

# Can a logographic script be meaningfully simplified and remain logographic?

Lessons from the 20th century Chinese writing  
reform informed by recent psycholinguistic  
research<sup>\*</sup>

Zev HANDEL

University of Washington, U.S.A.

## 1. Introduction

In the 1950s and 1960s, the government of the People's Republic of China undertook, in two stages, a carefully planned “simplification” of the logographic Chinese script. Drawing on a variety of historical precedents, over 2,000 individual graphs were modified in an attempt to make the script easier to learn and use. This was the first significant change in the official form of the Chinese script in nearly two millennia.

---

<sup>\*</sup> I would like to thank James Myers and Alex de Voogt for pointing me to helpful references in psycholinguistics, and the organizers and attendees of the Venice conference who heard and commented on an earlier version of this talk. This version is a rather disorganized draft which I intend to rework into something more coherent and compact.

This study begins with a broad overview of the structure and function of the Chinese script (section 2), followed by a detailed description of the history of simplification (section 3) and of the approaches and techniques that were used last century by the government's Committee on Language Reform to devise the simplified character forms (section 4). This description is followed by an examination of the relationship between the simplification of individual graphs and the complexity and coherence of the system as a whole (sections 5 and 6). The psycholinguistic research of the last two decades allows us to go beyond insightful but speculative criticisms of the simplification process (cf. Chen 1999, Ramsey 1987, DeFrancis 1984:214-216, Hannas 1997) and to re-evaluate the 20th-century simplification in objective terms related to reading strategies and psychological representation. It thus provides a more sophisticated framework for evaluating the claims that were made on behalf of simplification by early- and mid-20th century advocates, many of which had little or no scientific basis.

In the last part of the paper I consider the practical implications of the script reform in societal terms (section 7). The conclusion cannot be avoided that this historic script reform has, despite the great effort and cost involved in its implementation, had negligible practical effects.

This conclusion concerning the 20th century simplification leads to an interesting theoretical question: Can a logographic writing system with the basic structural properties of Chinese be effectively simplified—in a meaningful way that addresses both learning efficiency and reading efficiency—while remaining typologically logographic? Or are real gains in “simplification” only achievable through a typological shift (to, say, a syllabary or alphabet)? Is there a theoretical “simplification” different from the one that actually took place that might have been more effective? This

Can a logographic script be meaningfully simplified and remain logographic?

question is first engaged through a discussion of the unsuccessful 1976 proposals for further character simplification (section 8).

The question can be further elucidated through an understanding of the roles of structural elements of Chinese characters in the cognitive processes associated with recognition and reading. The final part of this study (section 9) begins with a review of the key psycholinguistic studies on the recognition and processing of Chinese characters in young learners and skilled adult readers. The hypotheses advanced by the authors of these studies suggest a mechanism for the construction of a Chinese logographic script, derived through a simplification of the existing script, that would present real advantages to its learner and users. Such a simplified script is described from a theoretical perspective, along with consideration of the practical impediments to implementation (section 10).

## **2. Chinese characters - structure and function**

In order to understand the methods of simplification that were employed in the 20<sup>th</sup> century and the motivations behind them, and to evaluate their effects, it is first necessary to review the basic structure of modern Chinese characters and the nature of the Chinese writing system as a whole. This section is intended for those readers who are not already familiar with the Chinese script.

Chinese writing is usually described as logographic.<sup>1</sup> Although the precise nature of the writing system from its formative period through its

---

<sup>1</sup> I use the term *logographic* in its widely accepted sense, referring to a writing system whose graphic units represent individual morphemes of the spoken language; these units are called *logographs* or *logograms*. Unger & DeFrancis 1995 object to this characterization of Chinese writing, but they do so by defining the term logographic anew, in a way that precludes any writing system from being logographic, and thereby render the term useless as a way of characterizing writing systems like Chinese whose graphic referents are primarily morphemes rather than phoneme

first systematization—roughly the 13<sup>th</sup> century BCE through the 3<sup>rd</sup> century BCE—remains a matter of some controversy, over the last 2,000 years it can be accurately characterized as *morphosyllabic* or perhaps more clumsily *morphosyllabographic* (DeFrancis 1984:88). By this is meant that the overwhelming majority of Chinese characters as conventionally employed in the Chinese writing system are logographs representing monosyllabic morphemes of the spoken language. Since the vast majority of spoken Chinese morphemes are monosyllabic, there exists a nearly isomorphic relationship between written graphs on the one hand and spoken syllables/morphemes on the other. If we are speaking specifically of Modern Standard Written Chinese, then the morphemes involved are those occurring in Modern Standard Mandarin.<sup>2</sup> Because a morpheme, by definition, has both phonological shape and semantic content, each Chinese character also has an associated pronunciation and meaning, namely the pronunciation and meaning of the morpheme with which it is or once was conventionally associated.<sup>3</sup> For native users of the script, these linguistic features are thought of as inhering within the written graph itself.

The monosyllabic and logographic nature of individual characters can be seen in the examples below. The fact that distinct but homophonous

---

strings. This slight of hand is achieved by redefining *logogram* in terms of its internal structure rather than its referent, as a unit of writing that contains no visual clue to its pronunciation (1995:45, 50). While this conception may be useful as a way of thinking about the internal structure and function of graphs, it simply confuses things to repurpose an existing technical term in a novel way.

<sup>2</sup> In practice, the writing system permits the imposition of different phonological systems onto the syntax and lexicon of the standard written language. Thus, a Hong Kong speaker of Cantonese can read Standard Written Chinese—which essentially reflects the lexicon and syntax of modern standard Mandarin—aloud using Cantonese phonological patterns. Note that Standard Written Chinese with Cantonese phonology is quite different from the written representation of spoken Cantonese.

<sup>3</sup> In the case of obsolete or rare characters, the modern phonological associations can in a sense be considered artificial in that no morpheme exists in the modern spoken language; nevertheless, the associations are not arbitrary as they are determined on a historical-textual basis.

Can a logographic script be meaningfully simplified and remain logographic?

morphemes are represented by distinct graphs demonstrates the logographic, rather than syllabographic, nature of the script. Although these examples are all of Standard Mandarin usage, the same principle applies for written forms of other Chinese varieties, such as Cantonese.<sup>4</sup>

人	<i>rén</i> 'person' (free morpheme)
仁	<i>rén</i> 'humane' (bound morpheme)
糖	<i>táng</i> 'sugar' (free morpheme)
堂	<i>táng</i> 'hall' (bound morpheme)
十	<i>shí</i> 'ten' (free morpheme)
食	<i>shí</i> 'eat, food' (bound morpheme)

Many words of Modern Standard Mandarin are bimorphemic and bisyllabic; it follows that such words are written with two Chinese characters, each representing a component morpheme of the compound:

食堂	<i>shítáng</i> 'dining hall'
----	------------------------------

A relatively small number of Chinese characters are regularly employed solely phonographically, as syllabographs, most often when writing foreign loanwords or in the transcription of non-Chinese proper names. In addition, many Chinese characters that are ordinarily employed logographically can optionally be employed phonographically in transcriptional context. For example

---

<sup>4</sup> In the text of this paper all written Chinese is in traditional Chinese characters, except when examples of simplified forms are explicitly given. In the bibliography, names and titles are given as they appear in the publications themselves. Romanizations are in *Hànyǔ Pīnyīn*, the official transcription system of the People's Republic of China.

加拿大 Jiānádà ‘Canada’

is written with three graphs employed phonographically; in default contexts they write the three free morphemes

加	<i>jiā</i> ‘to add’
拿	<i>ná</i> ‘to take, hold’
大	<i>dà</i> ‘big’

There are also a number of characters that are used to write individual syllables of bisyllabic morphemes.<sup>5</sup> These characters cannot however be considered purely phonographic syllabographs, since they are restricted in use to specific lexical items—they are, in other words, morphemically contingent. For this reason native speakers tend to view them as independently containing semantic content, and this view is reflected in—or perhaps conditioned and reinforced by—their lexicological treatment as head entries in character dictionaries. Another way of thinking about this aspect of the writing system is that because the vast majority of Chinese characters are employed to represent monosyllabic morphemes, there are strong systemic pressures to use and interpret *all* characters in this way. While monosyllabism generally trumps morphemicity—that is to say, a bisyllabic morpheme is nearly always written with two characters rather than one—there is an unmistakable tendency for script users to impose a morphemic identity on such characters.

As an example, consider the monomorphemic bisyllabic Chinese words *shānhú* ‘coral’, *húdié* ‘butterfly’, and *húlu* ‘gourd’. In all three words, *hú* constitutes a meaningless syllable, much like the syllable *cor* of English ‘coral’ and ‘rancor’. The words are written this way:

---

<sup>5</sup> Although bisyllabic morphemes constitute a small minority of the Chinese morpheme inventory, many of them occur with high frequency. They tend to be found most often in certain semantic domains, such as the names of plants and insects.

Can a logographic script be meaningfully simplified and remain logographic?

- (1) 珊瑚 *shānhú* ‘coral’
- (2) 蝴蝶 *húdié* ‘butterfly’
- (3) 葫蘆 *húlu* ‘gourd’

Each of the three *hú* characters (瑚, 蝴, 葫) are typically listed as separate entries in a Chinese dictionary, just like the majority of characters that represent morphemes. A typical entry for 瑚, for example, might read “瑚 ‘coral’ -- see 珊瑚”. Native speakers have a strong tendency to ascribe meanings to these characters in isolation, even though they only occur in collocated form.<sup>6</sup>

Finally, in modern usage there are also a handful of characters that write bimorphemic monosyllables (which are exceptionally rare in the spoken language; most are contracted forms) and bisyllabic compound words. For example, the modern standard Mandarin monosyllable *bié* ‘do not (imperative)’ is a contracted form of the bisyllabic compound *búyào*, and is written with the single character 別. The character 廿 is used to write the bisyllabic, two-morpheme word *èrshí* ‘20’.<sup>7</sup> There is also one character regularly employed in modern written Chinese to represent the sub-syllabic derivational morpheme *-r*. Very few characters are used in these ways (although some of these character appear with high frequency), and they can be considered marginal from the perspective of the writing system typology as a whole, in much the same way that the use of certain glyphs like ‘&’ can be considered marginal in Western alphabetic systems.

