paa		125	coo		144	di		163	naa	
107	vo		126	pi		145	geo		164	tei	
108	hae		127	xai		146	fuu		165	zaa	
109	xao		128	gau		147	hao		166	gie	
110	bu		129	tau		148	hau		167	tai	
111	zu		130	nae		149	ro		168	tia	
112	xu		131	ye		150	hai		169	tie	
113	teu		132	fao		151	lii		170	teo	
114	too		133	lee		152	foo		171	fae	
115	fai		134	zo		153	fau		172	pu	
116	xau		135	cuu		154	xuu		173	wii	
117	nuu		136	gao		155	geu		174	zi	
118	nii		137	muu		156	cae		175	gio	
119	maa		138	xea		157	loo		176	fea	

Classifications * **horizontal** * **vertical** * **straight legs**
* **curly tail** * **tail support** * **head** * **unclassified**

B.3 Ordered by Type

gee 21-2		fii 94-4		di 144-3		naa 163-4		lo 69-1	
ga 25-3		fi 102-4		moo 161-3		hoo 74-5		lee 133-1	
ge 64-3		fao 132-4		mii 162-3		hii 105-5		tao 59-2	
geu 155-3		fae 171-4		me 3-4		hae 108-5		we 78-2	
gu 2-4		fu 23-5		mee 122-4		nii 118-5		xai 127-2	
goo 57-4		ra 80-5		pu 172-4		ne 67-6		loo 157-2	
gii 87-4		ro 149-5		mu 32-5		hai 150-6		tie 169-2	
gaa 89-4		fea 176-5		ye 131-5		nee 93-7		teo 170-2	
gao 136-4		re 86-6		na 5-2		laa 65-1		tee 10-3	
go 16-5		fai 115-6		he 62-3		gi 101-2		po 12-3	
gae 58-5		fuu 146-6		noo 83-3		lii 151-2		pa 33-3	
guu 91-5		foo 152-6		nu 84-3		gai 77-4		ta 45-3	
gie 166-5		fau 153-6		hao 147-3		gea 99-4		te 48-3	
gau 128-6		ma 1-1		ho 49-4		baa 139-4		wu 75-3	
geo 145-6		de 31-2		ha 52-4		gia 159-4		waa 76-3	
fa 104-2		maa 119-2		hi 61-4		gei 71-5		too 114-3	
fo 18-3		muu 137-2		hu 70-4		bu 110-5		hau 148-3	
faa 27-3		cu 158-2		haa 98-4		qi 160-5		tei 164-3	
ri 82-3		mo 4-3		nuu 117-4		gio 175-6		tia 168-3	
fee 92-4		pi 126-3		nae 130-4		la 17-1		ti 51-4	

* gee 21-2 means the radical with transliteration *gee*, number 21 and a stroke count of 2

taa 60-4		se 38-2		wi 40-3		xee 68-4	
ka 72-4		ci 41-2		ze 42-3		no 81-4	
teu 113-4		to 43-2		mi 44-3		bo 85-4	
tii 121-4		vi 47-2		tuu 46-3		hee 88-4	
tea 124-4		ja 50-2		cee 90-3		xaa 96-4	
pii 142-4		za 66-2		xoo 95-3		pe 97-4	
zaa 165-4		caa 73-2		paa 106-3		xae 100-4	
tai 167-4		vo 107-2		coo 125-3		wee 103-4	
tu 29-5		xao 109-2		cuu 135-3		xu 112-4	
tau 129-5		zu 111-2		xea 138-3		cii 123-4	
tae 143-5		xau 116-2		cae 156-3		zo 134-4	
le 6-1		vu 141-2		wii 173-3		pee 140-4	
li 8-1		zi 174-2		xe 14-4		qe 19-5	
da 9-1		xa 13-3		ni 20-4		xo 54-5	
sa 22-1		wo 24-3		ya 28-4		ve 56-5	
xi 7-2		huu 26-3		va 34-4		fe 63-5	
ca 11-2		co 30-3		qa 35-4		bi 79-5	
ce 15-2		wa 37-3		xii 53-4		ba 120-5	
lu 36-2		si 39-3		be 55-4		xuu 154-5	

Classifications * horizontal * vertical * straight legs * curly tail * tail support * head * unclassified

B.4 Ordered by Stroke Count

ma 1-1		xai 127-2		zi 174-2		te 48-3		coo 125-3	
laa 65-1		loo 157-2		ga 25-3		wu 75-3		cuu 135-3	
la 17-1		tie 169-2		ge 64-3		waa 76-3		xea 138-3	
lo 69-1		teo 170-2		geu 155-3		too 114-3		cae 156-3	
lee 133-1		xi 7-2		fo 18-3		hau 148-3		wii 173-3	
le 6-1		ca 11-2		faa 27-3		tei 164-3		gu 2-4	
li 8-1		ce 15-2		ri 82-3		tia 168-3		goo 57-4	
da 9-1		lu 36-2		mo 4-3		xa 13-3		gii 87-4	
sa 22-1		se 38-2		pi 126-3		wo 24-3		gaa 89-4	
gee 21-2		ci 41-2		di 144-3		huu 26-3		gao 136-4	
fa 104-2		to 43-2		moo 161-3		co 30-3		fee 92-4	
de 31-2		vi 47-2		mii 162-3		wa 37-3		fii 94-4	
maa 119-2		ja 50-2		he 62-3		si 39-3		fi 102-4	
muu 137-2		za 66-2		noo 83-3		wi 40-3		fao 132-4	
cu 158-2		caa 73-2		nu 84-3		ze 42-3		fae 171-4	
na 5-2		vo 107-2		hao 147-3		mi 44-3		me 3-4	
gi 101-2		xao 109-2		tee 10-3		tuu 46-3		mee 122-4	
lii 151-2		zu 111-2		po 12-3		cee 90-3		pu 172-4	
tao 59-2		xau 116-2		pa 33-3		xoo 95-3		ho 49-4	
we 78-2		vu 141-2		ta 45-3		paa 106-3		ha 52-4	

