Viewpoints on the Structural Description of Chinese Characters

Tomohiko Morioka

Abstract. This paper is about our viewpoints and methodology concerning the description of the structure of Chinese characters. First, we describe how components can be detected in characters. When a character is used as a component of a compound character and its shape appears without significant change, then the component can be easily identified. However, in many cases, it is not so easy to find the components that build a character out of purely visual features. One of the factors is simplification of the graphic form when characters are assembled out of components. Such a change of glyph form reduces the connection with pronunciation and meaning of the original character and increases the symbolic aspect of the character. It is particularly complicated when multiple components are combined, transformed and demotivated into a symbolic component. Here we discuss these issues with respect to the productivity of components and to the relationships between components and characters.

1. Introduction

Most Chinese characters (漢字) can be represented by a combination of components. For example, the character "林" (forest) has the same component "木" (tree) on the left and right, and the character "雲" (cloud) has component "云" (phonetic part) placed under the component "雨" (rain). The structure of Chinese characters, which consists of a combination of components, is not only an abstract expression of its shape, but is also related to its semantic and phonetic values. Therefore, to understand a Chinese character, it is important to find out what components are used, where they are placed in the characters, and how they are combined.

In this paper, such a description of the structure of Chinese character is called a *Hanzi structure description* (漢字構造記述; structure description of Chinese character) (Morioka, 2018b). Various formalisms have been used to describe Hanzi structure. Nowadays a quite widespread formalism is the one of Ideographic Description Sequences (IDS) defined in

Tomohiko Morioka 0000-0001-5315-3383
47 Kitashirakawa-Higashioguramachi, Sakyo, Kyoto, 606–8265, Japan
E-mail: tomo@zinbun.kyoto-u.ac.jp

Y. Haralambous (Ed.), *Grapholinguistics in the 21st Century 2020. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 5.
Fluxus Editions, Brest, 2021, pp. 683–712. https://doi.org/10.36824/2020-graf-mori
ISBN: 978-2-9570549-7-8, e-ISBN: 978-2-9570549-9-2

ISO/IEC 10646 (*Information technology—Universal Coded Character Set (UCS)* 2014). In this paper, we use the IDS formalism to describe Hanzi structure. For example,

- 林 = ⿰木木
- 雲 = ⿱雨云

in IDS.

In these two examples we can easily detect components and infer structure. However, in some cases, the situation is ambiguous, e.g.,

- 旗 = ⿰方⿱𠂉其 or ⿸㫃其?
- 嬴 = ⿱吂⿲月女卂 or ⿵𣎆女?

How do we detect components and the corresponding structure in such cases?

2. Etymological View

When a character is used as a component of a compound character and its shape is simply inserted into it without significant change, the component can be easily detected via its shape. For example, the character "林" (forest) appears to have two components "木" (tree) arranged side-by-side. Therefore, the structure of "林" can be written as "⿰木木".

Some structures have been preserved even though the graphic forms of Chinese characters have changed significantly over the centuries. For example, "ᚼᚼ" is an Oracle-bone character corresponding to the modern character "林". Its structure can be written as "⿰ᚼᚼ," which is similar to "⿰木木" ("ᚼ" is an Oracle-bone character corresponding to the modern character "木").

Similarly, the character "雲" (cloud) seems to consist of a component "雨" (rain) placed above a component "云" (the structure being "⿱雨云"). The Oracle-bone character "𠄞" corresponds to the modern character "雲" *as a character*, but corresponds to "云" as a component of Hanzi structure.[1] That is, it is considered that "雲" was formed by adding the semantic component "雨" to distinguish it from morphemes other than "雲" because "云" was used phonetically. Anyway, in this case as well, the components can be easily found just by looking at the characters.

However, in many cases, it is not so easy to find the components of a character just by examining the graphic form of the character. One of the causes of the problem is simplification of the graphic form of components when they are assembled into characters. For example, "隆" (eminent, exalt) looks like a combination of "阝" and "夅" on the left and right,

1. In Mainland China, "云" is a simplified Chinese character corresponding to "雲".

namely "    ". However, "  " (Small Seal form (小篆) of Shuowen corresponding to "隆") seems to contain the component "  " ("生" (be born, living, raw)) inside the component "" ("降" (fall)), so that the structure could be "". Compared to the Small Seal form, Hanzi structure of "隆" should be described as "". Through this structural analysis, we see that "夅" can be considered as being a simplified form of "降" or a component of it. As described above, there are two approaches to structural description: a structural description based on appearance and a structural description based on etymological explanation.

To adopt the latter position, the etymological knowledge of the Chinese character is required. However, characters with clear etymology are only a small part of the whole, and the etymological data on many characters are unclear or unknown.

3. Component Models

Historically, Hanzi structure descriptions (or the underlying analysis) were written for humans. For example, *Shuowen Jiezi* (説文解字; Shuowen), which is considered to be the oldest radical-based Chinese character dictionary, describes the kinds of components that comprise each compound character. The analysis of Hanzi structures in Shuowen is based on the so-called six-categories classification model (六書) and focusing on components motivated by pronunciation and meaning. In this model, each component is considered to be derived from a character, and each component is considered to (partially) inherit the phonetic and/or semantic value of the original character.

In the twentieth century, Tang Lan (唐蘭) proposed a new research approach and the three-categories classification model (三書). He also focused on graphemes that were not motivated by pronunciation or meaning and named them "symbol characters" (記号字; unmotivated characters), see also Slaměníková (2019). Qiu Xigui (裘錫圭) also made great contributions to the study of symbol characters. Tatsuro Asahara (浅原達郎) greatly contributed to Qiu Xigui's approach by avoiding the classification of components, and by proposing a symbolization (demotivation) model based on a more relational viewpoint (Asahara, 1996). According to this theory, instead of considering the classification of semantic, pronunciation and symbolic components, it is assumed that associative keys connect characters and components motivated by meaning and pronunciation (or act as symbolization filters to remove them).

4. Productivity of Components

Based on Tatsuro Asahara's view, when studying the functionality of a component, it is important to focus on the set of characters that have the same component, and to explore this component's features (or the actual state of componentization) as a common tendency among them. In other words, this can be viewed as a position that focuses on the *productivity of the components*. Tatsuro Asahara gave up explaining the problem of hierarchical componentization, however, by focusing on the productivity of components, it may be possible to identify a composite component acting as a functional unit by analyzing the corpus of the Hanzi structure described visually.

For example, "嬴" can be described as "⿻𣎆女" (etymological decomposition: "𣎆" is a phonetic component and "女" (woman) is a semantic component) or as: "⿱吂⿲月女卂" (visual decomposition). When searching for characters containing "𣎆" in the CHISE-IDS Database[2] (Morioka, 2015), only 50 characters were found in those included in UCS (even limited to components, the following are found: 「嬴」「贏」(「嬴」)「𥢐」「羸」「[illegible]」「[illegible]」「羸」「[illegible]」「[illegible]」「[illegible]」「臝」「鸁」「[illegible]」「驘」「蠃」「[illegible]」「[illegible]」「[illegible]」「[illegible]」). In contrast, only fifteen characters containing "⿲月女卂" were found, all of which contained "嬴" or some variant of it. Similarly, in the case of "族," many characters containing "㫃" were found (such as "斻," "施," "斾," "斿," "旂," "旃," "旄," "旗," "旄," "旆," "旅," "旊," "旆," "旈," "旋," "旍," "旐," "旋," "旑," "旌," "旍," "旎," "族," "旐," "旛," "旟," "旝," …), whereas only three characters were found that contain "矦".

More detailed results are shown in Tables 1 to 12.

To perform this investigation, we introduced 12 rewriting rules, transforming functional structures into apparent structures. Each table corresponds to one of these rewriting rules:

Rule-111 ⿸⿰ LRB ↔ ⿰ L ⿱ RB
Rule-112 ⿸⿲ 亻丨 RB ↔ ⿰⿰ 亻丨⿱ RB
Rule-121 ⿸⿱ ATR ↔ ⿱ A ⿸ TR if T is *tare*
Rule-122 ⿸⿱ ADR ↔ ⿱ A ⿰ DR if D is not *tare*
Rule-131 ⿸⿸ E 彳 R ↔ ⿸ E ⿰彳 R
Rule-132 ⿸⿸ EAB ↔ ⿸ E ⿱ AB if A = non 彳
Rule-210 ⿹⿰ LRB ↔ ⿰⿱ LBR
Rule-411 ⿴⿱ AEM ↔ ⿱ A ⿴ EM if E is *kamae*
Rule-414 ⿴⿱ ABM ↔ ⿱ A ⿱ MB if both A and B are not enclosure
Rule-511 ⿶⿱⿰ LRBA ↔ ⿱⿲ LARB
Rule-611 ⿵⿱ A ⿰ LRC ↔ ⿱ A ⿲ LCR
Rule-612 ⿵⿳ AM ⿰ LRC ↔ ⿳ AM ⿲ LCR

2. `https://gitlab.chise.org/CHISE/ids`

Each table consists of five columns. The first column (“char”) display the characters, the second column (“structure”) its structure, the third column (“component”) the components, and the fourth column (p*n*) indicates the number of character objects in the CHISE character ontology that contain the given *component*. Columns other than *char* are divided into upper and lower subrows. The upper subrow contains information about functional structure and the lower one contains information about apparent structure.

The accuracy value is calculated by the following formula:

$$\mathcal{A}_x = \frac{{N_{px}}^2}{N_p(N_{pf} + N_{pa})} \cdot 100.$$

Let S_{pf} be a set of character objects in the CHISE character ontology that have a functional component, and let $N_{pf} = n(S_{pf})$ (p*n* value of the upper subline).

Similarly, let S_{pa} be a set of character objects in the CHISE character ontology that have a apparent component, and let $N_{pa} = n(S_{pa})$ (p*n* value of the lower subline).

Let $S_p = S_{pf} \cup S_{pa}$, and let $N_p = n(S_p)$.

Let x of $\mathcal{A}_x$ and N_{px} be a variable to select f (functional) or a (apparent).

Note that the CHISE character ontology is based on the Multiple Granularity Hanzi Structure Model (Morioka, 2015; 2018a), so that each number f-p*n* and a-p*n* denotes the plural glyph granularity of Chinese characters such as abstract character, unified-glyph, abstract-glyph (字體), etc. In addition, the CHISE character ontology also includes character objects that cannot be unified by the existing CJKV Unified Ideographs of UCS. However, these tables show only representative glyphs and abstract characters of UCS: <character> indicates abstract characters and characters without angle brackets denote representative glyphs. Some abstract components unify multiple abstract characters, which are expressed as: <人/入/𠆢>.