---

<sup>6</sup> Hannas (1997:176-178) has argued that the aspects of the writing system just described have had a profound effect on spoken Chinese, inhibiting changes to the basic typology of the language’s morphology. This is an interesting claim, but such a strong hypothesis cannot be accepted without a more rigorous evidence-based argument.

<sup>7</sup> The character also has a one-syllable reading, *niàn*; this reading is considered learned, and is not used in ordinary speech.

When Chinese orthography is considered in terms of its correlation with units of spoken language, the basic unit of the writing system is clearly the Chinese character. However, it is crucially important to recognize that the characters are themselves highly structured internally; many psycholinguistic studies have demonstrated that this internal structure is salient at both a conscious and unconscious level for script users, as will be elaborated below. Aspects of the internal structure bear some relationship with spoken language, but some internal elements can only be understood in purely graphic terms as units of a graphic system without consideration of the spoken language.<sup>8</sup>

At the most basic level, script users view Chinese characters as composed of “strokes” (*bǐhuà* 筆畫), which can be defined as movements of the writing utensil that cannot be interrupted by lifting it from the writing surface.<sup>9</sup> Strokes are highly salient for users of Chinese characters. When children or second-language learners first learn how to write characters, they are taught to do so stroke by stroke.

Some common methods of lexicographic ordering of Chinese characters depend on the number of strokes, the order of strokes, and/or the type of strokes found in each character. In practice, this means that not only do characters have standardized forms, but they also have highly salient and frequently reinforced normative ways of being written. While a user of the Latin alphabet may write the letter “t” by drawing the horizontal cross-stroke first or last, or by drawing the vertical stroke from top to bottom or from bottom to top, a user of the Chinese script is not granted the same leeway. Proper stroke direction and stroke order are

---

<sup>8</sup> Myers (2011) argues that the formal regularities seen in Chinese character structure, which psycholinguistic studies have demonstrated are “psychologically real”, justify the interpretation of these regularities as a sort of “phonology”, analogous to a grammatical system of spoken language.

<sup>9</sup> Within China there is a sophisticated native tradition of stroke analysis, closely tied to calligraphic practice.



Can a logographic script be meaningfully simplified and remain logographic?

taught in school. And while violations of these norms are not uncommon in practice, lexicographic conventions ensure that the normative stroke directions and orderings remain well known even in cases where they are not widely followed.

Figure 1



Figure 1 above is from a Chinese language textbook for American university students.<sup>10</sup> The top left box shows that the character 我 is composed of seven strokes; the positioning of the numerals ‘1’ through ‘7’ indicates the starting points, and thus the direction, of each stroke. For example, stroke 4 is to be drawn from left to right, while stroke 6 is to be drawn from right to left. In the seven boxes of the second-to-last row, the method of writing the character is illustrated by building it up stroke by stroke. (The other boxes are provided for the student to practice writing the character. The student is to start by tracing the character in the top row. The divided boxes in the next row serve as a guide to keep the character centered and balanced as it is written. After completing the first two rows, the student finally moves to the blank boxes in the third and fourth row.)

<sup>10</sup> Integrated Chinese, Traditional Character Edition, Level 1 Part 1 Character Workbook, first edition (Boston: Cheng & Tsui, 1997).

While strokes are a culturally salient and aesthetically important component of characters, they are not systemically significant. Generally speaking, stroke distinctions are not graphemic; in other words, switching out one stroke for another will not often produce a meaningfully contrastive graphic form.<sup>11</sup> There is a unit of graphic structure in between the individual stroke and the whole graph that plays a crucial role in the writing system; for want of a better term this unit is commonly referred to as a “character component” or “character element”; I will use the terms “component” and “element” interchangeably.<sup>12</sup> Components recur across multiple characters and are to a large degree *distinctive*, or *graphemic*: exchanging components yields contrasting graphic forms.<sup>13</sup> Moreover, while some components are simply frequently recurring meaningless groupings of strokes, many have an identifiable function within the character. The two most common functions are that of semantic component (also called *signific* or *determinative*)<sup>14</sup> and phonetic component

---

<sup>11</sup> I use the term “graphemic” here by analogy with “phonemic”; just as phonetic features are deemed phonemic if they are contrastive between phonemes, graphic features are deemed graphemic if they are contrastive between graphemes.

<sup>12</sup> Unfortunately, there is no scholarly consensus on the classification and terminology of components of Chinese characters. Usage varies widely, and one term (e.g. *radical*) may be employed with significant differences by different researchers.

<sup>13</sup> The contrasting graphic forms may not exist in the writing system, but they will appear to users of the script as plausible but unrecognized characters (often termed *pseudo-characters* in the literature), which could serve as potential new graphs to be added to the writing system.

<sup>14</sup> The semantic component is also frequently referred to as a *radical* or *classifier* (*bùshǒu* 部首). These terms properly refer to the character element under which the character is arranged in a lexicographic classification. This element is often, but not always, the character’s semantic element, which is the reason for the confusion and overlap in usage. Most linguists and philologists frown on the use of the term *radical* in graphic analysis. Nevertheless, it is a commonly used term among non-academic students of Chinese characters and among psycholinguists who study the cognitive aspects of Chinese writing.

I am tempted to coin a new term, *semaphoric*, to refer to semantic components, in parallel with *phonophoric*. This would allow the use of a cover term *phoric* to refer to both types of function-bearing components of Chinese characters, in distinction to recurring components that do not have an identifiable function (termed

Can a logographic script be meaningfully simplified and remain logographic?

(also called *phonophoric*). Many, but not all, components that have these functions are individual graphs in their own right—though allographic variation may obscure the graphic relationship to the untrained observer. I will refer to recognized semantic and phonetic components as *functional components*, as distinguished from purely graphical non-functional components.

Figure 2


	yuè / lè (713)		樂	樂	樂	樂	樂	樂
	music/happy		樂	乐				
白	幺	幺	樂					

Figure 2 above is from the same textbook as Figure 1. The top left box indicates that the character 樂 is composed of fifteen strokes; but note that the bottom-row illustration showing how to build up the character is given in terms of components, not strokes. In this case, the components are 幺 (twice), 白, and 木. By the time the student has reached this lesson, she is expected to already recognize these recurring components and be familiar with the normative way of writing each of them. (The simplified form of this character, 乐, is also shown in Figure 2, in the second row.)

The native Chinese tradition of character analysis depends in part on a theory of internal structure and of the functional role of individual character components, as do more sophisticated, linguistically-informed

---

*subcomponents* by Shu et al. (2003:28) and *stroke patterns* by Liu & Hsiao (2012:689) and Hsiao & Shillcock (2006)). However, given the confusion already engendered by the proliferation of terms in the literature, it seems advisable not to introduce any new terminology.

modern scholarly analyses.<sup>15</sup> For the purposes of this paper—i.e. a description and evaluation of 20<sup>th</sup>-century simplifications—a somewhat simplified categorization of character structure and component function will suffice. The following categorization is based on that found in Handel 2009.

In terms of internal structure and function, there are three types of graph: unit graphs, semantic-semantic (or *syssemantic*) compound graphs, and phonetic-semantic (or *phonosemantic*) compound graphs.<sup>16</sup> The latter type is also sometimes referred to as semantic-phonetic compound graphs.

Some unit graphs contain recognizably distinct graphic components, but these components are not functional. Most unit graphs derive from earlier pictographs or other iconic representations. Examples are 日 *rì* ‘sun’ and 馬 *mǎ* ‘horse’.<sup>17</sup> The latter is an example of a character with a recognizable graphic element: the four dots at the bottom of the graph, which recur in a number of other characters such as 魚 *yú* ‘fish’ and 鳥 *niǎo* ‘bird’. However, these four dots are purely graphic, not functional: they do not serve a semantic or phonetic function, and are not graphemic. Unit graphs are not especially numerous, but many of the highest-frequency graphs are of this type. We can see for example from Tables 1-3 in Shu et al. (2003:31-32) that of the 436 Chinese characters introduced to first-grade students in the 1996 *Elementary School Textbooks* prepared by

---

<sup>15</sup> A truly synchronic theory of character structure analysis would seem to be nearly impossible; both the native tradition and modern analyses appeal to an understanding of the origin and development of characters, as well as to diachronic developments in spoken language. See for example Boltz 1994 and Behr 2010.

<sup>16</sup> While this categorization appears to be straightforward, in practice it is complicated by numerous diachronic factors. Characters might shift from one category to the other over time, depending on changes in character form, changes in pronunciation, folk perceptions of script users, systemic realignments, and other factors. It is also possible to recognize some characters as belonging to the semantic-semantic and phonetic-semantic categories simultaneously. For a more sophisticated analysis that takes all these factors into account, see Behr 2010.

<sup>17</sup> The pictographic nature of the characters is obscured in the modern form of the graphs, but is readily evident in early forms from the first millennium BCE.

Can a logographic script be meaningfully simplified and remain logographic?

the Chinese Ministry of Education, 26% are unit graphs.<sup>18</sup> In sixth grade, as lower-frequency characters are introduced, only 4% of the 203 new graphs are unit graphs.

Semantic-semantic compounds are composed of two graphic elements. In many cases these graphic elements can themselves function as unit graphs, or are allographic variants of unit graphs. The meaning of the morpheme written by the compound is associated with, or suggested by, the meanings associated with each component graph and/or their juxtaposition. Examples are 尖 *jiān* ‘sharp’, composed of 小 *xiǎo* ‘small’ atop 大 *dà* ‘large’, and 体 *tǐ* ‘body’, composed of 人 *rén* ‘person’ (in its allographic left-side combining form 亻) and 本 *běn* ‘root’.<sup>19</sup> Graphs of this type are also relatively few in number.

Phonetic-semantic compounds are, like semantic-semantic compounds, also composed of two graphic elements, but in this case one is conventionally associated with a semantic area related to the morpheme represented by the compound character, and the other is conventionally associated with a pronunciation similar to that of the morpheme. Examples are 芳 *fāng* ‘fragrant’, composed of the semantic element 艸 *cǎo* ‘grass’ (in its allographic abbreviated form 艹) above and the phonetic element 方 *fāng* ‘square’ below; and 路 *lù* ‘road’, composed of the semantic element 足 *zú* ‘foot’ on the left and the phonetic element 各 *gè* ‘each’ on the right. In the latter example, the modern pronunciation of the phonetic element (*gè*) bears little relation to the pronunciation of the represented morpheme (*lù*). This discrepancy is the result of sound changes that have taken place over the more than 2,000 years since these characters first came into use. The reconstructed Old Chinese pronunciations are *\*kʰak* (for the morpheme ‘each’ represented by the phonetic element 各 when functioning as an

---

<sup>18</sup> I consider as unit graphs the types labeled by Shu et al. as “pictographs” and “ideographs”.