* gee 21-2 means the radical with transliteration *gee*, number 21 and a stroke count of 2

hi 61-4		xe 14-4		go 16-5		tae 143-5	
hu 70-4		ni 20-4		gae 58-5		qe 19-5	
haa 98-4		ya 28-4		guu 91-5		xo 54-5	
nuu 117-4		va 34-4		gie 166-5		ve 56-5	
nae 130-4		qa 35-4		fu 23-5		fe 63-5	
naa 163-4		xii 53-4		ra 80-5		bi 79-5	
gai 77-4		be 55-4		ro 149-5		ba 120-5	
gea 99-4		xee 68-4		fea 176-5		xuu 154-5	
baa 139-4		no 81-4		mu 32-5		gau 128-6	
gia 159-4		bo 85-4		ye 131-5		geo 145-6	
ti 51-4		hee 88-4		hoo 74-5		re 86-6	
taa 60-4		xaa 96-4		hii 105-5		fai 115-6	
ka 72-4		pe 97-4		hae 108-5		fuu 146-6	
teu 113-4		xae 100-4		nii 118-5		foo 152-6	
tii 121-4		wee 103-4		gei 71-5		fau 153-6	
tea 124-4		xu 112-4		bu 110-5		ne 67-6	
pii 142-4		cii 123-4		qi 160-5		hai 150-6	
zaa 165-4		zo 134-4		tu 29-5		gio 175-6	
tai 167-4		pee 140-4		tau 129-5		nee 93-7	

Classifications *** horizontal** *** vertical** *** straight legs**
*** curly tail** *** tail support** *** head** *** unclassified**

B.5 Ordered by Frequency

ma 1-1		ha 52-4		ne 67-6		ve 56-5		vo 107-2	
xa 13-3		xo 54-5		ze 42-3		xii 53-4		fee 92-4	
de 31-2		he 62-3		caa 73-2		tu 29-5		mee 122-4	
sa 22-1		wi 40-3		ni 20-4		mo 4-3		tii 121-4	
xe 14-4		za 66-2		hi 61-4		mu 32-5		haa 98-4	
la 17-1		wo 24-3		gi 101-2		taa 60-4		bi 79-5	
ca 11-2		te 48-3		xee 68-4		fo 18-3		goo 57-4	
na 5-2		xu 112-4		fi 102-4		fu 23-5		too 114-3	
me 3-4		qa 35-4		go 16-5		nu 84-3		hee 88-4	
da 9-1		co 30-3		wu 75-3		gaa 89-4		wee 103-4	
xi 7-2		fa 104-2		cee 90-3		maa 119-2		vu 141-2	
ce 15-2		cu 158-2		no 81-4		ra 80-5		li 8-1	
ci 41-2		fe 63-5		gu 2-4		gee 21-2		lo 69-1	
wa 37-3		ti 51-4		ho 49-4		ba 120-5		cii 123-4	
ya 28-4		ga 25-3		mi 44-3		tee 10-3		xoo 95-3	
ja 50-2		xaa 96-4		waa 76-3		gii 87-4		naa 163-4	
we 78-2		va 34-4		qe 19-5		faa 27-3		guu 91-5	
ta 45-3		ge 64-3		hu 70-4		vi 47-2		re 86-6	
le 6-1		pa 33-3		ka 72-4		be 55-4		hii 105-5	
se 38-2		to 43-2		si 39-3		pe 97-4		tuu 46-3	

* gee 21-2 means the radical with transliteration *gee*, number 21 and a stroke count of 2

pi 126-3		gai 77-4		gea 99-4		geu 155-3	
hoo 74-5		bu 110-5		baa 139-4		gie 166-5	
nee 93-7		xai 127-2		tea 124-4		fea 176-5	
bo 85-4		coo 125-3		tei 164-3		muu 137-2	
xuu 154-5		cuu 135-3		paa 106-3		hao 147-3	
xae 100-4		gao 136-4		xao 109-2		gei 71-5	
zi 174-2		tau 129-5		ri 82-3		lii 151-2	
tae 143-5		huu 26-3		fai 115-6		gia 159-4	
ye 131-5		zo 134-4		fao 132-4		qi 160-5	
po 12-3		gau 128-6		ro 149-5		gio 175-6	
tai 167-4		laa 65-1		fau 153-6		teu 113-4	
fii 94-4		fuu 146-6		moo 161-3		hau 148-3	
nii 118-5		fae 171-4		nae 130-4		loo 157-2	
lu 36-2		mii 162-3		pii 142-4		zaa 165-4	
gae 58-5		pu 172-4		teo 170-2		tia 168-3	
di 144-3		hae 108-5		xau 116-2		tie 169-2	
tao 59-2		nuu 117-4		xea 138-3		zu 111-2	
foo 152-6		lee 133-1		pee 140-4		cae 156-3	
noo 83-3		hai 150-6		geo 145-6		wii 173-3	