Table 1. ⿰⿰ LRB ↔ ⿰ L ⿱ RB (111)

char	structure	component	p*n*	accuracy
旅	⿸〈疒〉多	〈疒〉	167	98.8
	⿰方⿱〈𠂉〉多	⿱〈𠂉〉多	1	0.1
旋	⿸〈疒〉疋	〈疒〉	167	98.8
	⿰方定	定	1	0.1

Same as above	旓, 旕, 旟, **旃**, **旗**, 旚, 旞, **旚**, 旌, 旎, 旝, 旜, 旑, 旆, 旍, 〈旙〉, 〈旟〉, 〈旛〉, 〈旚〉, 〈旛〉, 〈旛〉, 〈旜〉, 〈旛〉, 〈旛〉, 〈旜〉, 〈旛〉, 〈旝〉, 〈旛〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旋〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旂〉, 〈旆〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈族〉, 〈旝〉, 〈旝〉, 〈旚〉, 〈旜〉, 〈旞〉, 〈旝〉, 〈旝〉, 〈旓〉, 〈旝〉, 〈旝〉, 〈旑〉, 〈旕〉, 〈旅〉, 〈旌〉, 〈旎〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旟〉, 〈旞〉, 〈旝〉, 〈旜〉, **旛**, 〈旚〉, 旖, 〈旒〉, 旐, 旎, **旌**, **旆**, 〈旚〉, 〈旜〉, 〈旆〉			
旝	⿸〈㫃〉**西** ⿰**方**〈覀〉	〈㫃〉 〈覀〉	167 2	97.6 0.1
〈旝〉	⿸〈㫃〉**言** ⿰**方**⿱〈𠂉〉**言**	〈㫃〉 ⿱〈𠂉〉**言**	167 2	97.6 0.1
Same as above	旝, 旝, 旝, 旝, 旝, 旝, 旝, 旝, 旝, 旝, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旝〉, 〈旓〉, 〈旍〉, 〈旝〉, **旄**, 〈旝〉, 〈旝〉			
旃	⿸〈㫃〉**冉** ⿰**方**⿱〈𠂉〉**冉**	〈㫃〉 ⿱〈𠂉〉**冉**	167 3	96.5 0.1
〈旋〉	⿸〈㫃〉〈疋〉 ⿰**方**疋	〈㫃〉 疋	167 3	96.5 0.1
Same as above	旝, **旒**, 〈旝〉, 〈旝〉, 〈旒〉			
旋	⿸〈㫃〉**疋** ⿰**方**疋	〈㫃〉 疋	167 4	95.4 0.1
〈旗〉	⿸〈㫃〉〈其〉 ⿰**方**㫷	〈㫃〉 㫷	167 4	95.4 0.1
族	⿸〈㫃〉**矢** ⿰**方**矦	〈㫃〉 矦	167 4	95.4 0.1
Same as above	旃, 〈旃〉, 〈旃〉, 旂			
旃	⿸〈㫃〉**几** ⿰**方**㐫	〈㫃〉 㐫	167 7	92.1 0.2
〈旃〉	⿸〈㫃〉**几** ⿰**方**㐫	〈㫃〉 㐫	167 7	92.1 0.2
旨	⿸〈㫃〉**日** ⿰**方**〈旨〉	〈㫃〉 〈旨〉	167 10	89.0 0.4
〈旨〉	⿸〈㫃〉**日** ⿰**方**〈旨〉	〈㫃〉 〈旨〉	167 10	89.0 0.4

斿	⿸〈㫃〉**子** ⿰**方**𭕄	〈㫃〉 𭕄		167 11	88.0 0.4
〈旆〉	⿸〈㫃〉**巾** ⿰**方**𰀁	〈㫃〉 𰀁		167 12	87.0 0.5
旅	⿸〈㫃〉𠂔 ⿰**方**𠂢	〈㫃〉 𠂢		167 15	84.2 0.7
〈旅〉	⿸〈㫃〉𠂔 ⿰**方**𠂢	〈㫃〉 𠂢		167 15	84.2 0.7
〈旋〉	⿸〈㫃〉〈疋〉 ⿰**方**〈疋〉	〈㫃〉 〈疋〉		167 18	81.5 1.0
施	⿸〈㫃〉**也** ⿰**方**〈㐌〉	〈㫃〉 〈㐌〉		167 88	42.9 11.9
〈滕〉	⿸〈胅〉氺 ⿰〈⺝/⺼/月/**月**〉⿱〈龹/关〉氺	〈胅〉	⿱〈龹/关〉氺	25 1	92.5 0.2
〈勝〉	⿸〈胅〉**力** ⿰〈⺝/⺼/月/**月**〉⿱〈龹/关〉**力**	〈胅〉	⿱〈龹/关〉**力**	25 1	92.5 0.2
Same as above	〈腾〉, 〈黱〉, 〈鰧〉				
〈⿰月⿱龹衣〉	⿸〈胅〉〈衣〉 ⿰〈⺝/⺼/月/**月**〉〈⿱龹衣〉	〈胅〉 〈⿱龹衣〉		25 2	85.7 0.6
〈縢〉	⿸〈胅〉〈糹/糸/糸〉 ⿰〈⺝/⺼/月/**月**〉〈⿱龹糸〉	〈胅〉 〈⿱龹糸〉		25 7	61.0 4.8
〈幐〉	⿸〈胅〉**巾** ⿰〈⺝/⺼/月/**月**〉〈帣〉	〈胅〉 〈帣〉		25 7	61.0 4.8
〈幐〉	⿸〈胅〉**巾** ⿰〈⺝/⺼/月/**月**〉〈帣〉	〈胅〉 〈帣〉		25 7	61.0 4.8
䲢	⿸胅**鱼** ⿰〈月〉⿱关**鱼**	胅 ⿱关**鱼**		20 1	90.7 0.3
⿰月⿱关衣	⿸胅**衣** ⿰〈月〉⿱关**衣**	胅 ⿱关**衣**		20 1	90.7 0.3
Same as above	𫆏, 朘				
⿰月⿱关田	⿸胅**田** ⿰〈月〉〈⿱关田〉	胅 〈⿱关田〉		20 2	82.6 0.9

謄	⿸朕言 ⿰〈月〉⿱关言	朕 ⿱关言	20 4	69.4 2.8
〈⿸数虫〉	⿸〈数〉虫 ⿰男⿱〈夊/夂〉虫	〈数〉 ⿱〈夊/夂〉虫	18 1	89.8 0.3
〈⿸数貝〉	⿸〈数〉貝 ⿰男⿱〈夊/夂〉貝	〈数〉 ⿱〈夊/夂〉貝	18 2	81.0 1.1
〈⿸朕足〉	⿸〈朕〉足 ⿰〈月〉⿱〈关〉足	〈朕〉 ⿱〈关〉足	15 1	87.9 0.4
〈⿸倝魚〉	⿸〈倝〉魚 ⿰〈𠦝〉⿱〈人/入/𠆢〉魚	〈倝〉 ⿱〈人/入/𠆢〉魚	15 1	87.9 0.4
Same as above	〈⿸倝飛〉, 〈⿸倝夋〉, 〈⿸倝火〉, 〈⿸倝日〉, 〈⿸倝戈〉, 〈⿸倝隹〉, 〈⿸倝虫〉, 〈翰〉, 〈⿸倝木〉, 〈斡〉, 〈⿸倝鳥〉, 〈⿸倝毛〉			
〈幹〉	⿸〈倝〉干 ⿰〈𠦝〉〈仐〉	〈倝〉 〈仐〉	15 10	36.0 16.1
⿸倝火	⿸倝火 ⿰〈𠦝〉⿱人火	倝 ⿱人火	15 1	87.9 0.4
⿸倝飛	⿸倝飛 ⿰〈𠦝〉⿱人飛	倝 ⿱人飛	15 1	87.9 0.4
Same as above	⿸倝魚, ⿸倝鳥, ⿸倝毛, ⿸倝虫, ⿸倝夋, **翰**, **幹**, **斡**			
⿸倝戈	⿸倝戈 ⿰〈𠦝〉⿱人戈	倝 ⿱人戈	15 2	77.9 1.4
⿸倝日	⿸倝日 ⿰〈𠦝〉⿱人日	倝 ⿱人日	15 2	77.9 1.4
⿸倝木	⿸倝木 ⿰〈𠦝〉⿱人木	倝 ⿱人木	15 2	77.9 1.4
黱	⿸朕黑 ⿰月⿱龹黑	朕 ⿱龹黑	14 1	87.1 0.5
鰧	⿸朕魚 ⿰月⿱龹魚	朕 ⿱龹魚	14 1	87.1 0.5
幐	⿸朕巾 ⿰月帣	朕 帣	14 1	87.1 0.5
榺	⿸朕木 ⿰月桊	朕 桊	14 2	76.6 1.6

賸	⿸[月龹]**貝** ⿰月⿱龹**貝**	[月龹] ⿱龹**貝**	14 5	54.3 7.0
媵	⿸[月龹]**女** ⿰月⿱龹**女**	[月龹] ⿱龹**女**	14 5	54.3 7.0
塍	⿸[月龹]**土** ⿰月⿱龹**土**	[月龹] ⿱龹**土**	14 5	54.3 7.0
〈隆〉	⿸〈降〉**生** ⿰〈阝 〉⿱〈夅〉**生**	〈降〉 ⿱〈夅〉**生**	13 1	86.2 0.6
⿸[爿厂]臣	⿸〈[爿厂]〉**臣** ⿰**爿**⿱**一臣**	〈[爿厂]〉 ⿱**一臣**	11 1	84.0 0.7
〈⿸[爿厂]萬〉	⿸〈[爿厂]〉〈萬〉 ⿰**爿**⿱**一**〈萬〉	〈[爿厂]〉 ⿱**一**〈萬〉	11 1	84.0 0.7
Same as above	〈⿸[爿厂]鮮〉, 〈⿸[爿厂]畱〉, 〈⿸[爿厂]非〉, 〈⿸[爿厂]叵〉, 〈⿸[爿厂]隹〉			
⿸[爿厂]矢	⿸〈[爿厂]〉**矢** ⿰**爿**⿱**一矢**	〈[爿厂]〉 ⿱**一矢**	11 2	71.6 2.4
〈⿸[爿厂]矢〉	⿸〈[爿厂]〉**矢** ⿰**爿**⿱**一矢**	〈[爿厂]〉 ⿱**一矢**	11 2	71.6 2.4
Same as above	⿸[爿厂]脊, 〈⿸[爿厂]脊〉			
糳	⿸[丵臼殳]**米** ⿰[丵臼]⿱**殳米**	[丵臼殳] ⿱**殳米**	10 1	82.6 0.9
⿸[甹攵]貝	⿸[甹攵]**貝** ⿰甹⿱**攵貝**	[甹攵] ⿱**攵貝**	19 2	81.9 0.9
⿸[甹攵]虫	⿸[甹攵]**虫** ⿰甹[夂虫]	[甹攵] [夂虫]	19 4	68.2 3.1
絛	⿸〈攸〉**糸** ⿰亻丨⿱**支糸**	〈攸〉 ⿱**支糸**	9 1	81.0 1.1
⿸[攸]足	⿸〈攸〉足 ⿰亻丨⿱**支**足	〈攸〉 ⿱**支**足	9 1	81.0 1.1
〈脩〉	⿸〈攸〉月 ⿰亻丨⿱**支**月	〈攸〉 ⿱**支**月	9 2	66.9 3.4
〈佞〉	⿸〈仁〉**女** ⿰〈亻 〉〈[二女]〉	〈仁〉 〈[二女]〉	16 2	79.0 1.3