<sup>19</sup> The character 体 is a long-attested variant of the standard form 體. It was designated an official simplified character in 1956.

independent graph) and \**gar<sup>s</sup>aks* (for the morpheme ‘road’ written by the character 路).<sup>20</sup> Although modern-day script users are unaware of the ancient pronunciations, they still intuit that the component 各 acts as a phonetic element in the character 路 *lù* because of its parallel occurrence in characters that write other *l*-initial morphemes such as 落 *luò* ‘fall’ and 駱 *luò* ‘camel’, as well as its positioning to the right of a familiar left-side semantic element.

Phonetic-semantic characters make up the vast majority of graphs in the writing system, upwards of 80%. (The exact figure depends on the set of graphs and on the counting method, including the degree to which character frequencies are considered and weighted; see e.g. DeFrancis 1984:84, 96 and Shu et al. 2003:41.)

### 3. The history of Chinese script simplification

Fully functional phonographic—i.e. alphabetic or near-alphabetic—representations of Chinese have existed for many centuries. Even setting aside *ad hoc* transcriptional representations of Chinese pronunciations in such writing systems as Tibetan, which date back to the first millennium CE, alphabetic scripts employed to systematically representing Chinese pronunciation are many hundreds of years old. To my knowledge, the first state-sanctioned official transcription of Chinese was the 'Phags-pa alphabet, developed around 1270 during the Yuán dynasty (1279-1368).<sup>21</sup> With increasing numbers of European missionaries, travelers, and

---

<sup>20</sup> Old Chinese reconstructions are adapted from the system of Baxter and Sagart, version 1.00 of February 20, 2011, accessed from <http://crlao.ehess.fr/document.php?id=1217>.

<sup>21</sup> The 'Phags-pa alphabet was used in a number of variants to transcribe different languages. Its use for Chinese was codified in the Chinese character syllabary *Měnggǔ Zìyùn* 蒙古字韻 of 1308. For a general introduction to 'Phags-pa and the associated Chinese textual corpus, see Coblin 2007.

Can a logographic script be meaningfully simplified and remain logographic?

statesmen sojourning in China through the 17<sup>th</sup>, 18<sup>th</sup>, and 19<sup>th</sup> centuries, a variety of transcriptional systems for different varieties of Chinese language were invented, and by the late 19<sup>th</sup> and early 20<sup>th</sup> centuries a number of standard transcriptions had been developed and codified in dictionaries.<sup>22</sup> It was also around this time that a number of influential Chinese intellectuals began seriously advocating for the abolition of Chinese characters and their replacement with an alphabetic writing system based on the Roman alphabet. In the early 20<sup>th</sup> century several new Romanization schemes were developed. One of them, known as Latinized New Writing (Lādinghuà Xīn Wénzì 拉丁化新文字, aka “Sin Wenz”) was closely associated with the Communist Party in the 1930s and 1940s.<sup>23</sup>

After the success of the Communist revolution in 1949, it was widely assumed that the government of the newly established People’s Republic of China [PRC] would lay the groundwork for eliminating Chinese characters in favor of an alphabetic writing system such as Sin Wenz, in support of the party’s goal to foster mass literacy. However, as described by Ramsey (1987:143-145), the new government quickly backtracked on this policy. In 1950 Chairman Mao Zedong, while affirming the ultimate goal of alphabetization, made it clear that writing reform in China would begin with the simplification, rather than the abolition, of characters. In the meantime a new system of alphabetic *transcription* for Modern Standard Chinese would be developed. By 1958 the government had clarified that writing reform would continue to concentrate on character simplification; that the newly devised alphabetic transcription, *Hànyǔ Pīnyīn* (漢語拼音, hereafter “Pinyin”) would function not as an orthography but primarily as

---

<sup>22</sup> The best-known example today is the Wade-Giles transcription, which was the standard in Western academia through much of the 20<sup>th</sup> century.

<sup>23</sup> The full name of the system, as represented in that system, is Latinxua Sin Wenz. For a brief explanation of this system and its history, see Norman 1989:260-263.

a pronunciation guide for characters<sup>24</sup>; and that resolution of the question of the ultimate fate of Chinese characters would be postponed indefinitely.

Variation in the written form of individual graphs in the Chinese writing system has existed throughout the history of writing in China—as indeed it exists in all writing systems. Over the last century, as increasing numbers of excavated texts from the first 1500 years of Chinese history have come to light, the nature and scope of early variation in Chinese writing has received a great deal of attention and analysis. (These texts include epigraphic material on bone, shell, bronze, and stone, as well as manuscripts written in ink on bamboo, silk, wood, and other materials.<sup>25</sup>) Orthographic variation has included forms that could be characterized as simplified, which are the natural result of changes in writing technology, the tendency toward speed and efficiency in writing, and aesthetic considerations (especially as related to calligraphic practice), among other factors.

Over this same time period, there has also always been a strong tendency toward official or elite identification of one of the variants—often, but not always, the most structurally or visually complex—as the standard form (*zhèngzì* 正字) suitable for formal or official writing and printing.

---

<sup>24</sup> In a speech delivered on January 10, 1958, Premier Zhou Enlai specified that language reform would have three components: character simplification, promulgation of the new national standard language, and development of a phonetic transcription. He stated unequivocally that the transcription “is to annotate the characters phonetically and to popularize the common speech. It is not to replace the Chinese characters” (Zhou 1958:17). The Resolution of the State Council on the Promulgation of the Draft Scheme for a Chinese Phonetic Alphabet, adopted by the plenary session of the State Council of the PRC on November 1, 1957, stated that the alphabet was “for the purpose of annotating the Chinese characters so as to make Chinese easier to learn and help unify pronunciation” and would “act as an impetus to [the] work of improving the teaching and learning of the Chinese language in schools, popularizing the common speech and wiping out illiteracy” (Anonymous 1958:61-62).

<sup>25</sup> The relevant scholarship on this subject is too vast to summarize here. The interested reader may consult Boltz 1994 and Qiu Xigui 2000 for general introductory overviews in English.



Can a logographic script be meaningfully simplified and remain logographic?

While recognition of the standard forms was to a large degree the product of tacit agreement among elite users of the script, those standards were not infrequently codified in state-sponsored dictionaries, where they were explicitly contrasted with variant forms (*yìtǐzì* 異體字) labeled as ‘vulgar’ or ‘popular’ (*sú* 俗), ‘ancient’ (*gǔ* 古), ‘odd’ (*guài* 怪), ‘wrong’ (*é* 訛), and so on.

These circumstances—protean variation overlaid by a stable veneer of officially sanctioned regularity—undergird the astonishing conservatism of the official Chinese character script from the 3<sup>rd</sup> century through the 1950s.<sup>26</sup> Any literate Chinese person living in the first half of the 20<sup>th</sup> century would have had no difficulty identifying characters in a formal text written 1500 years earlier. Indeed, one would be hard pressed to identify any formal orthographic distinctions at all across that 1500-year span.

At the end of the 19<sup>th</sup> century, when the modern movement for script reform began, the existence of simpler variant forms of characters was not an unfamiliar notion. Simpler variants were a regular part of daily life, just as they had been for millennia. They were widely used in calligraphic practice, casual handwriting, and informal situations. What was new was the idea that any of these non-standard forms might be employed in formal settings and be fully sanctioned by both the state and the literary elites. This is the radical idea that advocates of simplification were proposing.

---

<sup>26</sup> The detailed picture is not as clear-cut. Within the overall pattern of continuity and stability were periods of disruption and change, generally followed by earnest efforts to re-standardize the written language. A prominent and influential example of the latter is the dictionary *Gān Lù Zìshū* 干祿字書 of the 7<sup>th</sup> century by Yán Yuánsūn 顏元孫 (?-714). This work gathered together variant forms of a character and classified them, contrasting standard character forms (*zhèng* 正) with vulgar (*sú* 俗) and common (*tōng* 通) forms. Vulgar and common forms were distinguished in that the former were recently created, while the latter had been in circulation for some time. For example, the three variant forms 斷 斷 斷 (for *duàn* ‘break’) were labeled vulgar, popular, and standard, respectively.

Various plans for formal simplification of Chinese characters were put forth by individuals and groups through the early part of the 20<sup>th</sup> century.<sup>27</sup> The Ministry of Education of the Nationalist government of China developed a simplification program in the mid-1930s, which was never adopted.<sup>28</sup> Despite intense efforts by its advocates through the 1930s and 1940s, formal promulgation of government-sanctioned simplified characters, widely adopted and uniformly implemented by publishers and the education system, was not to occur until several years after the 1949 Communist revolution.<sup>29</sup>

In 1956, the government's Committee on Script Reform (Wénzì Gǎigé Wěiyuánhui 文字改革委員會) completed work on an initial set of character simplifications, which was published under the title *Hànzì Jiǎnhuà Fāng'àn* 漢字簡化方案 (*Scheme for Simplification of Chinese Characters*) by the State Council; it contained only 515 simplified forms (Chén 1956, Norman 1988:80, 256).<sup>30</sup> A revision and extension of these simplifications was undertaken over the next eight years, culminating in the 1964 release of a

---

<sup>27</sup> See, for example, Barnes 1988-1989 on Chén Guāngyáo, who spent several decades advocating for his own system of character simplification. Seybolt and Chiang 1979 contains as an appendix (pp. 385-390) a useful timeline of the major events in language reform from 1913 to 1978.

<sup>28</sup> In 1934 the Ministry released a list of 324 simplified characters for compulsory use (Chang 1976), but they were withdrawn in the face of opposition. Proposals were also made to the Ministry for more comprehensive simplifications involving thousands of characters (Barnes 1988-1989:156-157).

<sup>29</sup> For a comprehensive study of language reform planning in the PRC during the years following the revolution, see Milsky 1974.