Classifications * horizontal * vertical * straight legs * curly tail * tail support * head * unclassified

Appendix C

Chinese Recursive Radical Codes

C.1 Original Order

1	一	26	𠂇	51	尸	76	贝	101	甘	126	西
2	丨	27	干	52	己	77	水	102	石	127	而
3	丿	28	工	53	弓	78	见	103	龙	128	页
4	丶	29	土	54	子	79	牛	104	业	129	至
5	𠂆	30	艹	55	屮	80	手	105	目	130	虍
6	十	31	寸	56	女	81	气	106	田	131	虫
7	厂	32	廾	57	飞	82	毛	107	罒	132	肉
8	匚	33	大	58	马	83	长	108	皿	133	缶
9	卜	34	兀	59	幺	84	片	109	生	134	舌
10	冂	35	弋	60	巛	85	斤	110	矢	135	竹
11	八	36	小	61	王	86	爪	111	禾	136	臼
12	人	37	口	62	无	87	父	112	白	137	自
13	勹	38	口	63	韦	88	月	113	瓜	138	血
14	儿	39	山	64	木	89	氏	114	鸟	139	舟
15	匕	40	巾	65	支	90	欠	115	疒	140	色
16	几	41	彳	66	犭	91	风	116	立	141	齐
17	亠	42	彡	67	歹	92	殳	117	穴	142	衣
18	冫	43	夕	68	车	93	文	118	疋	143	羊
19	冖	44	夂	69	牙	94	方	119	皮	144	米
20	凵	45	犭	70	戈	95	火	120	癶	145	聿
21	卩	46	广	71	比	96	斗	121	矛	146	艮
22	刀	47	门	72	瓦	97	户	122	耒	147	羽
23	力	48	宀	73	止	98	心	123	老	148	糸
24	又	49	辶	74	攵	99	母	124	耳	149	麦
25	厶	50	彐	75	日	100	示	125	臣	150	走

C.2 Ordered by Stroke Count

1 (1)	一	24 (2)	又	415 (2)	乜	49 (3)	辶	271 (3)	戈	399 (3)	扌
2 (1)	丨	25 (2)	厶	417 (2)	ㅓ	50 (3)	彐	276 (3)	川	403 (3)	乇
3 (1)	丿	26 (2)	廴	420 (2)	ㅌ	51 (3)	尸	279 (3)	乡	407 (3)	冂
4 (1)	丶	207 (2)	丂	27 (3)	干	52 (3)	己	283 (3)	也	410 (3)	片
5 (1)	乛	231 (2)	七	28 (3)	工	52 (3)	已	288 (3)	于	413 (3)	丬
229 (1)	𠃊	233 (2)	丁	29 (3)	土	52 (3)	巳	289 (3)	下	414 (3)	卜
354 (1)	乙	251 (2)	丁	30 (3)	艹	53 (3)	弓	292 (3)	尸	416 (3)	五
6 (2)	十	254 (2)	⺀	31 (3)	寸	54 (3)	子	302 (3)	彑	61 (4)	王
7 (2)	厂	268 (2)	乂	32 (3)	廾	55 (3)	屮	304 (3)	及	62 (4)	无
8 (2)	匚	282 (2)	了	33 (3)	大	56 (3)	女	318 (3)	叉	63 (4)	韦
9 (2)	卜	290 (2)	乙	34 (3)	兀	57 (3)	飞	321 (3)	丰	64 (4)	木
10 (2)	冂	296 (2)	ス	35 (3)	弋	58 (3)	马	324 (3)	与	65 (4)	支
11 (2)	八	299 (2)	マ	36 (3)	小	59 (3)	幺	331 (3)	才	67 (4)	歹
12 (2)	人	300 (2)	乃	37 (3)	口	60 (3)	巛	343 (3)	万	68 (4)	车
13 (2)	勹	309 (2)	刂	38 (3)	囗	66 (3)	犭	355 (3)	乚	69 (4)	牙
14 (2)	儿	313 (2)	九	39 (3)	山	148 (3)	纟	361 (3)	丘	70 (4)	戈
15 (2)	匕	345 (2)	ユ	40 (3)	巾	185 (3)	饣	364 (3)	卫	71 (4)	比
16 (2)	几	348 (2)	丩	41 (3)	彳	202 (3)	廿	369 (3)	乇	72 (4)	瓦
17 (2)	亠	356 (2)	ㄡ	42 (3)	彡	204 (3)	丸	371 (3)	乍	73 (4)	止
18 (2)	冫	359 (2)	コ	43 (3)	夕	234 (3)	云	381 (3)	彐	74 (4)	攵
19 (2)	冖	373 (2)	乡	44 (3)	夂	255 (3)	⺍	382 (3)	丈	75 (4)	日
20 (2)	凵	375 (2)	乚	45 (3)	丬	256 (3)	牛	385 (3)	勿	75 (4)	曰
21 (2)	卩	393 (2)	ヌ	46 (3)	广	258 (3)	上	387 (3)	歹	76 (4)	贝
22 (2)	刀	406 (2)	力	47 (3)	门	261 (3)	久	390 (3)	丹	77 (4)	水
23 (2)	力	408 (2)	《	48 (3)	宀	263 (3)	门	394 (3)	乡	78 (4)	见

* 22 (2) means radical number 22 with a stroke count of 2.