佞	⿸**仁女** ⿰⟨亻 ⟩⟨妄⟩	**仁** ⟨妄⟩	16 2	79.0 1.3
⟨臨⟩	⿸⟨丨丨𠂉⟩⟨吅/罒⟩ ⿰丨丨⿱𠂉⟨吅/罒⟩	⟨丨丨𠂉⟩ ⿱𠂉⟨吅/罒⟩	7 1	76.6 1.6
⟨嶐⟩	⿸⟨陸⟩**山** ⿰⟨阝 ⟩⿱⟨夆⟩**山**	⟨陸⟩ ⿱⟨夆⟩**山**	7 1	76.6 1.6

Table 2. ⿸⿰亻丨 RB ↔ ⿰⿰亻丨⿱ RB (112)

char	structure	component	p*n*	accuracy
⿸攸足	⿸⟨攸⟩**足** ⿰亻丨⿱⟨**夂/夊**⟩**足**	⟨攸⟩ ⿱⟨**夂/夊**⟩**足**	54 1	96.4 0.1
⟨⿸攸里⟩	⿸⟨攸⟩⟨里⟩ ⿰亻丨⿱⟨**夂/夊**⟩⟨里⟩	⟨攸⟩ ⿱⟨**夂/夊**⟩⟨里⟩	54 1	96.4 0.1
Same as above	⟨⿸攸足⟩, ⟨⿸攸目⟩, ⟨儵⟩, ⿸攸車, **脩**, ⟨翛⟩, ⟨倐⟩, ⟨倏⟩, ⟨⿸攸白⟩			
⟨⿸攸貝⟩	⿸⟨攸⟩**貝** ⿰亻丨⿱⟨**夂/夊**⟩**貝**	⟨攸⟩ ⿱⟨**夂/夊**⟩**貝**	54 2	93.0 0.2
⟨絛⟩	⿸⟨攸⟩⟨糹 /糸/?糸⟩ ⿰亻丨⿱夂糸	⟨攸⟩ ⿱夂糸	54 2	93.0 0.2
修	⿸⟨攸⟩**彡** ⿰亻丨⟨⿱夂彡⟩	⟨攸⟩ ⟨⿱夂彡⟩	54 6	81.0 1.1
⟨修⟩	⿸⟨攸⟩**彡** ⿰亻丨⟨⿱夂彡⟩	⟨攸⟩ ⟨⿱夂彡⟩	54 6	81.0 1.1
⟨⿸攸田⟩	⿸⟨攸⟩**田** ⿰亻丨⟨备⟩	⟨攸⟩ ⟨备⟩	54 16	59.5 5.3
⟨條⟩	⿸⟨攸⟩⟨朩/**木**⟩ ⿰亻丨⟨条⟩	⟨攸⟩ ⟨条⟩	54 24	47.9 9.5
⿸攸足	⿸**攸足** ⿰亻丨⿱**夂足**	**攸** ⿱**夂足**	36 1	94.7 0.1
絛	⿸**攸糸** ⿰亻丨⿱夂糸	**攸** ⿱夂糸	36 1	94.7 0.1
Same as above	**翛**, **儵**, **條**, ⿸攸白, ⿸攸目, ⿸攸里			

⿸攸貝	⿸**攸貝** ⿰亻⿱**夊貝**	**攸** ⿱**夊貝**	36 2	89.8 0.3
倐	⿸攸**火** ⿰亻⿱**夊火**	攸 ⿱**夊火**	9 9	25.0 25.1
偹	⿸攸**田** ⿰亻备	攸 备	9 13	16.7 35.0

Table 3. ⿸⿱ ATR ↔ ⿱ A ⿸ TR if T is *tare* (121)

char	structure	component	p*n*	accuracy
⿸虍肎	⿸虍肎 ⿱⟨卜⟩⿸厂肎	虍 ⿸厂肎	65 1	97.0 0.1
⿸虍子	⿸虍**子** ⿱⟨卜⟩⿸厂**子**	虍 ⿸厂**子**	65 1	97.0 0.1
Same as above	⟨⿸虍肎⟩, ⟨⿸虍隶⟩, ⟨⿸虍臬⟩, ⟨⿸虍督⟩			
⿸虍又	⿸虍**又** ⿱⟨卜⟩⿸厂**又**	虍 ⿸厂**又**	65 2	94.1 0.1
⟨⿸虍又⟩	⿸虍**又** ⿱⟨卜⟩⿸厂**又**	虍 ⿸厂**又**	65 2	94.1 0.1
斄	⿸⟨𠩺⟩**來** ⿱枚⿸厂來	⟨𠩺⟩ ⿸厂來	26 1	92.7 0.2
⟨⿸𠩺貝⟩	⿸⟨𠩺⟩**貝** ⿱枚⟨⿸厂貝⟩	⟨𠩺⟩ ⟨⿸厂貝⟩	26 1	92.7 0.2
Same as above	⟨⿸𠩺来⟩, ⟨⿸𠩺金⟩, ⟨⿸𠩺黑⟩, ⟨⿸𠩺木⟩, ⟨⿸𠩺发⟩			
⟨⿸𠩺鳥⟩	⿸⟨𠩺⟩**鳥** ⿱枚**⿸厂鳥**	⟨𠩺⟩ **⿸厂鳥**	26 2	86.2 0.6
⟨斄⟩	⿸⟨𠩺⟩⟨來⟩ ⿱枚⟨⿸厂來⟩	⟨𠩺⟩ ⟨⿸厂來⟩	26 2	86.2 0.6
Same as above	⟨⿸𠩺友⟩, ⟨⿸𠩺心⟩, ⟨⿸𠩺子⟩			
⟨⿸𠩺巾⟩	⿸⟨𠩺⟩**巾** ⿱枚⟨⿸厂巾⟩	⟨𠩺⟩ ⟨⿸厂巾⟩	26 3	80.4 1.1

⟨⿸𠩺文⟩	⿸⟨𠩺⟩**文** ⿱枚⟨⿸厂文⟩	⟨𠩺⟩ ⟨⿸厂文⟩	26 4	75.1 1.8
漦	⿸⟨𠩺⟩**水** ⿱枚⟨⿸厂水⟩	⟨𠩺⟩ ⟨⿸厂水⟩	26 4	75.1 1.8
Same as above	⟨⿸𠩺木⟩, ⟨⿸𠩺文⟩, ⟨⿸𠩺万⟩			
⟨⿸𠩺干⟩	⿸⟨𠩺⟩**干** ⿱枚厈	⟨𠩺⟩ 厈	26 5	70.3 2.6
⟨嫠⟩	⿸⟨𠩺⟩**女** ⿱枚⿸**厂女**	⟨𠩺⟩ ⿸**厂女**	26 5	70.3 2.6
釐	⿸⟨𠩺⟩**里** ⿱枚**厘**	⟨𠩺⟩ **厘**	26 6	66.0 3.6
⟨犛⟩	⿸⟨𠩺⟩**牛** ⿱枚⿸**厂牛**	⟨𠩺⟩ ⿸**厂牛**	26 6	66.0 3.6
⟨氂⟩	⿸⟨𠩺⟩**毛** ⿱枚⿸**厂毛**	⟨𠩺⟩ ⿸**厂毛**	26 7	62.1 4.5
⟨⿸𠩺万⟩	⿸⟨𠩺⟩**万** ⿱枚厉	⟨𠩺⟩ 厉	26 13	44.4 11.2
⟨釐⟩	⿸⟨𠩺⟩⟨里⟩ ⿱枚⟨厘⟩	⟨𠩺⟩ ⟨厘⟩	26 44	13.8 39.6
⟨⿸产初⟩	⿸⟨产⟩**初** ⿱⟨亠/**文**⟩⿸**厂初**	⟨产⟩ ⿸**厂初**	22 1	91.5 0.2
⟨廉⟩	⿸⟨产⟩⟨兼⟩ ⿱⟨亠/**文**⟩⟨㢘⟩	⟨产⟩ ⟨㢘⟩	22 1	91.5 0.2
Same as above	⟨⿸产火⟩, ⟨⿸产子⟩, ⟨⿸产焦⟩			
⟨⿸产隶⟩	⿸⟨产⟩隶 ⿱⟨亠/**文**⟩⿸**厂**隶	⟨产⟩ ⿸**厂**隶	22 2	84.0 0.7
⟨⿸产言⟩	⿸⟨产⟩**言** ⿱⟨亠/**文**⟩⿸**厂言**	⟨产⟩ ⿸**厂言**	22 6	61.7 4.6
⟨⿸产火⟩	⿸⟨产⟩**火** ⿱⟨亠/**文**⟩**灰**	⟨产⟩ **灰**	22 24	22.9 27.3
⟨⿸产舌⟩	⿸⟨产⟩舌 ⿱⟨**刀**/⺈/力/ク/⺈⟩⿸**厂**舌	⟨产⟩ ⿸**厂**舌	21 1	91.1 0.2