<sup>30</sup> These 515 simplified characters replaced 544 traditional characters by eliminating (through consolidation with other characters) 29 characters. The 515 characters were presented in two lists, one of 230 characters to be implemented immediately, and one of 285 to be circulated for testing and evaluation. In addition, a separate list of simplified “radicals” (recurring character elements), whose implementation would potentially affect thousands of characters, was also included for consideration. Some characters in the latter two categories were put into effect on a trial basis. Thus the practical effect of the 1956 publication was the widespread use of notably fewer than 515 simplified forms. In January 1958 that number seems to have been 355 (Zhou 1956:8).

Can a logographic script be meaningfully simplified and remain logographic?

new list titled *Jiǎnhuàzì Zǒngbiǎo* 简化字總表 (*Comprehensive List of Simplified Characters*). This revised list contained 2,238 simplified character forms, and has remained the basis for the official writing system of the People's Republic of China up to the present day.<sup>31</sup> The 1964 simplification is known today as the first round of character simplification. Already at the time of its publication simplification was viewed as an ongoing, open-ended process. Wu Yu-chang (1958:36), Director of the Committee for Reforming the Chinese Written Language, reflected this view when, surveying the results two years after the first set of simplified characters were promulgated, said “the work of simplifying Chinese characters must be actively continued and pushed ahead so that all those characters which are relatively complex in structure but are rather commonly used can be gradually simplified.” It seems to have been widely assumed that the 1964 list was merely a first step that would be followed by additional rounds of simplification.

In 1977 the government proposed a new set of simplified forms, known as the second round of Chinese character simplification. Because of controversy surrounding this proposal, it was never fully adopted in practice, and in 1986 it was officially rescinded by the government.<sup>32</sup> The finalized official set of simplified characters dating from 1986—nearly identical to the 1964 set—can be easily found today as an appendix in most PRC-published dictionaries of the standard language.

---

<sup>31</sup> This 1964 list was essentially a full implementation of the 1956 proposals. To modern sensibilities, PRC publications in the intervening years 1956-1964 look quite odd, in that they appear to be a hybrid of today's simplified and traditional Chinese scripts. (Texts published in that period are particularly problematic for some OCR (Optical Character Recognition) engines, as I learned when attempting to create a text version of a scanned copy of Chen 1956. Choosing either “Chinese (Simplified)” or “Chinese (Traditional)” as my language option both yielded alarmingly poor results.)

<sup>32</sup> For more on the second round proposal, see Zhao and Baldauf 2008:51ff. We will look at the nature of these rejected proposals in more detail in section 8 below.

While in theory character simplification and writing reform remain ongoing processes in the PRC, in practice there have been no significant proposals since the late 1980s, and there appears to be little activity or interest in this area on the part of the Chinese government today. It is reasonable, therefore, for any analysis of modern-day writing reform in the PRC to treat the current state of reforms as complete.<sup>33</sup>

It is important to note that character simplification did not take place in either Taiwan or Hong Kong, which have continued to this day to use the long-established standard forms (termed “traditional”, “complex”, or “long-form” characters). Script use in these two regions can therefore serve as a useful control against which to measure the effects of the mainland Chinese simplification experiment. For an example of the two script types in use in running text, see appendix 1.

#### **4. Simplification methods**

As Barnes (1988-1989:147) notes, it had been long assumed by advocates of writing reform in the early 20<sup>th</sup> century that simplified characters would be drawn from three primary sources: the informal “vulgar” or “popular” character forms found in ancient dictionaries and/or still in use among the population at large; the cursivized forms found in calligraphy and handwriting; and obsolete character forms attested in earlier eras. (Obviously any such forms would only be useful if they were deemed simpler in structure than the standard forms; this was true of all cursive forms and of most vulgar forms, but not necessarily of ancient obsolete

---

<sup>33</sup> Media reports indicate that a proposal was made in 2009 to simplify 44 graphs, but that the proposal was rejected. See <http://www.telegraph.co.uk/news/worldnews/asia/china/6397611/Plans-to-simplify-Chinese-characters-provoke-anger.html> for a report in English and <http://yuweiban.scxxt.com.cn/ViewInfo.asp?id=220> for a report in Chinese with the full list of 44 characters (both accessed September 21, 2012).

Can a logographic script be meaningfully simplified and remain logographic?

forms.) The reason to rely on such historically attested “simplified” characters was that their familiarity and connection to Chinese culture would presumably make them more acceptable to script users and thus ensure a less disruptive transition. Moreover, because many vulgar forms had been disparaged by past elites but remained in use among the people at large, advocacy for their adoption fit well into the ruling Communist Party’s ideological narrative.

Two additional factors were important in the simplification process. Mao had made it clear that he wanted character simplification to include a reduction in the overall number of basic characters that had to be learned (Chang 1976:187). This meant that it was necessary not just to simplify individual graphs, but in some cases to consolidate one or more distinct graphs into a single simplified form. Second, it was clear that the three primary sources mentioned above would not be able to yield a sufficient number of simplified characters. If simplification were to occur on a large enough scale to be of benefit, it would be necessary to create some newly simplified characters that lacked any historical precedent.

The examples below illustrate the most common techniques of simplification.<sup>34</sup>

1. Replacement with a simpler graph writing a homophonous morpheme

(4) 裡 lǐ ‘inside’ ® 里 (cf. 里 lǐ ‘mile’)

(5) 穀 gǔ ‘grain’ ® 谷 (cf. 谷 gǔ ‘valley’)

(6) 係 xì ‘relation’, xì 繫 ‘tie’ ® 系 (cf. 系 xì ‘system’)

(7) 雲 yún ‘cloud’ ® 云 (cf. 云 yún ‘say’)

---

<sup>34</sup> This list of techniques has been adapted and revised from the categorizations given by Chen (1999:150-157), Ramsey (1987:147-150), Seybolt and Chiang (1979:11-12), and Yi (1956). When listing examples, I use the formulation “x ® y” to indicate that the traditional character x has been replaced with simplified character y.

The result of such simplifications is a consolidation of homophonous graphs and a reduction in the overall number of graphs in the system. There is of course an accompanying increase in polyvalency, since one graph ends up representing two or more morphemes. Generally, this method was only used when one of the graphs involved writes an obsolete morpheme or a bound morpheme occurring in limited contexts. For example, *yún* ‘to say’ is a common verb in the classical written language, but is not part of the modern spoken language. Moreover, it is a distinct part of speech from *yún* ‘cloud’. The possibility of ambiguity in the use of the simplified polyvalent graph 云 is therefore deemed to be minimal.

2. Adoption of vulgar forms commonly used for centuries

(8) 頭 ® 头 *tóu* ‘head’

(9) 個 ® 个 *gè* ‘[classifier]’

3. Adoption of established calligraphic or cursive forms

(10) 馬 ® 马 *mǎ* ‘horse’

(11) 見 ® 见 *jiàn* ‘see’

These cursive forms developed through combining and reducing strokes. The ductus of the new simplified forms was modified from their handwritten sources to give them the same ‘look and feel’ as printed forms. This was achieved by straightening and angularizing the individual strokes from their flowing, curved forms, and modifying the overall envelope of the character to match the proportions of standard forms.

4. Use of archaic forms employed at various times in history

(12) 從 ® 从 *cóng* ‘follow’

(13) 陽 ® 阳 *yáng* ‘yang principle’

Can a logographic script be meaningfully simplified and remain logographic?

Some of forms in this category are obscure, but all are attested. 从 *cóng* is the regular form found in the Han dynasty text *Shuōwén Jiězì* 說文解字. 阳 *yáng* is found in a number of texts from the Yuan, Ming, and Qing dynasties (Liu and Li 1930:101).

#### 5. Newly invented simplifications

For the most part, newly invented simplifications are based on the same processes that led to historically attested variants as reflected in categories 2 through 4 above. They are, in other words, analogical extensions of attested variation patterns.

##### 5a. Use of part for whole (i.e. elimination of one or more components)

(14) 習 ® 习 *xí* ‘practice’

(15) 醫 ® 医 *yī* ‘medical’

##### 5b. Replacement of phonetic element

(16) 讓 ® 让 *ràng* ‘let, make’

(17) 認 ® 认 *rèn* ‘recognize’

In both of the examples above, the right-side phonetic element has been replaced with a phonetic element with fewer strokes. The new phonetic element for *ràng* ‘let, make’ is 上 *shàng* ‘above’; the new phonetic element for *rèn* ‘recognize’ is 人 *rén* ‘person’. (In addition, the left-side component in both cases has been simplified according to technique 3, from 言 to 讠.)

##### 5c. Reduction of complex components to simple shapes (most commonly 又)

(18) 樹 ® 树 *shù* ‘tree’

(19) 難 ® 难 *nán* ‘difficult’

(20) 鄧 ® 邓 ‘Dèng’ [surname]

(21) 風 ® 风 *fēng* ‘wind’

There is, obviously, significant overlap in these simplification methods; to some degree my placement of particular examples in only one category is arbitrary. For example, “裡 *lǐ* ‘inside’ ® 里 (cf. 里 *lǐ* ‘mile’)” is categorized above as homophone replacement. But it is also an example of use of a part for the whole (in this case, that part was the phonetic component), as well as an example of the formal adoption of a common vulgar practice. One reason that this overlap of categories occurs is that these simplification techniques are simply recapitulations of natural processes that have been part of the ongoing development of the Chinese writing system from its inception, and therefore are the source of many historical variant forms. Chen 1956, an explanation of the principles underlying the simplification of all 515 forms in the first simplification scheme of 1956, makes clear that reformers were quite conscious of these overlapping mechanisms.

#### 6. Analogous extension of simplified components

Recurring components that are simplified in one character (through reduction, replacement, or elimination) can be simplified in the same fashion in other characters containing that component. It is this kind of analogous extension, especially when it comes to the treatment of the semantic components (or “radicals”), that enabled the increase in the number of simplified characters from the hundreds to the thousands in 1964. For example, in the dozens of characters that contain the component 馬 *mǎ* ‘horse’ either as a phonetic component or as a semantic component, that component is replaced with 马, as in 媽 ® 妈 *mā* ‘mother’. Two more examples of analogous extension are given below. The first is analogous extension of a type 5c simplification (replacement with 又):

(22) 難 ® 难 *nán* ‘difficult’

(23) 漢 ® 汉 *hàn* ‘Chinese ethnicity’

(24) 灘 ® 滩 *tān* ‘beach’



Can a logographic script be meaningfully simplified and remain logographic?