79 (4)	牛	317 (4)	𠂔	118 (5)	疋	379 (5)	丰	341 (6)	豕	168 (8)	青
80 (4)	手	319 (4)	内	119 (5)	皮	380 (5)	毌	350 (6)	且	169 (8)	卓
81 (4)	气	320 (4)	仒	120 (5)	癶	386 (5)	冊	352 (6)	亚	170 (8)	雨
82 (4)	毛	322 (4)	止	121 (5)	矛	388 (5)	匆	353 (6)	四	171 (8)	非
83 (4)	长	328 (4)	尸	205 (5)	弗	392 (5)	仌	365 (6)	州	172 (8)	齿
84 (4)	片	330 (4)	艮	208 (5)	电	397 (5)	仌	366 (6)	束	173 (8)	黾
85 (4)	斤	332 (4)	太	214 (5)	末	402 (5)	甩	374 (6)	曲	174 (8)	隹
86 (4)	爪	338 (4)	午	217 (5)	用	30 (6)	卄	396 (6)	兴	175 (8)	阜
87 (4)	父	339 (4)	书	221 (5)	民	99 (6)	母	404 (6)	凸	176 (8)	金
88 (4)	月	340 (4)	凵	222 (5)	甲	122 (6)	耒	405 (6)	凹	177 (8)	鱼
89 (4)	氏	346 (4)	卅	226 (5)	冉	123 (6)	老	419 (6)	※	178 (8)	录
90 (4)	欠	349 (4)	办	235 (5)	由	124 (6)	耳	149 (7)	麦	179 (8)	革
91 (4)	风	357 (4)	丯	236 (5)	阝	125 (6)	臣	150 (7)	走	210 (8)	直
92 (4)	殳	357 (4)	丯	239 (5)	疋	126 (6)	西	151 (7)	赤	213 (8)	垂
93 (4)	文	362 (4)	夂	240 (5)	世	127 (6)	而	152 (7)	豆	238 (8)	其
94 (4)	方	367 (4)	去	243 (5)	半	128 (6)	页	153 (7)	酉	316 (8)	豕
95 (4)	火	368 (4)	乌	252 (5)	且	129 (6)	至	154 (7)	辰	363 (8)	承
96 (4)	斗	377 (4)	冊	257 (5)	氏	130 (6)	虍	155 (7)	豕	180 (9)	面
97 (4)	户	383 (4)	丹	259 (5)	戋	131 (6)	虫	156 (7)	卤	181 (9)	韭
98 (4)	心	389 (4)	冊	262 (5)	术	132 (6)	肉	157 (7)	里	182 (9)	骨
216 (4)	天	391 (4)	匹	266 (5)	丙	133 (6)	缶	158 (7)	足	183 (9)	香
218 (4)	巴	395 (4)	卌	273 (5)	弗	134 (6)	舌	159 (7)	邑	184 (9)	鬼
220 (4)	屯	398 (4)	朩	278 (5)	禸	135 (6)	竹	160 (7)	身	185 (9)	食
230 (4)	冄	400 (4)	自	297 (5)	去	136 (6)	臼	161 (7)	釆	186 (9)	音
232 (4)	五	401 (4)	尤	301 (5)	包	137 (6)	自	162 (7)	谷	187 (9)	首
237 (4)	央	411 (4)	户	303 (5)	头	138 (6)	血	163 (7)	豸	215 (9)	柬
244 (4)	𠂤	412 (4)	卩	307 (5)	𠂢	139 (6)	舟	164 (7)	龟	219 (9)	禺
246 (4)	尹	100 (5)	示	308 (5)	兄	140 (6)	色	165 (7)	角	188 (10)	髟
247 (4)	勿	101 (5)	甘	310 (5)	平	141 (6)	齐	166 (7)	言	189 (10)	鬲
248 (4)	乌	102 (5)	石	310 (5)	平	142 (6)	衣	167 (7)	辛	190 (10)	鬥
250 (4)	反	103 (5)	龙	311 (5)	央	143 (6)	羊	206 (7)	甫	191 (10)	高
253 (4)	乡	104 (5)	业	323 (5)	本	144 (6)	米	212 (7)	肖	197 (10)	黍
260 (4)	丑	105 (5)	目	325 (5)	正	145 (6)	聿	223 (7)	我	249 (10)	离
264 (4)	川	106 (5)	田	327 (5)	东	146 (6)	艮	225 (7)	巫	192 (11)	黄
265 (4)	井	107 (5)	罒	329 (5)	申	147 (6)	羽	241 (7)	奂	193 (11)	麻
274 (4)	开	108 (5)	皿	334 (5)	戊	203 (6)	回	242 (7)	肖	194 (11)	鹿
275 (4)	业	109 (5)	生	336 (5)	必	209 (6)	亦	269 (7)	卵	195 (12)	鼎
277 (4)	中	110 (5)	矢	337 (5)	巨	224 (6)	出	272 (7)	朱	196 (12)	黑
280 (4)	尢	111 (5)	禾	344 (5)	专	291 (6)	年	284 (7)	来	198 (13)	鼓
281 (4)	不	112 (5)	白	358 (5)	斥	295 (6)	成	306 (7)	甶	199 (13)	鼠
285 (4)	为	113 (5)	瓜	360 (5)	阝	298 (6)	兆	326 (7)	两	200 (14)	鼻
286 (4)	以	114 (5)	鸟	370 (5)	东	305 (6)	农	342 (7)	求	201 (17)	龠
287 (4)	尤	115 (5)	疒	372 (5)	未	314 (6)	𠂤	351 (7)	束		
293 (4)	户	116 (5)	立	376 (5)	卌	315 (6)	畕	409 (7)	豸		
294 (4)	夕	117 (5)	穴	378 (5)	亚	333 (6)	再	418 (7)	𦥑		