⟨詹⟩	⿸⟨⺈厂⟩言 ⿱⟨**刀**/𠂉/⺆/ク/⺈⟩⿸**厂**言	⟨⺈厂⟩ ⿸**厂**言	21 1	91.1 0.2
Same as above	⟨㑨⟩, ⟨危⟩			
⟨㑅⟩	⿸⟨⺈厂⟩**失** ⿱⟨**刀**/𠂉/⺆/ク/⺈⟩⿸**厂失**	⟨⺈厂⟩ ⿸**厂失**	21 2	83.4 0.8
⟨矦⟩	⿸⟨⺈厂⟩**矢** ⿱⟨**刀**/𠂉/⺆/ク/⺈⟩厌	⟨⺈厂⟩ 厌	21 7	56.3 6.3
危	⿸⟨⺈厂⟩⟨㔾⟩ ⿱⟨**刀**/𠂉/⺆/ク/⺈⟩**厄**	⟨⺈厂⟩ **厄**	21 42	11.1 44.5
詹	⿸⺈厂舌 ⿱⟨⺈⟩⿸**厂**舌	⺈厂 ⿸**厂**舌	21 1	91.1 0.2
詹	⿸⺈厂言 ⿱⟨⺈⟩⿸**厂**言	⺈厂 ⿸**厂**言	21 1	91.1 0.2
㑅	⿸⺈厂**失** ⿱⟨⺈⟩⿸**厂失**	⺈厂 ⿸**厂失**	21 2	83.4 0.8
矦	⿸⺈厂**矢** ⿱⟨⺈⟩厌	⺈厂 厌	21 7	56.3 6.3
⟨巖⟩	⿸⟨岸⟩⟨磊⟩ ⿱**山**⿸⟨厂/丆⟩⟨磊⟩	⟨岸⟩ ⿸⟨厂/丆⟩⟨磊⟩	14 1	87.1 0.5
⟨嶢⟩	⿸⟨岸⟩⟨堯⟩ ⿱**山**⿸⟨厂/丆⟩⟨堯⟩	⟨岸⟩ ⿸⟨厂/丆⟩⟨堯⟩	14 1	87.1 0.5
Same as above	⟨㠣⟩, ⟨崵⟩, ⟨巕⟩, ⟨嵺⟩, ⟨㟪⟩, ⟨崀⟩, ⟨嵪⟩, ⟨崖⟩, ⟨岂⟩, ⟨㟜⟩, 崖			
㢋	⿸产⟨久⟩ ⿱亠⿸**厂**⟨久⟩	产 ⿸**厂**⟨久⟩	11 1	84.0 0.7
廉	⿸产兼 ⿱亠⿸**厂**兼	产 ⿸**厂**兼	11 1	84.0 0.7
廉	⿸产隶 ⿱亠⿸**厂**隶	产 ⿸**厂**隶	11 2	71.6 2.4
彦	⿸产彡 ⿱亠彥	产 彥	11 5	47.3 9.8

産	⿸产言	产	11	41.9
	⿱亠⿸厂言	⿸厂言	6	12.5

TABLE 4. ⿸⿱ ADR ↔ ⿱ A ⿰ DR if D is not *tare* (122)

char	structure	component	p*n*	accuracy
死	⿸⟨歹⟩巳	⟨歹⟩	451	99.6
	⿱一⿰⟨夕/夂/夊⟩巳	⿰⟨夕/夂/夊⟩巳	1	0.1
⟨死⟩	⿸⟨歹⟩⟨已/己/巳/㔾⟩	⟨歹⟩	451	99.6
	⿱一⿰⟨夕/夂/夊⟩⟨已/己/巳/㔾⟩		1	0.1
		⿰⟨夕/夂/夊⟩⟨已/己/巳/㔾⟩		
⟨死⟩	⿸⟨歹⟩⟨⟨匕⟩/七⟩	⟨歹⟩	451	99.6
	⿱一⿰⟨夕/夂/夊⟩⟨⟨匕⟩/七⟩		1	0.1
		⿰⟨夕/夂/夊⟩⟨⟨匕⟩/七⟩		
死	⿸歹匕	歹	113	93.3
	⿱一⿰夕匕	⿰夕匕	4	0.2
皍	⿸皀卩	皀	33	94.2
	⿱白卬	卬	1	0.1
⟨皍⟩	⿸⟨皀⟩卩	⟨皀⟩	23	91.8
	⿱白⿰⟨匕⟩卩	⿰⟨匕⟩卩	1	0.2
石	⿸⟨丆⟩口	⟨丆⟩	22	91.5
	⿱一⿰丿口	⿰丿口	1	0.2
⟨布⟩	⿸⟨丆⟩巾	⟨丆⟩	22	91.5
	⿱一⿰丿巾	⿰丿巾	1	0.2
⿸宀爿蔓	⿸宀爿蔓	宀爿	19	90.3
	⿱宀⿰爿蔓	⿰爿蔓	1	0.3
⿸宀爿采	⿸宀爿采	宀爿	19	90.3
	⿱宀⟨⿰爿采⟩	⟨⿰爿采⟩	1	0.3
Same as above	寢, 𡪫, 寤, 寀, 寥, 𡬑, 寤, ⟨寢⟩			
寢	⿸宀爿𠬶	宀爿	19	81.9
	⿱宀⟨⿰爿𠬶⟩	⟨⿰爿𠬶⟩	2	0.9
寱	⿸宀爿臬	宀爿	19	74.6
	⿱宀⿰爿臬	⿰爿臬	3	1.9
寤	⿸宀爿吾	宀爿	19	74.6
	⿱宀⟨⿰爿吾⟩	⟨⿰爿吾⟩	3	1.9

寎	⿴〈宵〉丙 ⿱宀⿰〈爿/丬〉丙	〈宵〉 ⿰〈爿/丬〉丙	17 1	89.2 0.3
〈寢〉	⿴〈宵〉叟 ⿱宀⿰〈爿/丬〉叟	〈宵〉 ⿰〈爿/丬〉叟	17 1	89.2 0.3
Same as above	〈癵〉, 〈癮〉, 〈寐〉, 〈寤〉, 〈寢〉, 〈寤〉, 〈寐〉, 〈寎〉, 〈寥〉, 〈寐〉			
寣	⿴〈宵〉言 ⿱宀⿰〈爿/丬〉言	〈宵〉 ⿰〈爿/丬〉言	17 2	80.1 1.1
〈寣〉	⿴〈宵〉言 ⿱宀⿰〈爿/丬〉言	〈宵〉 ⿰〈爿/丬〉言	17 2	80.1 1.1
着	⿴羊目 ⿱〈⺶〉⿰丿 目	羊 ⿰丿 目	12 1	85.2 0.6
羞	⿴羊丑 ⿱〈⺶〉⿰丿 丑	羊 ⿰丿 丑	12 1	85.2 0.6
Same as above	〈羹〉, 〈羹〉, 〈羹〉			

Table 5. ⿰⿰ E 彳 R ↔ ⿰ E ⿰彳 R (131)

char	structure	component	p*n*	accuracy
屩	⿰尸曳 ⿰尸⿰彳曳	尸 ⿰彳曳	9 1	81.0 1.1
屨	⿰尸婁 ⿰尸⿰彳婁	尸 ⿰彳婁	9 1	81.0 1.1
屩	⿰尸喬 ⿰尸僑	尸 僑	9 1	81.0 1.1
〈屩〉	⿰尸〈曳/曳〉 ⿰尸⿰彳〈曳/曳〉	尸 ⿰彳〈曳/曳〉	9 1	81.0 1.1
〈屨〉	⿰尸〈婁/婁/婁〉 ⿰尸⿰彳〈婁/婁/婁〉	尸 ⿰彳〈婁/婁/婁〉	9 1	81.0 1.1
屧	⿰尸枼 ⿰尸⿰彳枼	尸 ⿰彳枼	9 1	81.0 1.1
〈屧〉	⿰尸娄 ⿰尸⿰彳娄	尸 ⿰彳娄	9 1	81.0 1.1

屜	⿸𡰯**世** ⿸**尸**⿰**彳世**	𡰯 ⿰**彳世**	9 1	81.0 1.1
〈屩〉	⿸𡰯〈喬〉 ⿸**尸**〈⿰彳喬〉	𡰯 〈⿰彳喬〉	9 2	66.9 3.4

TABLE 6. ⿸⿸ EAB ↔ ⿸ E ⿱ AB if A = non 彳 (132)

char	structure	component	p*n*	accuracy
屘	⿸**尾水** ⿸**尸**⿱**毛水**	**尾** ⿱**毛水**	145 1	98.6 0.1
〈⿸麻亽〉	⿸〈麻〉亽 ⿸**广**⿱〈**林**/𣏟〉亽	〈麻〉 ⿱〈**林**/𣏟〉亽	133 1	98.5 0.1
〈⿸麻木〉	⿸〈麻〉**木** ⿸**广**⿱〈**林**/𣏟〉**木**	〈麻〉 ⿱〈**林**/𣏟〉**木**	133 1	98.5 0.1
Same as above	〈⿸麻黽〉, 〈⿸麻賛〉, 〈⿸麻參〉, 〈⿸麻云〉, 〈⿸麻吉〉, 〈⿸麻亘〉, 〈⿸麻龍〉, 〈⿸麻禺〉, 〈⿸麻西〉, 〈⿸麻留〉, 〈⿸麻弁〉, 〈⿸麻几〉, 〈⿸麻而〉, 〈⿸麻目〉, 〈⿸麻皿〉, 〈⿸麻育〉, 〈⿸麻氏〉, 〈⿸麻隹〉, 〈⿸麻虫〉, 〈摩〉, 〈⿸麻口〉, 〈⿸麻男〉, 〈魔〉, 〈靡〉, 〈糜〉, 〈⿸麻木〉, 〈⿸麻手〉, 〈⿸麻鳥〉, 〈⿸麻骨〉			
〈⿸尼木〉	⿸〈尼〉**木** ⿸**尸**⿱〈**匕**〉**木**	〈尼〉 ⿱〈**匕**〉**木**	97 1	98.0 0.1
⿸麻多	⿸麻**多** ⿸**广**⿱〈𣏟〉**多**	麻 ⿱〈𣏟〉**多**	77 1	97.5 0.1
糜	⿸麻**米** ⿸**广**⿱〈𣏟〉**米**	麻 ⿱〈𣏟〉**米**	77 1	97.5 0.1
Same as above	黂, ⿸麻賈, ⿸麻爲, ⿸麻易, ⿸麻电, ⿸麻鬼, ⿸麻英, ⿸麻奉, ⿸麻旁, ⿸麻弓, ⿸麻豕, ⿸麻火, ⿸麻云, ⿸麻木, **⿸麻系**, **⿸麻电**, **麼**, ⿸麻隹, ⿸麻足, ⿸麻革, ⿸麻育, ⿸麻乙, ⿸麻鳥, 〈⿸麻食〉, 〈⿸麻育〉, 〈⿸麻肓〉			
⿸麻日	⿸麻**日** ⿸**广**𣈆	麻 𣈆	77 2	95.0 0.1
⿸麻女	⿸麻**女** ⿸**广**⿱〈𣏟〉**女**	麻 ⿱〈𣏟〉**女**	77 2	95.0 0.1
靡	⿸麻**非** ⿸**广**⿱〈𣏟〉**非**	麻 ⿱〈𣏟〉**非**	77 2	95.0 0.1
〈⿸厥鸟〉	⿸〈厥〉〈鸟〉 ⿸〈厂/丆〉⿱〈欮〉〈鸟〉	〈厥〉 ⿱〈欮〉〈鸟〉	67 1	97.1 0.1
〈⿸厥肉〉	⿸〈厥〉**肉** ⿸〈厂/丆〉⿱〈欮〉**肉**	〈厥〉 ⿱〈欮〉**肉**	67 1	97.1 0.1
〈⿸厥鳥〉	⿸〈厥〉**鳥** ⿸〈厂/丆〉⿱〈欮〉**鳥**	〈厥〉 ⿱〈欮〉**鳥**	67 1	97.1 0.1