The second is an analogous extension of a type 3 simplification:

(25) 語 ® 语 *yǔ* ‘spoken language’

(26) 讓 ® 让 *ràng* ‘let, make’

(27) 認 ® 认 *rèn* ‘recognize’

The categorization given above is not exhaustive, but accounts for the vast majority of simplified forms.

## 5. Analysis of simplification

Although the question may seem naïve, it is worth asking what exactly makes these simplified character forms simpler. The answer to this question is not as straightforward as might be initially assumed, and will be the focus of this section of the discussion.

In their discussion of character simplification, early 20<sup>th</sup>-century reform advocates focused on two basic aims: simplifying a large number of characters (especially those of high frequency), and reducing the number of strokes in each character. It was widely, and unquestioningly, assumed that this was the best way to achieve the ultimate goal of simplification: increased literacy. After all, the reasoning went, reducing the number of strokes would make characters easier to learn and remember; the more characters simplified in this way, the greater the impact would be. Once characters became easier to learn and remember, more people could achieve literacy with fewer years of formal education, a key concern for a poor, resource-strapped nation desperate to modernize. Some advocates of reform were guided by arbitrary stroke reduction targets. For example,

Qian Xuantong in 1920 proposed that only characters having more than 10 strokes should be targeted for simplification (Barnes 1988-89:145).

Looking at the examples given above in section 4, it is easy to see that in all cases the simplified characters have fewer strokes than their traditional counterparts, in some cases greatly so. For example, 讓 ® 让 *ràng* ‘let, make’ is a change from 24 strokes to 5, a reduction by 19 strokes. Following the 1956 simplification, reformers touted the reduced number of strokes in characters as a self-evidently significant achievement, and drew a direct causal link between stroke number and ease of learning. The 515 simplified forms from 1956 had an average of 8.16 strokes per character, while the 544 characters that they replaced had an average of 16.08 strokes per character; these statistics were widely cited (Chen 1956:54, Wu 1958).<sup>35</sup> Wu Yu-chang (1958:34) states: “The popularization of simplified characters greatly facilitates children’s education, elimination of illiteracy, and writing in general”. Chen (1956:54) concludes by claiming that the use of simplified instead of traditional characters will make the task of eliminating illiteracy “much easier” (*shěng shì hěn duō* 省事很多). But only anecdotal evidence (generally in the form of poorly educated Chinese speakers praising the simpler forms<sup>36</sup>) is provided as justification for these claims. Wu, as part of a criticism of “rightists” who oppose writing reform, opines “we must say that the simplification of Chinese characters has definitely benefited hundreds of millions of children and illiterates, and that it is a success, not a failure” (1958:34). Such a sweeping claim seems excessive on the face of it, given the small number of simplified characters that had

---

<sup>35</sup> A more recent analysis summarizes the results of simplification this way: “among the most frequently used 3,500 characters, around 40% were simplified, which have approximately 22.5% fewer strokes than the traditional counterparts” (Liu & Hsiao 2012:689, citing Gao & Kao, 2002).

<sup>36</sup> See, for example, Zhou 1958:8-9 on the enthusiasm that students and peasants expressed for the simpler forms.

Can a logographic script be meaningfully simplified and remain logographic?

been put into general use at the time. In section 9 below, we will investigate these claims anew from the perspective of recent psycholinguistic studies.

Another metric by which reformers judged simplification proposals was reduction in the overall number of characters to be learned. Barnes (1988-1989:146) reports that a number of surveys carried out in the 1920s revealed that no more than 5,500 characters were in use in contemporary reading material. If that number could be brought down significantly, it would reduce barriers to education and literacy. But the simplifications undertaken in 1956 and 1964 did very little to reduce the overall number of characters. Chang (1976:189) calculates that the number of characters reduced was 294, and concludes that “Mao [Zedong]’s request that the number of characters be greatly reduced does not seem to have borne much fruit”.

The reason that the number of characters ultimately eliminated from the system is so low probably has to do with the fundamental logographic nature of the script. Once the decision is taken not to replace the logographic script with a syllabary, extensive character elimination would naturally be viewed as an invitation to chaos, since it would eliminate the idealized one-to-one ratio of graph to morpheme and lead to increasing ambiguity in the writing system, at least in theory.<sup>37</sup> Wu (1958:34-35) notes that there was already some dissatisfaction with this aspect of the 1956 simplification scheme, a concern that “the substitution of some characters by the simpler form of their homophones ... [is] either rather inappropriate in use or likely to cause ambiguity”. Wu then offers an

---

<sup>37</sup> Both the natural tendency toward syllabary-like use of Chinese characters, and the normative counter-tendency of establishing a one-to-one correspondence between graph and morpheme, have been hallmarks of the history of Chinese writing for millennia. See Boltz 1994 and Handel 2009. Given these historical tendencies, it would not be surprising if Chinese elites over-estimated the threat of ambiguity and under-estimated the role of context in practical disambiguation of polyvalent graphs.

example involving the following two characters, the first of which has been simplified to be identical to the second:

(28) 隻 *zhī* ‘[classifier]’ ® 只

(29) 只 *zhǐ* ‘only’

Wu points out that the use of the character 只 to write the classifier *zhī* was already common practice in informal situations. For this reason, and no doubt also because the parts of speech are distinct, the 1956 simplification scheme formally adopted this replacement. However, Wu points out that it is possible to devise written sentences that are ambiguous when the simplified forms are used, such as:

(30) 許多船只通过苏伊士运河

The two possible readings are:

(31) 許多 船只 通过 苏伊士 运河。  
Xǔduō chuánzhǐ<sup>38</sup> tōngguò Sūyīshì yùnhé  
many vessels pass Suez canal  
Many vessels passed through the Suez Canal.

(32) 許多 船 只 通过 苏伊士 运河。  
Xǔduō chuán zhǐ tōngguò Sūyīshì yùnhé  
many boats only pass Suez canal  
Many boats only passed through the Suez Canal.

---

<sup>38</sup> This word is derived according to a morphological pattern in which a noun and its classifier form a compound; the result is a collective noun with generalized semantics. For example, *zhǐ* ‘paper’ compounded with *zhāng* ‘[classifier for sheets of paper]’ yields *zhǐzhāng* ‘paper products, stationery’; *chē* ‘automobile’ compounded with *liàng* ‘[classifier for automobiles]’ yields *chēliàng* ‘vehicles’. In certain contexts *chuánzhǐ* might be rendered in English as ‘fleet (of ships)’. See *A grammar of spoken Chinese* by Yuen-Ren Chao, p. 396 (Berkeley and Los Angeles: University of California Press, 1968).

Can a logographic script be meaningfully simplified and remain logographic?

In speech, the two sentences are easily distinguished by intonation and phrasal grouping, not to mention the difference in tone between the two morphemes *zhī* and *zhǐ*, but if lacking context the two sentences cannot be distinguished in written form. If the writing system employed lexical spacing, or if the original form of the character for the classifier *zhī* were retained, the sentences would be distinguishable in writing. (Despite Wu's concern, which was clearly of more theoretical than practical import, the simplification of 隻 to 只 was retained in the 1964 list and is part of the standard orthography today.)

It is fair to say that reduction of the overall number of Chinese characters has not had a significant impact on the writing system as a whole, and has almost certainly had a negligible effect on literacy rates. This is because the number of characters that have been eliminated in the transition to the simplified character script is so minimal. Norman (1988:73) cites studies indicating that educated Chinese without particular academic expertise probably know between 3,500 and 4,000 characters. Next to this figure, a reduction of about 300 cannot be considered significant. In evaluating the efficacy of the simplified character reform, therefore, we should turn our attention to the effects of stroke reduction.

As noted earlier, there is no doubt that the simplified characters have, on average, significantly fewer strokes than their traditional counterparts. But more linguistically sophisticated observers of the Chinese script have long recognized that this metric can be misleading. A concern that has frequently been raised is that stroke reduction can lead to difficulty in distinguishing graphs.<sup>39</sup> This can be readily illustrated with a few concrete examples:

---

<sup>39</sup> For example, Chen (1999:162) says: "When a reduction in the number of strokes makes characters easier to write, it may also make them less differentiated from each other, and thus less easy to recognize."

(33) 龍 ® 龙 *lóng* 'dragon'

(34) 發 ® 发 *fā* 'emit'

(35) 環 ® 环 *huán* 'loop'

(36) 壞 ® 坏 *huài* 'bad'

The concern is that as the features that distinguish pairs of graphemes become fewer and smaller, the reader will need to expend more cognitive effort, making use of context and other cues, to identify characters and the morphemes that they represent, with the result that reading becomes slower and more laborious, especially at smaller font sizes. While this is certainly a legitimate theoretical concern, I am not aware of any experimental evidence suggesting that users of the simplified character script read more slowly than users of the traditional script. However, there is persuasive evidence that it has affected cognitive processes involved in reading that are measurable in both student learners and adult users; this evidence will be discussed below in section 9.

The larger problem with using stroke number as the metric by which to judge the efficacy of simplification is that it is based on a fundamental misjudgment about Chinese characters, one that should already be apparent given the analysis of section 3 above: namely, that the stroke is the basic cognitive unit by which script users learn and remember characters. This misjudgment is in turn predicated on a view of Chinese characters that fails to take into account their systemic embedding, i.e. the patterns that are observed across characters within the writing system considered as a whole.


To anyone unfamiliar with the Chinese script, the complexity of a character seems to be intimately associated with the number of strokes, which certainly contributes to its visual density. For example, consider these two traditional characters:


Can a logographic script be meaningfully simplified and remain logographic?

(37) 龕 *dá* ‘appearance of flying dragons’ (48 strokes)<sup>40</sup>

(38) 黽 *mǐn* ‘frog’ (13 strokes)

By enlarging the characters we can see the individual strokes more clearly:

 *dá* ‘appearance of flying dragons’ (48 strokes)

 *mǐn* ‘frog’ (13 strokes)

The top character is denser and appears to have greater internal complexity. But to anyone who has even a minimal working knowledge of the script and language, the impression of complexity shifts markedly. Beginners memorize the first characters they learn as arbitrary collections of individual strokes. But once a small inventory of characters has been mastered, the learner begins, consciously or unconsciously, to identify recurring components. Because of the connection to morphemes, these components are recognized over time as having semantic or phonetic properties. Because many components are characters in their own right, they also have pronunciations which can serve as labels. The salience of these forms is further reinforced by lexicographic and other cultural conventions. As a result, after a few dozen characters have been mastered, newly encountered characters start to be learned not as collections of strokes, but as structured groupings of higher-order elements. This is also how characters are conceptualized and described by script users, and thus how their structures are communicated to learners.