Appendix D

Software Outline

One of the major issues in working with Xixia characters is the difficulty in searching through the 6000 characters to find a character of interest. Unlike the major living languages there are no elementary textbooks that can be used to quickly build up efficiency in the language, so a large amount of time needs to be spent with dictionaries. In order to improve efficiency in working with the character set, an Excel spreadsheet tool has been developed.

The *Lookup* sheet allows Xixia characters to be searched for by meaning. Once the English meaning is typed in, all Xixia characters with that meaning are returned. The dictionary meanings have been taken from Li FanWen's dictionary [41]. The partial/complete match toggle helps to refine a search. For example, searching for the word *cat* under a partial match will return 35 characters, having meanings such as cattail, catch and even duplicate - any words which contain the letters *cat*. Selecting the complete match option narrows the search and only two characters, both meaning *cat* are returned. There is also the possibility of returning the characters within a given category - these categories are the same as those used in the categorical dictionary in the dictionaries supplement.

Figure D.1: Dictionary Tool - Lookup by Meaning

Word to Find:	cat	Category	
Partial Match		animal	
Complete Match			
Number	**Character**	**Meaning**	**Grouping**
1767		cat	animal
4964		cat	animal

The *Rad4Corners* sheet can be used to search for characters by structure. There are five different methods of searching. The first method is by radical number, using the radicals shown in Li Fanwen's dictionary. These radicals can also be selected by clicking on a picture of a radical on the *RadLookup* sheet.

Figure D.2: Dictionary Tool - Lookup by Character Structure

		Input				
Radical No:			☑ Code Segment			
Four corners no:			☑ Full Match			
Character no:						
Recursive code:						
Transliteration:		mmc				
Mojikyo No.	**Character**	**Meaning**	**Four Corners**	**Rad No**	**Recursive**	**Transliteration**
4054		surname	452120	67	[44,{3,30}]	mimeco

The second method is the four corners method, also outlined in Li Fanwen's dictionary. Any number of corners from one to six can be used. If a particular corner is illegible, this can also be skipped over by using a question mark *?* to represent the relevant position.

The third method provides a facility to quickly lookup a character with a given Mojikyo/Li Fanwen dictionary number.

The fourth method allows the use of the recursive index developed in Chapter 2 on page 13. If the *Code Segment* toggle on the side is selected, then the recursive code can be entered, eg. $\{1, 3, [40, \{11, 1, [14, 17, 14]\}]\}$. If the *Code Segment* toggle isn't selected, the radical numbers together with (optionally) their multiplicities can be entered, eg. 1,2;14,2;17 - this searches for all characters with exactly two instances of radical number 1, 2 instances of radical number 14 and one or more instances of radical number 17 (multiplicity is unspecified here for radical number 17). Note that the radical numbers here are those presented in Appendix B, which are different to the ones presented in Li Fanwen's dictionary. With this method it is also possible to turn the *Full Match* toggle on so only characters which exactly match the input code are returned. When the *Full Match* toggle is turned off it is possible to enter only part of the recursive code and have all characters returned that contain that segment/list of radicals somewhere.

Finally, the fifth method allows the transliteration method to be used, developed in Chapter 2 on page 20. This uses the reduced transliteration code (without vowels), which is all that is necessary for searching. The *Full Match* toggle can also be used under this method to specify a partial or a full match.

The various categories are also cumulative. For example, it is possible to search for a character by typing in a radical number and then further restricting the search by typing in part or the whole of the four corners number.

Bibliography

[1] *Xinhua Character Dictionary* 新华字典. Commercial Press 商务印书出版社, Beijing, China, 10th edition, 2004.

[2] Bian Ren 邊人, editor. *Xixia: A Country Extinguished in History* 西夏*:* 消逝在歷史記憶中的國度. Peace Book Co. Ltd. 和平圖書有限公司, Hong Kong, China, 2006.

[3] Tom Bishop and Richard Cook. "A Specification for CDL - Character Description Language". Technical report, 2003. http://www.wenlin.com/cdl (accessed 30-Jul-2008).

[4] Tom Bishop and Richard Cook. "Character Description Language (CDL): The Set of Basic CJK Unified Stroke Types". Technical report, 2004. http://www.wenlin.com/cdl (accessed 30-Jul-2008).

[5] Chen Yongsheng 陈永胜. *Research on the Xixia Legal System* 西夏法律制度研究. Minority Nationalities Press 民族出版社, Beijing, China, 2006.