〈蟨〉	⿸〈厥〉**虫** ⿸〈厂/丆〉⿱〈欮〉**虫**	〈厥〉 ⿱〈欮〉**虫**	67 2	94.3 0.1
〈⿸康心〉	⿸**康心** ⿸**广**⿱**隶心**	**康** ⿱**隶心**	55 1	96.5 0.1
〈⿸庚凡〉	⿸**庚**〈凡〉 ⿸**广**⿱肀〈凡〉	**庚** ⿱肀〈凡〉	51 1	96.2 0.1
〈⿸庚十〉	⿸**庚十** ⿸**广**⿱肀**十**	**庚** ⿱肀**十**	51 1	96.2 0.1
Same as above	〈⿸庚衣〉, 〈⿸庚寿〉, 〈赓〉, **賡**			
〈⿸原水〉	⿸〈原〉**水** ⿸〈厂/丆〉⿱泉**水**	〈原〉 ⿱泉**水**	48 1	96.0 0.1
⿸麻鳥	⿸**麻鳥** ⿸**广**〈⿱林鳥〉	**麻** 〈⿱林鳥〉	45 1	95.7 0.1
⿸麻口	⿸**麻口** ⿸**广**〈⿱林口〉	**麻** 〈⿱林口〉	45 1	95.7 0.1
Same as above	靡, ⿸麻共, **摩**, **魔**, **磨**, ⿸麻黽, ⿸麻巟, ⿸麻去, ⿸麻子			
⿸麻言	⿸**麻**言 ⿸**广**〈⿱林言〉	**麻** 〈⿱林言〉	45 2	91.7 0.2
⿸麻呂	⿸**麻呂** ⿸**广**⿱**林呂**	**麻** ⿱**林呂**	45 3	87.9 0.4
⿸声匃	⿸声匃 ⿸**广**⿱廿匃	声 ⿱廿匃	43 1	95.5 0.1
⿸声衣	⿸声**衣** ⿸**广**⿱廿**衣**	声 ⿱廿**衣**	43 1	95.5 0.1
Same as above	〈⿸声貝〉, 〈⿸声殳〉, 〈⿸声鬲〉, 〈⿸声鬼〉, 〈⿸声鳥〉, 〈⿸声用〉, ⿸声鳥			
⿸声且	⿸声**且** ⿸**广**⿱廿**且**	声 ⿱廿**且**	43 2	91.3 0.2
〈庶〉	⿸声灬 ⿸**广**⿱廿灬	声 ⿱廿灬	43 2	91.3 0.2
Same as above	〈⿸声衣〉, 〈⿸声且〉			
⿸声厸	⿸声厸 ⿸**广**⿱廿厸	声 ⿱廿厸	43 3	87.4 0.5

鹿	⿸𢉖**比** ⿸**广**⿱龷**比**	𢉖 ⿱龷**比**	43 3	87.4 0.5
⟨麁⟩	⿸𢉖厸 ⿸**广**⿱龷厸	𢉖 ⿱龷厸	43 3	87.4 0.5
⟨鹿⟩	⿸𢉖⟨比⟩ ⿸**广**⿱龷⟨比⟩	𢉖 ⿱龷⟨比⟩	43 5	80.3 1.1
⟨饜⟩	⿸⟨猒⟩**食** ⿸**厂**⟨餍⟩	⟨猒⟩ ⟨餍⟩	35 1	94.5 0.1
⟨靨⟩	⿸⟨猒⟩**面** ⿸**厂**⟨靥⟩	⟨猒⟩ ⟨靥⟩	35 1	94.5 0.1
Same as above	**魘**, ⟨魘⟩, 檿, ⟨曆⟩			
歷	⿸⟨厤⟩**止** ⿸⟨厂/丆⟩⿱秝**止**	⟨厤⟩ ⿱秝**止**	34 1	94.4 0.1
⟨曆⟩	⿸⟨厤⟩**田** ⿸⟨厂/丆⟩⟨畨⟩	⟨厤⟩ ⟨畨⟩	34 1	94.4 0.1
Same as above	⟨厯⟩, ⟨鷉⟩, ⟨鬲⟩, ⟨磿⟩, ⟨檿⟩, ⟨厭⟩, ⟨壢⟩, ⟨歴⟩, ⟨劦⟩, ⟨曆⟩			
⟨曆⟩	⿸⟨厤⟩**甘** ⿸⟨厂/丆⟩⟨香⟩	⟨厤⟩ ⟨香⟩	34 2	89.2 0.3
磿	⿸⟨厤⟩⟨石⟩ ⿸⟨厂/丆⟩⿱秝⟨石⟩	⟨厤⟩ ⿱秝⟨石⟩	34 2	89.2 0.3
⟨歴⟩	⿸⟨厤⟩⟨止⟩ ⿸⟨厂/丆⟩⿱秝⟨止⟩	⟨厤⟩ ⿱秝⟨止⟩	34 2	89.2 0.3
䧹	⿸⟨雁⟩**目** ⿸**广**⿱⟨倠⟩**目**	⟨雁⟩ ⿱⟨倠⟩**目**	31 1	93.8 0.1
雁	⿸⟨雁⟩**口** ⿸**广**⿱⟨倠⟩**口**	⟨雁⟩ ⿱⟨倠⟩**口**	31 1	93.8 0.1
Same as above	鷹, **膺**, 譍, 鷹, 赝, ⟨鷹⟩			
應	⿸⟨雁⟩**心** ⿸**广**⿱⟨倠⟩**心**	⟨雁⟩ ⿱⟨倠⟩**心**	31 2	88.2 0.4
鷹	⿸⟨雁⟩**鳥** ⿸**广**⿱⟨倠⟩**鳥**	⟨雁⟩ ⿱⟨倠⟩**鳥**	31 2	88.2 0.4
Same as above	譍, 贗			

腐	⿴**府衣** ⿴**广**⿱**付衣**	**府** ⿱**付衣**	29 1	93.4 0.2
腐	⿴**府肉** ⿴**广**⿱**付肉**	**府** ⿱**付肉**	29 1	93.4 0.2
〈賨〉	⿴**府貝** ⿴**广**⿱**付貝**	**府** ⿱**付貝**	29 2	87.5 0.5
庾	⿴**府天** ⿴**广**⿱**付天**	**府** ⿱**付天**	29 3	82.1 0.9
〈庾〉	⿴**府天** ⿴**广**⿱**付天**	**府** ⿱**付天**	29 3	82.1 0.9
靨	⿴**厭甲** ⿴**厂**⿱**猒甲**	**厭** ⿱**猒甲**	27 1	93.0 0.2
屔	⿴**灰匕** ⿴**厂**⿱**火匕**	**灰** ⿱**火匕**	24 1	92.2 0.2
〈屔〉	⿴**灰**〈匕〉 ⿴**厂**⿱**火**〈匕〉	**灰** ⿱**火**〈匕〉	24 1	92.2 0.2
〈愿〉	⿴**厚心** ⿴**厂**⿱〈旱〉**心**	**厚** ⿱〈旱〉**心**	21 1	91.1 0.2
卮	⿴〈厂〉〈巳〉 ⿴〈厂〉⿱一〈巳〉	〈厂〉 ⿱一〈巳〉	20 1	90.7 0.3
〈卮〉	⿴〈厂〉〈巳〉 ⿴〈厂〉⿱一〈巳〉	〈厂〉 ⿱一〈巳〉	20 1	90.7 0.3
〈卮〉	⿴〈厂〉**已/己/巳/**〈巳〉 ⿴〈厂〉⿱一〈**已/己/巳/**〈巳〉	〈厂〉 ⿱一〈**已/己/巳/**〈巳〉	20 1	90.7 0.3
〈斤〉	⿴〈厂〉丁 ⿴〈厂〉〈干〉	〈厂〉 〈干〉	20 2	82.6 0.9
〈乭〉	⿴〈厂〉**乙** ⿴〈厂〉〈乙〉	〈厂〉 〈乙〉	20 2	82.6 0.9
卮	⿴〈厂〉**巴** ⿴〈厂〉⿱一**巴**	〈厂〉 ⿱一**巴**	20 2	82.6 0.9
后	⿴〈厂〉**口** ⿴〈厂〉〈口〉	〈厂〉 〈口〉	20 68	5.2 59.8
〈帠〉	⿴〈厂〉**巾** ⿴〈厂〉巾	〈厂〉 巾	20 80	4.0 64.0
犀	⿴**尿牛** ⿴**尸**⿱**水牛**	**尿** ⿱**水牛**	16 1	88.6 0.4
〈廱〉	⿴**庫**〈蚰〉 ⿴**广**⿱**車**〈蚰〉	**庫** ⿱**車**〈蚰〉	16 1	88.6 0.4
麝	⿴麻鬲 ⿴广⿱秝鬲	麻 ⿱秝鬲	16 1	88.6 0.4