---

<sup>40</sup> Also pronounced *tà*.

As an illustration, suppose that one script user wished to explain to another how to write an unusual character with a specialized usage: 嶗 Láo, the name of a mountain not far from the city of Qingdao in Shandong province. Were she to give the following instructions it could only be viewed as some sort of joke: “Start on the left: make a vertical stroke, then make a second vertical stroke to its left that hooks around into a horizontal stroke passing underneath and touching the first vertical stroke, then make another vertical stroke to its right ....”<sup>41</sup>

The normal and natural way to explain how to write the character does not attempt to describe the location, shape, and direction of fifteen ordered strokes, but simply names two components: “*shān* ‘mountain’ on the left, and *láo* ‘labor’ on the right”. And indeed, it is in terms of these two components that the character would be interpreted by any script user encountering it for the first time. The conscious or unconscious analysis that a script user would make runs something along these lines: “Ah, I see. The name of mountain Láo is written with *láo* ‘labor’ to match its pronunciation and *shān* ‘mountain’ to categorize its meaning.” Since both of those components are givens in the system—they are already familiar and automatic to the script user—the character is readily viewed as composed of two elements, not fifteen. Moreover, because each of those elements is logically connected to the linguistic unit represented by the character—i.e., because the character’s structure is well motivated—memorization and recall is even easier than it would be if the character

---

<sup>41</sup> This sentence describes the writing of the element 山 that appears on the left side of the character 嶗. It is rather remarkable that this intuitively obvious fact about character strokes was not part of the standard psycholinguistic literature until fairly recently. Hsiao & Shillcock (2006:407) observe “In early attempts to model Chinese language processing, researchers usually used strokes to encode orthographic representations [studies from 1999 and 2002 are cited]. In recent years, studies have shown that recognition by skilled readers is based upon well-defined orthographic constituents, i.e., single bodies, which are integral stroke patterns that cannot be further decomposed into other units, instead of individual strokes as previously thought ....” See section 9 for an elaboration of this point.



Can a logographic script be meaningfully simplified and remain logographic?

were made up of two completely arbitrary (but common) character components.<sup>42</sup>

Returning to the two example characters (37) 龘 and (38) 𪚩 with this in mind, and knowing that the high-frequency character 龍 *lóng* ‘dragon’ is already familiar to script users, we can now see that the first character is far simpler than the second. It is simply composed of three dragons.<sup>43</sup> There are few characters that are easier to learn within the systemic context of the script. The second character, by way of contrast, does not break down into any easily recognizable components; it must be learned stroke by stroke, and should be considered more complex.

Let us now consider the simplified forms of some of the characters we have been discussing.

(39) 嶗 ® 嶗 Láo [name of mountain in Shandong province]

---

<sup>42</sup> This intuitive analysis is confirmed by psycholinguistic studies. Liu & Hsiao (2012:689) note: “...expertise in recognizing Chinese characters is marked by reduced holistic processing (Hsiao & Cottrell, 2009). This effect may be due to expert Chinese readers’ knowledge about Chinese orthography. Chinese characters are composed of strokes, which combine to form over a thousand different stroke patterns ... and stroke patterns are the smallest functional units in Chinese character recognition (Chen, Allport, & Marshall, 1996). For expert Chinese readers, when recognizing Chinese characters, they may be more sensitive to the internal constituent components... compared with novices .... Consequently, expert readers may process Chinese characters less holistically than novices.” Chen, Allport & Marshall (1996:1038) are more explicit: “The performance of skilled Chinese readers, in simultaneous ‘same’—‘different’ judgements on pairs of Chinese characters, is affected, on the ‘same’ trials, by the number of orthographic units [i.e. character components] and, on ‘different’ trials, by the number of mismatching units .... the demonstration of these two effects suggests that these orthographic units are (implicitly) recognized by skilled Chinese readers and play a functional role in the visual processing of composite Chinese characters .... In contrast, the number of individual strokes had no influence upon the speed or accuracy of performance, independent of the number of stroke patterns.” Moreover, other studies show that Chinese children gradually acquire a more analytic apprehension of characters as they learn to read and write. This is discussed in section 9.

<sup>43</sup> Moreover, the arrangement is the same as for all characters composed of three repeating elements: one on the top, two on the bottom.

(40) cf. 勞 ® 劳 *láo* ‘labor’

While the simplified form of 勞 *Láo* has five fewer strokes than the traditional form 勞, given that the high-frequency character for *láo* ‘labor’ is already known to the literate script user, then one can reasonably suppose that *Láo* is equally easy to learn whether one is using the traditional or the simplified script.

(41) 龕 ≠ 龕 *dá* ‘appearance of flying dragons’<sup>44</sup>

(42) cf. 龍 ® 龙 *lóng* ‘dragon’

In contrast, the character 龕 for *dá*, so easy for a user of the traditional script to learn, presents a challenge for the simplified script user. It bears no relationship to the well-known character 龙 *lóng* ‘dragon’.

As noted by Ramsey (1987:152), this kind of inconsistency is rampant in the simplified character script. In the traditional script, learning the character 龍 for ‘dragon’ with its 16 strokes is somewhat laborious. It is one of a relatively small number of characters (fewer than 20%) that does not decompose into easily recognizable components. But once learned, characters that contain it as a component, such as 龕, are easy to acquire. In contrast, the user of the simplified script has presumably had an easier time learning the five-stroke simplified character 龙 for ‘dragon’. But she must still learn, in addition to it, the traditional sixteen-stroke form 龍, since it occurs as a component in a number of other characters (most of which are low-frequency). Thus the overall number of *components* to be learned in order to master the entire system has increased with simplification. And since those characters that contain the component 龍

---

<sup>44</sup> I will use the formula “*x* ≠ *x*” to indicate that the traditional character *x* remains unchanged, i.e. that no distinct simplified character form has been designated.

Can a logographic script be meaningfully simplified and remain logographic?

in the simplified system are of low frequency, it may be more difficult to retain the ability to write and recognize it.

A counter-argument, of course, is that the character 𪛗 is archaic and exceedingly rare; it does not write a morpheme that is part of the modern spoken language, and is likely only to be encountered in ancient poetry or obscure dictionaries. Thus the difficulty of acquiring this character is largely irrelevant to general literacy, and has no direct bearing on any analysis of simplified character system. But that counter-argument does not apply when we look beyond this particular example. The same situation is seen with the most common components and characters in the simplified system.

Let us now look a bit more systematically at some of the factors that arguably increase complexity of individual characters and of the system as a whole, even as they reduce (or leave unchanged) the number of strokes in every character considered individually. The cognitive effects of these factors will be discussed afterwards, in section 9.

## 6. Inconsistencies of 20th-century simplification

A number of typical character simplifications are listed below, with discussion of the ways in which they have led to increased systemic complexity.

### 1. Incomplete replacement of phonetic elements

(43) 讓 ® 让 ràng 'let, make'

(44) cf. 嚷 ≠ 嚷 rǎng 'howl', 瓢 ≠ 瓢 ráng 'pulp', 攘 ≠ 攘 rǎng 'seize', 壤 ≠ 壤 rǎng 'soil', 穰 ≠ 穰 ráng 'grain stalks', 囊 ≠ 囊 náng 'sack', 鑲 ® 镶 xiāng 'inlay', etc.

The high-frequency character 讓 *ràng* ‘let, make’ was simplified, but none of the other characters containing the phonetic element 襄 *xiāng* ‘assist’ were analogously simplified. Most of these characters are low frequency, but several are among the most frequent 3,000 characters, i.e. are necessary for basic literacy, e.g.: 壤 *rǎng* ‘soil’ (#2131), 囊 *náng* ‘sack’ (#2232), 镶 *xiāng* ‘inlay’ (#2625).<sup>45</sup>

In the remaining examples, all the listed characters are high frequency.

## 2. Inconsistent replacement of phonetic elements

(45) 燈 ® 灯 *dēng* ‘lamp’

(46) 鄧 ® 邓 *Dèng* [surname]

(47) cf. 登 *dēng* ‘ascend’

Although both simplified characters have fewer strokes, they are arguably both less systematic (and therefore harder to learn and remember) in simplified form. In the case of 燈 *dēng* ‘lamp’, the homophonous phonetic element 登 *dēng* has been replaced with the semi-homophonous 丁 *dīng* ‘fourth heavenly stem’. In the case of 鄧 ‘Dèng’ [surname], the simplified form 邓 lacks a phonetic element (having been simplified according to

---

<sup>45</sup> Character frequency numbers are from Wenlin software version 3.4 for Macintosh. The software guide at [http://guide.wenlininstitute.org/wiki/Frequency\\_Statistics](http://guide.wenlininstitute.org/wiki/Frequency_Statistics) (accessed July 12, 2011) states that the frequency numbers “were derived by combining and averaging these five sets of statistics”:

- (1) Modern Chinese Frequency Dictionary 《现代汉语频率词典》 Xiàndài Hànyǔ Pínǜ Cídiǎn (Beijing Language Institute, 1986) [Character and word frequencies from a sample of 1,807,389 characters.]
- (2) Which are the Most Commonly Used Chinese Characters? 《“最常用的汉子是哪些?” Zuì Chángyòng de Hànzì shì Nǎxiē? (Chinese Writing Reform Committee and National Standards Office, 1982) [Character frequencies from a sample of 11,080,000 characters.]
- (3) Cracking the Chinese Puzzles (T. K. Ann, © 1982 Stockflows Co., Ltd., 37 Queen’s Road, Central, Hong Kong) [Character frequencies from a sample of 1,408,573 characters.]
- (4) 华夏文摘 Huáxià Wénzhāi (HXWZ) [Sample of 4,189,874 characters.]
- (5) 枫华园 Fēng Huá Yuán (FHY) and 联谊通讯 Liányì Tōngxùn (LYTX) [Combined sample of 1,227,883 characters.]

Can a logographic script be meaningfully simplified and remain logographic?

principle 5c). And, of course, the graphic connection between the two nearly-homophonous morphemes 燈 *dēng* ‘lamp’ and 鄧 ‘Dèng’ has been broken by the inconsistency in simplification methods.