[6] Sir Gerard Clauson. "The Future of Tangut (Hsi Hsia) Studies". *Asia Major*, 11:54–77, 1964.

[7] Bernard Comrie, Stephen Matthews and Maria Polinsky, editors. *The Atlas of Languages: The origin and development of languages throughout the world.* ABC Books, Sydney, Australia, 2003.

[8] Richard Cook. "Proposal to Encode Tangut Characters in UCS Plane 1". Technical report, UC Berkeley Script Encoding Initiative, 2007.

[9] Tashi Daknewa. *A Basic Grammar of Modern Spoken Tibetan.* Library of Tibetan Works and Archives, Dharamsala, India, 1990.

[10] Ruth Dunnel[l]. "Chapter 2. The Hsi Hsia". In *The Cambridge History of China. Volume 6. Alien Regimes and Border States, 907-1368*, pages 154–214. Cambridge University Press, 1994.

[11] Ruth W. Dunnell. *The Great State of White and High. Buddhism and State Formation in Eleventh-Century Xia.* University of Hawaii Press, Honolulu, Hawaii, 1996.

[12] Ruth Wilton Dunnell. *Tanguts and the Tangut State of Ta Hsia.* PhD Thesis, Princeton, 1983.

[13] Gao Yaqi 高雅琪 and Zhuang Deming 庄德明. "The Design of Software for Lookup and Browsing of Xixia Characters 西夏字部件检索系统的设计". In Li Fanwen 李范文 [45], pages 527–537.

[14] Gong Hwang-Cherng. "Chapter 37. Tangut". In Graham Thurgood and Randy J. LaPolla [68], pages 602–620.

[15] Eric Grinstead. *Analysis of the Tangut Script.* Curzon Press, London, UK, 1972.

[16] Han Yinsheng 韓蔭晟, editor. *Collected Material on the Dangxiang and Xixia, Volumes 1-9* 党項與西夏資料匯編. Ningxia People's Press 宁夏人民出版社, 2000.

[17] Ray Huang. *China A Macro History*. M.E. Sharpe, New York, 1988.

[18] Guillaume Jacques. *Tangut Text I. Love and Piety. Textes tangoutes I. Le nouveau recueil sur l'amour parental et la piété filiale*. München: Lincom Europa, 2007.

[19] Juha Janhunen. "On the names of the Tangut". *Studia Etymologica Cracoviensia*, 12:63–73, 2007.

[20] Jing Yongshi 景永时 and Jia Changye 贾常业, editors. *System for Processing Xixia Characters* 西夏文字处理系统. Ningxia People's Press 蓼夏人民出版社, 2007.

[21] K.B. Kepping (К.Б. Кепинг). *Morphology of the Tangut Language* Тангутский Язык. Морфология. Scientific Publishing House Издательство "наука", Moscow, 1985.

[22] K.B. Kepping, V.S. Kolokolov, E.I. Kychanov and A.P. Terentev-Katansky (К.Б.Кепинг, В.С.Колоколов, Е.И.Кычанов И А.П.Терентьев-Катанский). *Sea of Characters - Facsimile of Tangut Woodcut* Море Письмен - факсимиле тангутских ксилографов. Eastern Literature Press Издательство восточной литературы, Moscow, USSR, 1969.

[23] Ksenia Kepping (Ксения Кепинг). *Sun Zi in Tangut Translation* Сунь цзы в тангутском переводе. Scientific Publishing House Издательство "наука", Moscow, 1979.

[24] Ksenia Kepping (Ксения Кепинг). *The Grove of Classifications: A Lost Chinese Encyclopedia in Tangut Translation* Лес категории — утраченная китайская лэйшу в тангутском переводе. Scientific Publishing House Издательство "наука", Moscow, Russia, 1983.

[25] Ksenia Kepping (Ксения Кепинг). *Ksenia Kepping Last Works and Documents* Ксения Кепинг Последние статьи и документы. Publishing House Omega Издательство "Омега", St-Petersburg, Russia, 2003.

[26] Ksenia Kepping (Ксения Кепинг). "Mi-Nia (Tangut) Self-appellation and Self-portraiture in Khara-Khoto Materials". In *Ksenia Kepping Last Works and Documents* Ксения Кепинг Последние статьи и документы [25], pages 97–115.

[27] Ksenia Kepping (Ксения Кепинг). "Tangut Ritual Language". In *Ksenia Kepping Last Works and Documents* Ксения Кепинг Последние статьи и документы [25], pages 24–28.

[28] Ksenia Kepping (Ксения Кепинг). "The Black-headed and the Red-faced in Tangut Indigenous Texts". In *Ksenia Kepping Last Works and Documents* Ксения Кепинг Последние статьи и документы [25], pages 116–136.

[29] P.K. Kozlov (П.К. Козлов). *Mongolia, Amdo and the Dead City of Khara-Khoto* Монголия и Амдо и мертвый город Хара-Хото. Tocudarctvenoye Publishers тосударственное издательство, Moscow, Russia, 2nd edition, 1948.

[30] Luc Kwanten. *The Timely Pearl. A 12th Century Tangut Chinese Glossary. Volume I, The Chinese Glosses*. Research Institute for Inner Asian Studies, Indiana University, Bloomington, USA, 1982.

[31] Luc Kwanten. "The Structure of the Tangut [Hsi Hsia] Characters". *T'oung Pao*, 75:1–42, 1989.