曆	⿸厤日 ⿸厂⿱秝日	厤 ⿱秝日	16 1	88.6 0.4
Same as above	曊, 曆, 曆, 曆, 曆, 曆, 曆, 曆, 曆			
燃	⿸厤火 ⿸厂⿱秝火	厤 ⿱秝火	16 2	79.0 1.3
歷	⿸厤心 ⿸厂⿱秝心	厤 ⿱秝心	16 2	79.0 1.3
𩪘	⿸厥骨 ⿸厂⿱欮骨	厥 ⿱欮骨	16 1	88.6 0.4
蟨	⿸厥虫 ⿸厂⿱欮虫	厥 ⿱欮虫	16 1	88.6 0.4
〈𫐐〉	⿸雁〈直〉 ⿸厂⿱〈倠〉〈直〉	雁 ⿱〈倠〉〈直〉	10 1	82.6 0.9
〈赝〉	⿸雁〈贝〉 ⿸厂⿱〈倠〉〈贝〉	雁 ⿱〈倠〉〈贝〉	10 1	82.6 0.9
贋	⿸雁貝 ⿸厂⿱〈倠〉貝	雁 ⿱〈倠〉貝	10 2	69.4 2.8
〈檿〉	⿸〈厌〉木 ⿸厂⿱犬木	〈厌〉 ⿱犬木	10 1	82.6 0.9
〈擪〉	⿸〈厌〉手 ⿸厂⿱犬手	〈厌〉 ⿱犬手	10 1	82.6 0.9
Same as above	〈懕〉, 〈㷳〉			
〈壓〉	⿸〈厌〉土 ⿸厂⿱犬土	〈厌〉 ⿱犬土	10 2	69.4 2.8
庶	⿸⿸广廿灬 ⿸广⿱廿灬	⿸广廿 ⿱廿灬	9 1	81.0 1.1
度	⿸⿸广廿又 ⿸广⿱廿又	⿸广廿 ⿱廿又	9 1	81.0 1.1
Same as above	廗, 庻			
庻	⿸⿸广廿火 ⿸广炗	⿸广廿 炗	9 3	56.3 6.3
席	⿸⿸广廿巾 ⿸广⿱廿巾	⿸广廿 ⿱廿巾	9 4	47.9 9.5
屬	⿸𡳿〈蜀〉 ⿸尸⿱丰〈蜀〉	𡳿 ⿱丰〈蜀〉	9 1	81.0 1.1

犀	⿸屖**牛** ⿸**尸**⿱丰**牛**	屖 ⿱丰**牛**	9 1	81.0 1.1
Same as above	犀, 〈屖〉, 〈犀〉, 〈屖〉, 〈犀〉			
愿	⿸**原心** ⿸**厂**⿱泉**心**	**原** ⿱泉**心**	8 1	79.0 1.3
〈⿸鴈真〉	⿸**鴈真** ⿸**厂**⿱〈⿰亻鳥〉**真**	**鴈** ⿱〈⿰亻鳥〉**真**	7 1	76.6 1.6
〈⿸鴈火〉	⿸**鴈火** ⿸**厂**⿱〈⿰亻鳥〉**火**	**鴈** ⿱〈⿰亻鳥〉**火**	7 1	76.6 1.6
〈庶〉	⿸⿸广廿灬 ⿸**广**⿱〈廾/廾/**廿**/卄/**龷**〉灬	⿸广廿 ⿱〈廾/廾/**廿**/卄/**龷**〉灬	7 1	76.6 1.6
〈度〉	⿸⿸广廿**又** ⿸**广**⿱〈廾/廾/**廿**/卄/**龷**〉**又**	⿸广廿 ⿱〈廾/廾/**廿**/卄/**龷**〉**又**	7 1	76.6 1.6
Same as above	〈庻〉, 〈廗〉			
〈席〉	⿸⿸广廿**巾** ⿸**广**〈帯/帯/巿/帯〉	⿸广廿 〈帯/帯/巿/帯〉	7 3	49.0 9.1
〈庻〉	⿸⿸广廿**火** ⿸**广**〈芡〉	⿸广廿 〈芡〉	7 5	34.0 17.4

TABLE 7. ⿹⿰ LRB ↔ ⿰⿱ LBR (210)

char	structure	component	p*n*	accuracy
〈⿹𣪊赤〉	⿹〈𣪊〉**赤** ⿰⿱𡉉**赤**〈殳〉	〈𣪊〉 ⿱𡉉**赤**	40 1	95.2 0.1
〈⿹𣪊羊〉	⿹〈𣪊〉**羊** ⿰⿱𡉉**羊**〈殳〉	〈𣪊〉 ⿱𡉉**羊**	40 1	95.2 0.1
Same as above	〈⿹𣪊口〉, 〈⿹𣪊皮〉, 〈⿹𣪊東〉, 〈⿹𣪊子〉, 鷇, **轂**, 縠, 〈⿹𣪊月〉, 〈⿹𣪊豆〉			
〈⿹𣪊豕〉	⿹〈𣪊〉〈豕〉 ⿰⿱𡉉〈豕〉〈殳〉	〈𣪊〉 ⿱𡉉〈豕〉	40 2	90.7 0.3
榖	⿹〈𣪊〉**木** ⿰𣓌〈殳〉	〈𣪊〉 𣓌	40 2	90.7 0.3
Same as above	**彀**, 〈⿹𣪊米〉, 〈⿹𣪊耳〉, 〈⿹𣪊豆〉, 〈⿹𣪊弓〉, 〈⿹𣪊豆〉, 〈⿹𣪊女〉			

〈㱿〉	⿹〈𣪊〉出 ⿰⿱士冖出〈殳〉	〈𣪊〉 ⿱士冖出	40 3	86.5 0.5
〈㝅〉	⿹〈𣪊〉子 ⿰孛〈殳〉	〈𣪊〉 孛	40 3	86.5 0.5
Same as above	〈㲉〉, 〈𤌊〉			
〈㲉〉	⿹〈𣪊〉卵 ⿰⿱士冖卵〈殳〉	〈𣪊〉 ⿱士冖卵	40 4	82.6 0.9
𣫃	⿹𣪊缶 ⿰⿱士冖缶殳	𣪊 ⿱士冖缶	32 1	94.0 0.1
嗀	⿹𣪊口 ⿰壳殳	𣪊 壳	32 1	94.0 0.1
Same as above	榖, 㝅, 𤌊, 嗀			
榖	⿹𣪊林 ⿰⿱士冖林殳	𣪊 ⿱士冖林	32 2	88.6 0.4
瞉	⿹𣪊目 ⿰⿱士冖目殳	𣪊 ⿱士冖目	32 2	88.6 0.4
Same as above	𣫃, 榖, 瞉, 穀			
㝅	⿹𣪊子 ⿰孛殳	𣪊 孛	32 3	83.6 0.8
𤌊	⿹𣪊火 ⿰⿱士冖火殳	𣪊 ⿱士冖火	32 3	83.6 0.8
㱿	⿹𣪊出 ⿰⿱士冖出殳	𣪊 ⿱士冖出	32 3	83.6 0.8
㲉	⿹𣪊卵 ⿰⿱士冖卵殳	𣪊 ⿱士冖卵	32 4	79.0 1.3
穀	⿹𣪊禾 ⿰〈稾〉殳	𣪊 〈稾〉	32 4	79.0 1.3
〈䫌〉	⿹〈須〉女 ⿰⿱彡女〈頁/頁〉	〈須〉 ⿱彡女	28 1	93.2 0.2
〈穎〉	⿹〈頃〉〈耒〉 ⿰⿱〈〈匕〉/七〉〈耒〉〈頁/頁〉	〈頃〉 ⿱〈〈匕〉/七〉〈耒〉	20 1	90.7 0.3
〈穎〉	⿹〈頃〉禾 ⿰⿱〈〈匕〉/七〉禾〈頁/頁〉	〈頃〉 ⿱〈〈匕〉/七〉禾	20 2	82.6 0.9
〈頴〉	⿹〈頃〉示 ⿰〈⿱匕示〉〈頁/頁〉	〈頃〉 〈⿱匕示〉	20 8	51.0 8.2

〈㲉〉	⿹⿰⿱士冖〈殳〉〈孚〉 ⿰⿱⿱士冖〈孚〉〈殳〉	⿰⿱士冖〈殳〉 ⿱⿱士冖〈孚〉	11 1	84.0 0.7
〈𣪊〉	⿹⿰⿱士冖〈殳〉土 ⿰⿱⿱士冖土〈殳〉	⿰⿱士冖〈殳〉 ⿱⿱士冖土	11 1	84.0 0.7
Same as above	〈𨣊〉, 〈縠〉, 〈觳〉, 〈嗀〉, 〈𣫫〉, 〈轂〉			
〈穀〉	⿹⿰⿱士冖〈殳〉⿱口米 ⿰⿱⿱士冖⿱口米〈殳〉	⿰⿱士冖〈殳〉 ⿱⿱士冖⿱口米	11 2	71.6 2.4
〈嚳〉	⿹⿰⿱士冖〈殳〉告 ⿰⿱⿱士冖告〈殳〉	⿰⿱士冖〈殳〉 ⿱⿱士冖告	11 2	71.6 2.4
〈犨〉	⿹⿰⿱士冖〈殳〉牛 ⿰⿱⿱士冖牛〈殳〉	⿰⿱士冖〈殳〉 ⿱⿱士冖牛	11 2	71.6 2.4
嚳	⿹𣪊告 ⿰⿱壳告殳	𣪊 ⿱壳告	10 1	82.6 0.9
犨	⿹𣪊牛 ⿰⿱壳牛殳	𣪊 ⿱壳牛	10 1	82.6 0.9
Same as above	嗀, 轂, 毂, 觳, 穀, 縠			
叡	⿹〈叔〉谷 ⿰⿱〈卢〉谷又	〈叔〉 ⿱〈卢〉谷	7 1	76.6 1.6
〈叡〉	⿹〈叔〉〈谷/谷〉 ⿰⿱〈卢〉〈谷/谷〉又	〈叔〉 ⿱〈卢〉〈谷/谷〉	7 1	76.6 1.6
䜣	⿹〈叔〉〈贝〉 ⿰⿱〈卢〉〈贝〉又	〈叔〉 ⿱〈卢〉〈贝〉	7 2	60.5 5.0
〈䜣〉	⿹〈叔〉〈贝〉 ⿰⿱〈卢〉〈贝〉又	〈叔〉 ⿱〈卢〉〈贝〉	7 2	60.5 5.0
〈叡〉	⿹〈叔〉貝 ⿰⿱〈卢〉貝又	〈叔〉 ⿱〈卢〉貝	7 2	60.5 5.0
〈𨔐〉	⿹〈台彐〉立 ⿰⿱台立彐	〈台彐〉 ⿱台立	7 1	76.6 1.6