### 3. Inconsistent simplification of identical elements

(48) 賈 ® 贾 *Jiǎ* ‘[surname]’

(49) 價 ® 价 *jià* ‘price’

(50) cf. 貝 ® 贝 *bèi* ‘cowrie shell’ (a common semantic component)

The character 賈 *Jiǎ* serves as phonetic element in 價 *jià* ‘price’. The simplified form of 賈 *Jiǎ* is created by simplification of its 貝 element, analogous to simplifications like 財 ® 财 *cái* ‘wealth’ and 貪 ® 贪 *tān* ‘corrupt’. But 價 is simplified by replacing its phonetic element entirely with a new, less homophonous element 介 *jiè* ‘introduce’.

### 4. Conflation of phonetic elements

(51) 環 ® 环 *huán* ‘loop’

(52) 懷 ® 怀 *huái* ‘long for’

In traditional forms, there are two distinct phonetic elements. The first, 環, occurs in several high-frequency words pronounced *huan*, such as 環 *huán* ‘loop’. The second, 懷, occurs in several high-frequency words pronounced *huai*, such as 懷 *huái* ‘long for’ and 壞 *huài* ‘bad’. These two phonetic elements, graphically similar, have been conflated to simplified element 不 *bù* ‘not’. The element 不 now serves as phonetic in words with pronunciations related to *bu*, *huan*, and *huai*; this increases the complexity of the relationship between phonetic elements and the sounds they correlate with.

### 5. Positional inconsistency of radical simplification

(53) 言 ≠ 言 *yán* ‘speech’

(54) 警 ≠ 警 *jǐng* ‘warn’, 譬 ≠ 譬 *pì* ‘example, analogy’, etc.

(55) 語 ⇒ 语 yǔ ‘spoken language’, 課 ⇒ 课 kè ‘course, class’, 諷 ⇒ 讽 fèng ‘satirize’, etc.

As noted above, the semantic element 言 has been simplified to the calligraphic form 讠 in dozens of common characters that write morphemes with semantics related to the notion of speech, including the three examples 語 yǔ ‘spoken language’, 課 kè ‘course, class’, 諷 fèng ‘satirize’. This semantic element is a character in its own right, writing the morpheme 言 yán ‘speech’. As a character, however, it has not been simplified. Furthermore, as a semantic element, it is not simplified when positioned at the bottom of a character. As a result, what in the traditional system is a single element 言 with consistent semantics has become two distinct elements 言 and 讠, both of which must be learned, and both of which occur with the same functional role; the choice is determined positionally.<sup>46</sup>

A number of very common semantic elements have been simplified analogously to 言: 食 ‘food’ and 金 ‘metal’ remain unsimplified as individual characters writing the morphemes *shí* ‘food’ and *jīn* ‘metal, gold’ respectively, and when positioned at the bottom of a character, but are simplified (食® 饣 and 金® 钅) in dozens of common characters when positioned on the left.

---

<sup>46</sup> It should be noted, however, that positional allography of this type is already a basic feature of the traditional script, so that the situation described here can be viewed as an extension of an existing pattern. Myers (2011) notes that the left and top parts of characters are the typical locations for semantic elements, where they tend to show allographic reduction, while the right and bottom parts are atypical and do not show similar reduction. Compare 忘 *wàng* ‘forget’ with 忙 *máng* ‘busy’. Both have the semantic element 心 *xīn* ‘heart’, which appears in an abbreviated form on the left side. Even traditional elements like 食 and 金 have slightly different forms in left and bottom position in the traditional script, although the difference is not so dramatic as in the simplified script. Compare the appearance of the semantic element on the left in 飯 *fàn* ‘rice, food’ with its appearance on the bottom in 養 *yǎng* ‘raise, rear’.

Can a logographic script be meaningfully simplified and remain logographic?

6. Ambiguous character envelopes

(56) 龍 lóng ‘dragon’, 發 fā ‘emit’ ® 龙, 发

(57) 活 huó ‘living’, 話 huà ‘speech’ ® 活, 话

As noted above, stroke reduction in a number of cases results in two or more high-frequency characters having similar overall shapes and densities, presumably increasing the cognitive burden on differentiation and reliance on context.

## 7. Measuring the success of character simplification

The analysis in the previous section looked at systematicity and consistency, and hypothesized about cognitive burden. Only carefully constructed psycholinguistic and neurolinguistic studies can attempt to definitively answer questions about whether the traditional or simplified writing system is “easier” for the human brain to learn and to use. (What we know so far about this is discussed below in section 9.) But, while of interest in their own right, the answers to those questions may be irrelevant to the practical considerations of language policy and education. Thanks to the example of Taiwan and Hong Kong, we have a real-world control against which to evaluate the goals of the writing reform movement in mainland China.

Since the implementation, in two phases, of simplified characters, literacy rates and education levels in mainland China have soared, rising from about 20% at mid-century to 92.2% in 2008. But in Hong Kong and Taiwan, where traditional characters were never replaced, literacy rates also rose dramatically in the same period. Today’s literacy rate in Taiwan

is 96.1% and in Hong Kong is 93.5%.<sup>47</sup> <sup>48</sup> There is therefore no reason to ascribe the increase in China's literacy to writing reform. As the economic success of Taiwan and Hong Kong have shown, traditional characters and an education system based on a Chinese-character medium are no impediment to the development of a highly skilled, productive, and entrepreneurial workforce. The historical evidence clearly indicates that traditional Chinese writing, while arguably more inefficient and time-consuming to learn than many other writing systems, was not the primary—and perhaps not even any—obstacle to modernization, as reformers assumed in the early 20<sup>th</sup> century.

At the same time, the practical effects of the potential deficiencies of simplified characters, as outlined in the preceding section, are clearly also minimal or non-existent. Mainland China appears to be neither at an advantage or disadvantage to other Chinese-speaking polities as a result of its reformed writing system. At a practical level, one is forced to conclude that official simplification of Chinese writing, as it has been implemented so far, has probably been an enormous waste of time and resources.

## **8. Theoretical simplification - the 1976 proposals and beyond**

From a theoretical perspective, it is worth comparing the current simplified system of characters with a hypothetical simplification based on more consistent application of the principles that have been described

---

<sup>47</sup> Literacy figures are from the CIA World Factbook (<https://www.cia.gov/library/publications/the-world-factbook/>, accessed July 12, 2011). The Hong Kong figure is from 2002, the Taiwan figure from 2003, the PRC figure from 2008.

<sup>48</sup> It is also worth pointing out that most Hong Kong residents are literate in both standard written Chinese and in written Cantonese, and that mastery of the latter requires command of several hundred additional Chinese characters.



Can a logographic script be meaningfully simplified and remain logographic?

above. If number of strokes, character frequency, historical attestation, and aesthetic considerations were devalued in favor of other factors, could a practical set of simplified characters for modern written Chinese be developed that undeniably involves greater systemic simplicity overall? It is not too difficult to imagine what such a system might look like. The phonetic element 襄 (examples 43-44) for example, could be replaced with 上 in all characters in which it appears. The components 言, 食, and 金 could be simplified consistently to 讠 饣 钅 in all characters in which they occur, and as stand-alone characters as well.

In fact, a system of simplified characters developed along these lines is not purely hypothetical. The rejected simplifications of 1976 go some way toward implementing such a system. For example, the 1976 proposal simplified 嚷 *rǎng* ‘howl’, 壤 *rǎng* ‘soil’, and 鑲 *xiāng* ‘inlay’ to forms with 上 on the right side as phonetic. It also proposed to reduce the total number of characters further, for example replacing four graphs pronounced *hú* (葫, 獼, 蝴, 糊) with the single character 胡.<sup>49</sup>

Why were these simplifications rejected? The views reflected in one short essay, Yu 1978, may be taken as indicative of the opposition that the proposals generated. In general terms, Yu complains that the proposals were not developed in consultation with language experts and intellectuals (p. 127)<sup>50</sup> and that, since the number of literate Chinese has greatly increased since the last simplification, a change in the writing system would necessarily be highly disruptive (p. 128). More specifically, Yu argues that the proposed changes lead to graphic and semantic confusion, and would interfere with attempts to promote standard pronunciation. The

---

<sup>49</sup> The first three of these graphs occur in the written form of the bisyllabic morphemes *húlu* ‘gourd’, *húsūn* ‘macaque’, and *húdié* ‘butterfly’, respectively. See section 3 above. The fourth writes a morpheme with a range of related meanings: ‘paste; sticky; vague’.

<sup>50</sup> This is not surprising, considering that the proposals were developed during the Cultural Revolution.

proposal replaces the traditional character 家 *jiā* ‘family’ with (宀 ‘roof’ over 人 ‘person’); Yu says the new character will be too easily confused with the graphically similar character 穴 *xué* ‘cave, hole’ (p. 127). The proposal replaces the character 蔡 ‘Cài’ [surname] with the homophonous existing character 菜 *cài* ‘food, prepared dish’; Yu says that the resulting homography of the words *Xiǎo Cài* (term of address for someone surnamed Cài) and *xiǎocài* ‘snack’ as 小菜 would be confusing (p. 128). The proposal replaces the character 寨 *zhài* ‘stockade, camp’ with (宀 ‘roof’ over 在 *zài* ‘be at’), where 在 *zài* functions as a phonetic element. Because the word *zhài* begins with *zh-* but would be written with a phonetic element beginning with *z-*, and because a lack of distinction between *zh-* [tʃ] and *z-* [ts] is a common non-standard dialectal feature, Yu argues that this will impede acquisition of standard pronunciation (p. 129).

If the 1976 proposals had been adopted, and if the simplification process had continued beyond that, resulting in something close to a writing system containing characters with fewer strokes, but equally as or more systematic and consistent than the traditional system, what would have been accomplished? Put another way, would the benefits of such a system be worth the disruption of changing the writing system every few years, with the attendant social and financial cost?

From a purely practical standpoint, the answer would seem to be no. As noted in the previous section, the practical effects of simplification seem to have been minimal. Even if some effects exist in small measure, and would be magnified (in the case of positive effects) or minimized (in the case of negative effects) by a more systematic simplification, it is difficult to see how the effects would ever become large enough to justify the expense and disruption that would be involved. This is especially true given the much larger percentage of the current population that is literate today, and would be forced to adjust to any changes.