[32] Luc Kwanten. "Tangut Miscellanea: I. On the Inventor of the Tangut Script". *Journal of the American Oriental Society*, 97(3):333–335, Jul–Sep 1977.

[33] E.I. Kychanov (Е.И. Кычанов). *The History of the Tangut State* Очерк истории тангутского государства. Scientific Publishing House Издательство "наука", Moscow, 1968.

[34] E.I. Kychanov (Е.И. Кычанов). *New and Reformed Law Codes of the Prosperous Kingdom of Heaven (1149-1169), Translated from Tangut in four parts.* Измененный и заново утвержденный кодекс девиза царствования Небесное процветание *(1149-1169).* Издание текста, перевод с тангутского, в *4*–х книгах. Scientific Publishing House Издательство "наука", Moscow, 1987–1989.

[35] E.I. Kychanov (Е.И. Кычанов) and S. Arakawa (С. Аракава), editors. *Tangut Dictionary. Tangut-Russian-English-Chinese Dictionary.* Словарь Тангутского *(*Си Ся*)* Языка. Тангутско-русско-англо-китайский словарь. Faculty of Letters, Kyoto University, Kyoto, Japan, 2006.

[36] E.I. Kychanov 克恰諾夫, Li Fanwen 李範文 and Luo Maokun 羅矛昆. *Research on the Sea of Sacred Information* 聖立義海研究. Ningxia People's Press 寧夏人民出版社, 1995.

[37] Randy J. LaPolla. "Chapter 2. Overview of Sino-Tibetan Morphosyntax". In Graham Thurgood and Randy J. LaPolla [68], pages 22–42.

[38] Berthold Laufer. "The Si-hia language, a study in Indo-Chinese philology". *T'oung-pao*, 17(2):1–126, 1916.

[39] Li Fanwen 李范文. *Homophones Research* 同音研究. Ningxia People's Press 宁夏人民出版社, Yinchuan, China, 1986.

[40] Li Fanwen 李范文. *The North-Western Chinese Dialect in Song Times* 宋代西北方音. China Social Science Press 中国社会科学出版社, Beijing, China, 1994.

[41] Li Fanwen 李范文, editor. *Tangut-Chinese Dictionary* 夏漢字典. China Social Science Press 中国社会科学出版社, Beijing, China, 1997.

[42] Li Fanwen 李范文. *Xi Xia History* 西夏通史. Ningxia People's Press 宁夏人民出版社, Yinchuan, China, 2005.

[43] Li Fanwen 李范文. "Appendix: Index of Xixia Research 附录:西夏研究论著目录索引". In *Xixia Research - The Proceedings of the Second International Conference on Xixia Studies* 西夏研究 - 第二届西夏学国际学术研讨会论文集 [45], pages 633–767.

[44] Li Fanwen 李范文. *Xixia Research 2. Comparison of The Rhyme Tables of Five Sound Categories and The Precious Rhymes of the Sea of Characters* 西夏研究 *2.*《五音切韵》与《文海宝韵》比较研究. China Social Science Press 中国社会科学出版社, Beijing, China, 2006.

[45] Li Fanwen 李范文, editor. *Xixia Research 3. The Proceedings of the Second International Conference on Xixia Studies* 西夏研究 *3.* 第二届西夏学国际学术研讨会论文集. China Social Science Press 中国社会科学出版社, Beijing, China, 2006.

[46] Li Fanwen 李范文, editor. *Xixia Research 6. A Study of the Tangut Language* 西夏研究 *6.* 西夏语文学 *(by N.A. Nevsky (*Н.А. Невский*)).* China Social Science Press 中国社会科学出版社, Beijing, China, 2007.

[47] Lin Yingchin 林英津. *Research on the Xixia Translation of* The Art of War, *2 volumes*夏譯《孫子兵法》研究. Academia Sinica Department of Historical Linguistics 中央研究院歷史語言研究所, Taipei, Taiwan, 1994.

[48] Lin Yingchin 林英津. "Linguistic Observations on *Xixia Version of the Lotus Sutra* Edited by Nishida Tatsuo (2005) 從語言學的觀點初探西夏語譯《法華經》-兼評西田龍2005編譯《西夏文《妙法蓮華經》》". In Lillian M. Huang Henry Y. Chang and Dah an Ho 張永利, 黃美金, 何大安, editors, *Streams Converging into an Ocean: Festschrift in Honour of Professor Paul Jen-kuei Li on His 70th Birthday (Language and Linguistics Monograph Series Number W-5* 百川匯海：李壬癸先生七秩壽慶論文集*(*《語言暨語言學》專刊外編之五*)*, pages 687–724. Institute of Linguistics, Academia Sinica 中央研究院語言學研究所, Taipei, Taiwan, 2006.

[49] Lin Yingchin 林英津. *Tangut Version of "Chanting the Names of Manjusri"* 西夏語譯《真實名經》釋文研究. Institute of Linguistics, Academia Sinica 中央研究院語言學研究所, Taipei, Taiwan, 2006.

[50] P.H. Matthews, editor. *Oxford Concise Dictionary of Linguistics*. Oxford University Press, Oxford, UK, 2nd edition, 2007.

[51] Nie Hongyin 聶鴻音. "Tangutology During the Past Decades". *Monumenta Serica*, 41:329–347, 1993.