TABLE 8. ⿴⿱ AEM ↔ ⿱ A ⿴ EM if E is *kamae* (411)

char	structure	component	p*n*	accuracy
〈卣〉	⿴占〈冂〉 ⿱〈⺊〉⿴口〈冂〉	占 ⿴口〈冂〉	199 1	99.0 0.1
鹵	⿴占〈※〉 ⿱〈⺊〉㐅	占 㐅	199 18	84.1 0.7

char	structure	component	p*n*	accuracy
⟨卤⟩	⿴**占**⟨乂⟩ ⿱⟨⺊⟩⟨囟⟩	**占** ⟨囟⟩	199 40	69.3 2.9
⟨互⟩	⿴⟨互⟩丶 ⿱一⿴⟨彑⟩丶	⟨互⟩ ⿴⟨彑⟩丶	58 1	96.6 0.1
⟨囱⟩	⿴白⟨夕/夂/夊⟩ ⿱丿 ⿴**囗**⟨夕/夂/夊⟩	白 ⿴**囗**⟨夕/夂/夊⟩	33 1	94.2 0.1
囱	⿴白**夂** ⿱丿 囱	白 囱	33 4	79.5 1.2
⟨囪⟩	⿴白㐅 ⿱丿 囟	白 囟	33 4	79.5 1.2
囟	⿴白⟨乂⟩ ⿱丿 ⟨囟⟩	白 ⟨囟⟩	33 40	20.4 30.1

Table 9. ⿴⿱ ABM ↔ ⿱ A ⿱ MB if both A and B are not enclosure (414)

char	structure	component	p*n*	accuracy
⟨衰⟩	⿴⟨衣⟩⟨丑⟩ ⿱⟨亠⟩⿱⟨丑⟩⟨𧘇⟩	⟨衣⟩ ⿱⟨丑⟩⟨𧘇⟩	389 1	99.5 0.1
⟨褩⟩	⿴⟨衣⟩⟨殳⟩ ⿱⟨亠⟩⿱⟨殳⟩⟨𧘇⟩	⟨衣⟩ ⿱⟨殳⟩⟨𧘇⟩	389 1	99.5 0.1
Same as above	⟨襲⟩, ⟨衰⟩, ⟨褻⟩, ⟨褒⟩, ⟨褏⟩, ⟨褩⟩, ⟨褵⟩, ⟨褱⟩, ⟨褩⟩, ⟨褢⟩, ⟨褻⟩, ⟨褭⟩, ⟨褭⟩, ⟨褏⟩, ⟨褰⟩, ⟨褩⟩, ⟨褻⟩, ⟨褢⟩, ⟨褻⟩, ⟨褢⟩, ⟨褻⟩, ⟨褢⟩, ⟨褢⟩, ⟨褒⟩, ⟨衺⟩, ⟨衰⟩, ⟨衰⟩, ⟨表⟩, ⟨衰⟩, ⟨裂⟩, ⟨衷⟩, ⟨衷⟩, ⟨衰⟩, ⟨衷⟩, ⟨衰⟩, ⟨衰⟩, ⟨衮⟩, ⟨衰⟩, ⟨衾⟩, ⟨衰⟩, ⟨衷⟩, ⟨表⟩, ⟨衮⟩, ⟨衰⟩, ⟨衮⟩, ⟨表⟩, ⟨衰⟩, ⟨衮⟩, ⟨褭⟩, ⟨褻⟩, 裹, ⟨褭⟩, ⟨褭⟩, ⟨褒⟩, ⟨褭⟩, ⟨褭⟩, **喪**, ⟨褭⟩, 裛, ⟨衰⟩, ⟨裏⟩, ⟨褰⟩, ⟨衰⟩, 衰, ⟨衰⟩, ⟨褭⟩, ⟨衰⟩, ⟨衷⟩, ⟨衰⟩, ⟨衮⟩, ⟨褻⟩, ⟨衮⟩			
⟨褻⟩	⿴⟨衣⟩⟨珏⟩ ⿱⟨亠⟩褻	⟨衣⟩ 褻	389 2	99.0 0.1
⟨襄⟩	⿴⟨衣⟩罡 ⿱⟨亠⟩⿱罡⟨𧘇⟩	⟨衣⟩ ⿱罡⟨𧘇⟩	389 2	99.0 0.1
Same as above	⟨哀⟩, ⟨襄⟩			
褻	⿴**衣埶** ⿱亠⟨埶⟩	**衣** ⟨埶⟩	185 1	98.9 0.1
裧	⿴**衣火** ⿱亠⟨炏⟩	**衣** ⟨炏⟩	185 1	98.9 0.1

Same as above	褏, 襃, 褰, 褰, 衮, 裘, 裛, 裦, 衾, 衺, 裵, 衷, 表, 袞, 褻, 亵, 衮, 衮, 豪, 褏, 褑, 裒, 衮, 衰, 衰, 裒, 褻, 褢, 袞, 裹, 褱, 衷, 衰, 囊, 㲦, 䙪, 䙫, 䙬, 褭, 褻, 褷, 褻, 褭, 褧, 褩, 褻, 褒, 褒, 裴, 袁, 袌, 衮, 衮, 表, 衮, 衷, 裹, 裹, 衮			
襄	⿴衣𠀎 ⿱亠⿱𠀎𧘇	衣 ⿱𠀎𧘇	185 2	97.9 0.1
襄	⿴衣㗊 ⿱亠⿱㗊𧘇	衣 ⿱㗊𧘇	185 2	97.9 0.1
Same as above	褭, 褰			
哀	⿴衣口 ⿱亠𠮛	衣 𠮛	185 7	92.8 0.2
〈啻〉	⿴音中 ⿱立𠯑	音 𠯑	153 1	98.7 0.1
〈龠〉	⿴〈侖〉〈⺍〉 ⿱〈亼/𠆢/今/亼/厶〉⿱〈⺍〉〈冊/册/冊〉	〈侖〉 ⿱〈⺍〉〈冊/册/冊〉	83 1	97.6 0.1
〈亟〉	⿴〈二〉⿲口㇉口 ⿱一⿱⿲口㇉口一	〈二〉 ⿱⿲口㇉口一	31 1	93.8 0.1
〈亙〉	⿴〈二〉舟 ⿱一⿱舟一	〈二〉 ⿱舟一	31 1	93.8 0.1
亘	⿴〈二〉回 ⿱一⿱回一	〈二〉 ⿱回一	31 2	88.2 0.4
〈亙〉	⿴〈二〉⺕ ⿱一⿱⺕一	〈二〉 ⿱⺕一	31 2	88.2 0.4
〈亘〉	⿴〈二〉回 ⿱一⿱回一	〈二〉 ⿱回一	31 2	88.2 0.4
橐	⿴柬包 ⿱声⿱包木	柬 ⿱包木	21 1	91.1 0.2
櫜	⿴柬毀 ⿱声⿱毀木	柬 ⿱毀木	21 1	91.1 0.2
Same as above	橐, 橐, 〈櫜〉, 〈橐〉, 〈橐〉, 〈橐〉, 〈橐〉, 〈橐〉, 〈橐〉, 〈橐〉			
橐	⿴柬〈㗊〉 ⿱声⿱〈㗊〉木	柬 ⿱〈㗊〉木	21 2	83.4 0.8
〈橐〉	⿴柬〈㗊〉 ⿱声⿱〈㗊〉木	柬 ⿱〈㗊〉木	21 2	83.4 0.8
橐	⿴柬〈臼〉 ⿱声〈梟〉	柬 〈梟〉	21 3	76.6 1.6

橐	⿴㯻〈石〉 ⿱声〈橐〉	㯻 〈橐〉	21 9	49.0 9.1
〈櫜〉	⿴㯻**非** ⿱声**棐**	㯻 **棐**	21 13	38.1 14.7
𦾰	⿴廾**黎** ⿱〈艹〉⿱**黎**廾	廾 ⿱**黎**廾	17 1	89.2 0.3
莾	⿴廾**大** ⿱〈艹〉⿱**大**廾	廾 ⿱**大**廾	17 1	89.2 0.3
Same as above	蕢, 藻, 葆, 蕉, 菲, 蓐, 葺, 菲, 蒸, 茸, 〈葬〉			
櫑	⿴〈畾〉桐 ⿱⿰**田田**⿱桐⿰**田田**	〈畾〉 ⿱桐⿰**田田**	12 1	85.2 0.6
〈罍〉	⿴〈畾〉〈缶/**缶**/缶〉 ⿱⿰**田田**⿱〈缶/**缶**/缶〉⿰**田田**	〈畾〉 ⿱〈缶/**缶**/缶〉⿰**田田**	12 1	85.2 0.6
Same as above	罍, 〈罍〉, 〈櫑〉			
䨻	⿴〈畾〉**辰** ⿱⿰**田田**⿱**辰**⿰**田田**	〈畾〉 ⿱**辰**⿰**田田**	12 2	73.5 2.1
〈䨻〉	⿴〈畾〉**辰** ⿱⿰**田田**⿱**辰**⿰**田田**	〈畾〉 ⿱**辰**⿰**田田**	12 2	73.5 2.1
罍	⿴畾**缶** ⿱⿰**田田**⿱**缶**⿰**田田**	畾 ⿱**缶**⿰**田田**	12 1	85.2 0.6
〈囂〉	⿴〈㗊〉莫 ⿱〈吅〉⿱莫〈吅〉	〈㗊〉 ⿱莫〈吅〉	11 1	84.0 0.7
〈器〉	⿴〈㗊〉⿰**太太** ⿱〈吅〉⿱⿰**太太**〈吅〉	〈㗊〉 ⿱⿰**太太**〈吅〉	11 1	84.0 0.7

TABLE 10. ⿵⿱⿰ LRBA -> ⿱⿲ LARB (511)

char	structure	component	p*n*	accuracy
〈鬻〉	⿵〈鬻〉〈咸〉 ⿱⿲**弓**〈咸〉**弓**〈鬲〉	〈鬻〉 ⿲**弓**〈咸〉**弓**	43 1	95.5 0.1
〈鬻〉	⿵〈鬻〉〈去〉 ⿱⿲**弓**〈去〉**弓**〈鬲〉	〈鬻〉 ⿲**弓**〈去〉**弓**	43 1	95.5 0.1
Same as above	〈鬻〉, 〈鬻〉			

〈⿱⿲弓束弓鬲〉	⿶〈⿱弜鬲〉束 ⿱〈⿲弓束弓〉〈鬲〉	〈⿱弜鬲〉 〈⿲弓束弓〉	43 2	91.3 0.2
〈⿱⿲弓米弓鬲〉	⿶〈⿱弜鬲〉米 ⿱⿲弓米弓〈鬲〉	〈⿱弜鬲〉 ⿲弓米弓	43 2	91.3 0.2
Same as above	〈⿱⿲弓粦弓鬲〉, 〈⿱⿲弓付弓鬲〉, 〈⿱⿲弓翟弓鬲〉, 〈⿱⿲弓建弓鬲〉, 〈[illegible]〉, 〈⿱⿲弓酉弓鬲〉, 〈⿱⿲弓沸弓鬲〉, 〈⿱⿲弓圭弓鬲〉			
〈⿱⿲弓芻弓鬲〉	⿶〈⿱弜鬲〉〈芻〉 ⿱〈⿲弓芻弓〉〈鬲〉	〈⿱弜鬲〉 〈⿲弓芻弓〉	43 3	87.4 0.5
〈⿱⿲弓会弓鬲〉	⿶〈⿱弜鬲〉〈会〉 ⿱⿲弓〈会〉弓〈鬲〉	〈⿱弜鬲〉 ⿲弓〈会〉弓	43 3	87.4 0.5
〈⿱⿲弓每弓鬲〉	⿶〈⿱弜鬲〉每 ⿱⿲弓每弓〈鬲〉	〈⿱弜鬲〉 ⿲弓每弓	43 3	87.4 0.5
〈⿱⿲弓孝弓鬲〉	⿶〈⿱弜鬲〉孝 ⿱⿲弓孝弓〈鬲〉	〈⿱弜鬲〉 ⿲弓孝弓	43 3	87.4 0.5
⿱⿲弓曾弓鬲	⿶⿱弜鬲曾 ⿱⿲弓曾弓鬲	⿱弜鬲 ⿲弓曾弓	19 1	90.3 0.3
⿱⿲弓辱弓鬲	⿶⿱弜鬲辱 ⿱⿲弓辱弓鬲	⿱弜鬲 ⿲弓辱弓	19 1	90.3 0.3
Same as above	[illegible], ⿱⿲弓焦弓鬲, [illegible], [illegible]			
⿱⿲弓弗弓鬲	⿶⿱弜鬲弗 ⿱⿲弓弗弓鬲	⿱弜鬲 ⿲弓弗弓	19 2	81.9 0.9
⿱⿲弓酉弓鬲	⿶⿱弜鬲酉 ⿱⿲弓酉弓鬲	⿱弜鬲 ⿲弓酉弓	19 2	81.9 0.9
Same as above	⿱⿲弓建弓鬲, ⿱⿲弓沸弓鬲, ⿱⿲弓圭弓鬲, ⿱⿲弓翟弓鬲, [illegible], [illegible]			
⿱⿲弓会弓鬲	⿶⿱弜鬲〈会〉 ⿱⿲弓〈会〉弓鬲	⿱弜鬲 ⿲弓〈会〉弓	19 3	74.6 1.9
⿱⿲弓每弓鬲	⿶⿱弜鬲每 ⿱⿲弓每弓鬲	⿱弜鬲 ⿲弓每弓	19 3	74.6 1.9
⿱⿲弓孝弓鬲	⿶⿱弜鬲孝 ⿱⿲弓孝弓鬲	⿱弜鬲 ⿲弓孝弓	19 3	74.6 1.9
⿱⿲弓咸弓鬲	⿶⿱弜鬲咸 ⿱⿲弓咸弓鬲	⿱弜鬲 ⿲弓咸弓	10 1	82.6 0.9
⿱⿲弓天天弓鬲	⿶⿱弜鬲天天 ⿱⿲弓天天弓鬲	⿱弜鬲 ⿲弓天天弓	10 1	82.6 0.9
⿱⿲弓采弓鬲	⿶⿱弜鬲采 ⿱⿲弓采弓鬲	⿱弜鬲 ⿲弓采弓	10 1	82.6 0.9

鬻	⿰鬻**去** ⿱⿰**弓去弓**鬲	鬻 ⿰**弓去弓**	10 1	82.6 0.9
鬻	⿰鬻〈粥〉 ⿱⿰**弓**〈粥〉**弓**鬲	鬻 ⿰**弓**〈粥〉**弓**	10 2	69.4 2.8
鬻	⿰鬻**付** ⿱⿰**弓付弓**鬲	鬻 ⿰**弓付弓**	10 2	69.4 2.8

Table 11. ⿰⿱ A ⿰ LRC ↔ ⿱ A ⿰ LCR (611)

char	structure	component	p*n*	accuracy
〈斎〉	⿰〈齐〉**貝** ⿱**文**⿰**丿 貝 丨**	〈齐〉 ⿰**丿 貝 丨**	30 1	93.7 0.1
〈斖〉	⿰〈齐〉**韭** ⿱**文**⿰**丿 韭 丨**	〈齐〉 ⿰**丿 韭 丨**	30 1	93.7 0.1
Same as above	〈斎〉, 〈斎〉, 〈斎〉			
斎	⿰〈齐〉**示** ⿱**文**⿰**丿 示 丨**	〈齐〉 ⿰**丿 示 丨**	30 3	82.6 0.9
〈羸〉	⿰〈嬴〉〈糸〉 ⿱〈言〉⿰〈月〉〈糸〉〈凡〉	〈嬴〉 ⿰〈月〉〈糸〉〈凡〉	23 1	91.8 0.2
〈贏〉	⿰〈嬴〉〈贝〉 ⿱〈言〉⿰〈月〉〈贝〉〈凡〉	〈嬴〉 ⿰〈月〉〈贝〉〈凡〉	23 1	91.8 0.2
Same as above	贏, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉, 〈贏〉			
〈羸〉	⿰〈嬴〉**馬** ⿱〈言〉⿰〈月〉**馬**〈凡〉	〈嬴〉 ⿰〈月〉**馬**〈凡〉	23 2	84.6 0.7
羸	⿰嬴**魚** ⿱言⿰〈月〉**魚凡**	嬴 ⿰〈月〉**魚凡**	16 1	88.6 0.4
羸	⿰嬴叟 ⿱言⿰〈月〉叟**凡**	嬴 ⿰〈月〉叟**凡**	16 1	88.6 0.4
Same as above	羸, 羸, 羸, 羸, 羸, 羸, 羸			
〈齋〉	⿰齊〈月/冃/月/**月**〉 ⿱亝⿰**丿**〈月/冃/月/**月**〉**丨**	齊 ⿰**丿**〈月/冃/月/**月**〉**丨**	7 1	76.6 1.6
齋	⿰齊**魚** ⿱亝⿰**丿 魚 丨**	齊 ⿰**丿 魚 丨**	7 2	60.5 5.0

〈齾〉	⿰齊魚 ⿱亝⿰丿魚丨	齊 ⿰丿魚丨	7 2	60.5 5.0
贏	⿰嬴虫 ⿱亡⿰〈月〉虫孔	嬴 ⿰〈月〉虫孔	7 1	76.6 1.6
羸	⿰嬴羊 ⿱亡⿰〈月〉羊孔	嬴 ⿰〈月〉羊孔	7 1	76.6 1.6
贏	⿰嬴〈果〉 ⿱亡⿰〈月〉〈果〉孔	嬴 ⿰〈月〉〈果〉孔	7 2	60.5 5.0

Table 12. ⿰⿱ AM ⿰ LRC ↔ ⿱ AM ⿰ LCR (612)

char	structure	component	p*n*	accuracy
〈贏〉	⿰〈嬴〉貝 ⿱〈亡〉口⿰〈月〉貝〈耳〉	〈嬴〉 ⿰〈月〉貝〈耳〉	4 1	64.0 4.0
〈羸〉	⿰〈嬴〉羊 ⿱〈亡〉口⿰〈月〉羊〈耳〉	〈嬴〉 ⿰〈月〉羊〈耳〉	4 1	64.0 4.0
〈贏〉	⿰〈嬴〉女 ⿱〈亡〉口⿰〈月〉女〈耳〉	〈嬴〉 ⿰〈月〉女〈耳〉	4 1	64.0 4.0

In many cases, especially in the trivial ones, the accuracy of the functional structures is high. Often these results are also consistent with the etymological structure. For example, in the case of "脩" (table 2), the accuracy of ⿰攸田 is lower than the accuracy of ⿰亻备. In fact, ⿰亻备 seems to correspond to the etymological structure. However, since these results were calculated with respect to a simple count of the number of CHISE character objects, productivity of a component was divided by the number of glyph-variants of the component. For example, component 〈嬴〉 is written in various forms such as "嬴," "嬴," "嬴," "嬴," "嬴," "嬴," "嬴," "嬴," "嬴," "嬴," etc. As the result, the productivity of apparent components (⿰月*x*凡, ⿰月*x*耳, etc.) increases and the accuracy of the functional structure decreases. Therefore, if information on variant-relations is available, it is better to normalize the glyphs (close to the original forms), its structure and components.

5. Conclusion

Structural description of Chinese character should be based on Chinese character analysis (Chinese character studies), like grammatical analysis of natural language. For this reason, it is desirable to describe Hanzi structure based on etymological explanations. Etymological knowledge

is required to perform the task, however this information is missing for many atypical Chinese characters, in fact the *majority* of CJKV Unified Ideographs of UCS.

On the other hand, if we consider the "(degree of) componentness" of Chinese characters, a (candidate for) component that produces more Chinese characters is more likely to be considered as an actual component. Considering components with respect to productivity does *not* require any etymological knowledge on characters and therefore can be calculated whenever we have a dataset of Hanzi structure descriptions at our disposal.

Based on this hypothesis, we conducted an experiment using the structural data on Chinese characters from the CHISE character ontology. As a result, we found that components based on etymological knowledge are more likely to have a higher productivity.

References

Asahara, Tatsuro [浅原達郎] (1996). 漢字の字符 *[Chinese Character Graphemes]*. URL: http://yuetgu.zinbun.kyoto-u.ac.jp:8098/yg/rs/jihu.pdf.

Information technology—Universal Coded Character Set (UCS) (2014). ISO/IEC 10646:2014. International Organization for Standardization (ISO).

Morioka, Tomohiko [守岡知彦] (2015). "Multiple-policy Character Annotation based on CHISE." In: *Journal of the Japanese Association for Digital Humanities* 1.1, pp. 86–106.

——— (2018a). "Integration of a Chinese Character Ontology Title and Historical Glyph Examples." In: *9th International Conference of Digital Archives and Digital Humanities (DADH 2018)*. Taiwanese Association for Digital Humanities / Dharma Drum Institute of Liberal Arts, pp. 287–300.

——— (2018b). "項書き換え系を用いた漢字字体の包摂規準の形式化の試み [An Attempt to Formalize Unification Rules of Chinese Characters Based on Term Rewriting System]." In: 情報処理学会論文誌 *[Journal of the Information Processing Society of Japan]* 59.2, pp. 332–340.

Slaměníková, Tereza (2019). "On the Nature of Unmotivated Components in Modern Chinese Characters." In: *Proceedings of Graphemics in the 21st Century, Brest 2018*. Ed. by Yannis Haralambous. Brest: Fluxus Editions, pp. 209–226.