[52] Nie Hongyin 聶鴻音, Huang ZhenHua 黃振華 and Shi Jinbo 史金波. *Lei Lin Research* 類林研究. Ningxia People's Press 寧夏人民出版社, Ningxia, China, 1993.

[53] Tatsuo Nishida (translated by James A. Matisoff). "The Structure of the Hsi-Hsia (Tangut) Characters". *Monumenta Serindica*, 8:42 pages, 1979.

[54] Nishida Tatsuo 西田龍雄. *A Study of the Hsi-Hsia Language. Reconstruction of the Hsi-Hsia Language and Decipherment of the Hsi-Hsia Script.* 西夏語の研究. 西夏語の再構成と西夏文字の解讀, volume 1. Zauho Kankokai 座右宝刊行会, Tokyo, Japan, 1964.

[55] Nishida Tatsuo 西田龍雄. *A Study of the Hsi-Hsia Language. Reconstruction of the Hsi-Hsia Language and Decipherment of the Hsi-Hsia Script.* 西夏語の研究. 西夏語の再構成と西夏文字の解讀, volume 2. Zauho Kankokai 座右宝刊行会, Tokyo, Japan, 1966.

[56] Nishida Tatsuo, editor. *Xixia Version of the Lotus Sutra*. St. Petersburg Branch, Institute of Oriental Studies, Russian Academy of Sciences and Soka Gakkai, 2005.

[57] Niu Dasheng 牛达生. *Xixia Remnants* 西夏遗迹. Historical Relics Press 文物出版社, Beijing, China, 2007.

[58] Daniel G. Peebles. *SCML: A Structural Representation for Chinese Characters*. Senior Honours Thesis, Dartmouth College, 2007.

[59] Conrad Schirokauer. *A Brief History of Chinese Civilisation*. Thomson Learning, Inc., United States, 1991.

[60] Behula Shah. "The Checkered Career of Ailanthus altissima". *Arnoldia*, pages 21–27, Fall, 1997.

[61] Shi Jinbo 史金波. "Xixia Buddhism, Part 1 西夏的佛教（上）". http://www.plm.org.hk/qikan/fayin/dharma/2k0508/g2k0508f07.htm, Week 8 2005, p.34 (accessed 11-Nov-2008).

[62] Shi Jinbo 史金波. *Xixia Culture* 西夏文化. Jilin Educational Press 吉林教育出版社, 1986.

[63] Shi Jinbo 史金波, Bai Bin 白滨 and Huang Zhenhua 黄振华. *Research on the Sea of Characters* 文海研究. China Social Science Press 中国社会科学出版社, Beijing, China, 1983.

[64] Sofronov M.V. (Софронов М.В.). *Grammar of the Tangut Language, 2 volumes* Грамматика Тангутского Языка. Scientific Publishing House Издательство "наука", Moscow, 1968.

[65] Michael Spang. "Your Own Hall of Memories". *Scientific American Mind*, 16(2):60–65, 2005.

[66] Tang Rongyao 唐荣尧. *Kingdom Extinguished* 王朝湮灭. Guang Ming Daily Newspaper Press 光明日报出版社, Beijing, China, 2006.

[67] Graham Thurgood. "Chapter 1. A Subgrouping of the Sino-Tibetan Languages: The Interaction between Language Contact, Change and Inheritance". In Graham Thurgood and Randy J. LaPolla [68], pages 3–21.

[68] Graham Thurgood and Randy J. LaPolla, editor. *The Sino–Tibetan Languages*. Routledge, Abingdon, UK, 2007.

[69] Oreste Vaccari and Mrs Enko Elisa Vaccari. *A.B.C. Japanese-English Dictionary*. Vaccari's Language Institute, Tokyo, Japan, 3rd edition, 1962.

[70] George van Driem. "Tangut Verbal Agreement and the Patient Category in Tibeto-Burman". *Bulletin of the School of Oriental and African Studies*, 54(3):520–534.

[71] Andrew West. "Re: Comment on PRI 98: IVD Adobe-Japan1 (pt.2)". http://www.unicode.org/mail-arch/unicode-ml/y2007-m03/0130.html, 22-Mar-2007 (accessed 2-Mar-2008).

[72] John Wilkins. *An Essay Towards a Real Character and a Philosophical Language*. Gellibrand, London, United Kingdom, 1668.

[73] Wai Wong, Candy L.K. Yiu and Kelvin C.F. Ng. "Typesetting Rare Chinese Characters in LaTeX". *TUGBoat*, 24(1):1001–1007, 2003.

[74] Wu Tianchi 吴天墀. *Xixia History Manuscript* 西夏史稿. Guangxi Normal University Press 广西师范大学出版社, Guilin, Guangxi, China, 2006.

[75] Yang Jitang 杨积堂. *Xixia Culture in Legal Statutes: Research on the Xixia "New and Reformed Law Codes of the Prosperous Kingdom of Heaven"* 法典中的西夏文化 *:* 西夏《天盛改旧新定律令》研究. Legal Press 法律出版社, Beijing, China, 2003.

[76] Zhang Gongjin 张公瑾 Huang Jianming 黄建明 and Ling Fuxiang 岭福祥, editors. *Summary of Ancient Minority Scripts* 民族古文献概览. Minority Nationalities Press 民族出版社, Beijing, China, 1997.