Can a logographic script be meaningfully simplified and remain logographic?

The complete replacement of the Chinese writing system with a syllabary or alphabet—i.e., the abolition of logographic writing in favor of phonographic writing—would certainly have larger implications. For example, suppose that all syllables and morphemes pronounced *hú* were to be written with a single graph (say 胡), and that a single graph were similarly chosen for every syllable in the modern standard spoken language. While at first glance this might appear to be merely the endpoint of a long simplification process, in fact, it would represent a nearly complete break with the current system. For one thing, the syllabographs, despite a historical connection to earlier logographs, would no longer have any functional internal structure.<sup>51</sup> The merits and challenges of a phonographic writing system for Chinese are a subject worth consideration, and on which much ink has already been spilled; but these are questions that are beyond the scope of the current study.<sup>52</sup>

What we wish to consider here is whether, from a theoretical rather than a practical perspective, a highly regularized and systematic logographic writing system for Chinese would have any scientifically measurable advantages over the current systems; and whether by the same measure it would be better or worse than a purely phonographic system. As it happens, the initial results from psycholinguistic studies do suggest an answer to this question—one that is somewhat surprising.

---

<sup>51</sup> And, indeed, developments along these lines are exactly those that seem to have elicited the greatest opposition from Chinese intellectuals, as we have seen above.

<sup>52</sup> The practical problems of logographic writing in the computer age seem now to have been largely resolved with the advent of Unicode and ongoing processes of digital standardization (see Zhao 2008). And, as the examples of Japan and Taiwan show, there is no reason that a society employing logographic writing cannot compete effectively in the modern high-tech world. Moreover, as the Japanese example demonstrates, there are certain advantages to logographic writing that may motivate its retention even when a phonographic alternative is readily available.

## 9. Evidence from psycholinguistic studies

A great deal of progress has been made in the last 20 years in the psycholinguistic study of the learning, recognition, and reading of Chinese script; it is fair to say, however, that the field is still in its infancy (Shu et al. 2003:46). Many studies have yielded statistically significant results in answering the research questions posed, but it is not always clear to what degree the tasks that researchers test are directly correlated with real-world reading skills; the correlations are largely a matter of reasoned speculation. Most of the studies involve highly constrained tasks. Typically these involve images presented for brief durations of time on a computer screen, about which test subjects must make rapid judgments. Among the tasks are lexical decision tasks (deciding whether a graphic form represents a real word or morpheme of the language), naming tasks (pronouncing a graph aloud), identity decision tasks (determining whether two graphic forms are the same or not), and so on. Carefully designed studies can reveal statistically significant differences among groups of test subjects or sets of stimuli, which in turn permit the generation of hypotheses about cognitive functioning, which can in turn be further tested.

From the perspective of a descriptive linguist, most of these studies ignore or over-simplify important Chinese-language issues that may have a bearing on the research results. All but one of the studies discussed here fail to take explicit account of the Chinese languages and dialects involved, even though this is directly relevant to any studies involving naming tasks that make use of statistics on the degree of consistency of phonetic elements in characters.<sup>53</sup> None of the studies that I have seen make explicit

---

<sup>53</sup> Indeed, even if we could be sure that subjects in mainland China and in Taiwan all use standard Mandarin pronunciations (as Chen & Yuen (1991) tacitly assume), we would still need to adjust for the fact that standard Mandarin readings in both regions differ. (For example, the standard reading of 圾 in mainland China is *jī*, while in Taiwan it is *sè*. This difference is relevant when judging the regularity of the phonetic element 及 *jí*.) A notable exception to this trend is Shu et al. 2003,

Can a logographic script be meaningfully simplified and remain logographic?

mention of historical sound change as a source of irregularity, indeed they seem ignorant of it, e.g. Hsiao & Shillcock 2006:418: “[I]nterestingly, more than half of irregular characters still share some segments with their phonetic radicals”. Indeed, there is a markedly ahistorical understanding of the script (2006:419): “The distribution of types within the irregular phonetic radicals may be understood in terms of priming relationships between words. Note that one of the largest categories is the one in which the phonetic radical rhymes with the pronunciation of the whole character. There is a much smaller category of alliterating phonetic radicals, which share an onset with the pronunciation of the whole character. There is a substantial literature showing the salience of the rime in the phonological representation of words .... In phonological priming experiments, overlap at offset tends to lead to facilitation of the target .... If, we equate the rhyming phonetic radicals with such facilitation, then their preponderance in the irregular phonetic radicals can be understood.” This line of argumentation seems to arise from the assumption that Chinese characters were created based on modern Mandarin pronunciation.

Despite these problems, we can learn a great deal from the studies that have been done, much of which is relevant to our question about the theoretical implications of character simplification. Potential advantages and disadvantages of the simplification of Chinese characters have been noted in previous sections. But, as DeFrancis (1984:215) pointed out, “The true extent to which simplification has eased the burden of learning characters must remain a matter of subjective evaluation until there is firm supporting evidence.” Psycholinguistic studies carried out beginning in the 1990s offer us the first glimmerings of such evidence.

The psycholinguistic studies can be broadly categorized into three types. The first is concerned with measuring the “orthographic depth” of

---

which notes that Cantonese speakers have a harder time using phonetic components to learn and retrieve Mandarin pronunciations of characters (pp. 43-44).

the Chinese script and placing Chinese on a continuum with phonographic writing systems (e.g. Sproat 2000; Ellis et al. 2004); this type of study is inherently comparative across scripts. The second is concerned with the cognitive abilities of Chinese script readers and the role of Chinese script elements in the cognitive processing of Chinese characters (e.g. Chen et al. 1996; Hsiao & Shillcock 2006; Hsiao & Cottrell 2009; Williams & Bever 2010). The third is concerned with cognitive development in Chinese-speaking children as they memorize characters and learn to read (e.g. Chen & Yuen 1991; Ho et al. 2003; Shu et al. 2003; McBride-Chang et al. 2005). All of these studies are relevant to the issue of script simplification; some of them explicitly compare and contrast the cognitive processes of traditional and simplified script users (e.g. Chen & Yuen 1991; Liu & Hsiao 2012; McBride-Chang et al. 2005).

Orthographic depth is defined as “the directness and simplicity with which a writing system represents the phonology of a language” (Frost & Katz 1989:302). Alphabetic orthographies with transparent relationships to the language’s phonology—i.e. for which pronunciation is predictable from written form—are termed *shallow*. Those which deviate greatly from a one-to-one letter to phoneme correspondence are termed *deep*. The orthographic depth hypothesis predicts that “shallow orthographies should be easier to read using word-recognition processes that involve the language’s phonology” (Ellis et al. 2004:438), and therefore that users of such orthographies will be more likely to use phonological approaches, with “transparent orthographies supporting word recognition involving phonology, and opaque orthographies encouraging readers to process words by accessing the lexicon and meaning via the word’s visual appearance” (Ellis et al. 2004:446).

There appears to be a consensus among psycholinguists that lexical recognition can take place along multiple pathways, and that orthographic depth does indeed play a role in determining which pathways readers most

Can a logographic script be meaningfully simplified and remain logographic?

often make use of; but that these pathways are also influenced by the nature of the particular task at hand (Frost 1989). Sproat (2000:158), surveying the literature published in the 1990s, concludes that “Multiple routes from written form to pronunciation are available” and “all writing systems can be shown to make use of both a ‘lexical’, and a ‘phonological’ (i.e, rule-based) route.” The lexical route involves identification of a word without making use of phonological information; the pronunciation is determined after the word is identified. The phonological route involves making use of orthographic representation of pronunciation to identify the word.<sup>54</sup> Under the right conditions, Sproat concludes (2000:165ff), “deep” orthographies can show shallow processing effects, and shallow orthographies can show deep processing effects.

This is consistent with the intuitively obvious notion that Chinese script is “deep” in the sense that there is not a transparent one-to-one relationship between grapheme and phoneme, but that nevertheless the presence of phonetic elements render the script not completely opaque.<sup>55</sup> A number of studies, notably Hsiao & Shillcock 2006 and Shu et al. 2003, have compiled comprehensive statistics on the degree of regularity of phonetic elements in Chinese writing (though only from the perspective of Standard Mandarin pronunciation) in order to attempt to quantify the orthographic depth of the script. Their studies and others (e.g. Chen et al. 1996) have shown that readers make use of the phonetic components in certain reading-related tasks, and that the ability of student learners to do so increases over time as they master the script. What is not clear is how applicable this is to actual reading, as opposed to the highly specific tasks examined in the psycholinguistic experiments. As Williams & Bever (2010:591) say: “There is evidence of separate search patterns [i.e. lexical-

---

<sup>54</sup> See the Figure 5.1 and accompanying discussion at Sproat 2000:160.

<sup>55</sup> The placement of Chinese characters at the “deep” end of the continuum is confirmed experimentally by Ellis et al. 2004, albeit in studies of Japanese readers, not Chinese.

route search and phonology-route search], depending upon task type ... but there is much debate as to whether there is a default reading strategy that tips towards semantic or phonetic interpretation.” The studies also show, however, that the low degree of phonetic consistency makes this task quite difficult, and that students only become good at it gradually over time. This is because of the poor degree of phonetic consistency within Chinese characters, which we might also characterize as deepness of the orthography.<sup>56</sup>

It is clear from several studies (Chen et al. 1996, Williams & Bever 2010) that semantic and phonetic components of characters are recognized by Chinese readers and that they are used in reading-related tasks such as character recognition. Confirming and making more concrete the position quoted above from Sproat 2000, Williams & Bever 2010:589 observe in summarizing the results of their careful experiments:

Our results suggest that semantic and phonetic radicals [i.e. the two types of functional components] are each available for access when a corresponding task emphasizes one or the other kind of radical. But in a more neutral lexical recognition task, the semantic radical is more informative. Semantic radicals that correctly pertain to character meaning facilitated reaction time in semantic categorization tasks (Experiment #1), while radicals that had no immediately interpretable relation to character meaning had a strong inhibitory effect. Likewise, phonetic radicals that accurately indicated a character’s pronunciation facilitated a homonym recognition task (Experiment #2), whereas phonetic radicals that differed significantly in pronunciation from their character inhibited homonym recognition. In a lexical decision task (Experiment #3) where each character had either a blurred semantic radical or a blurred phonetic radical, the characters with a blurred semantic radical elicited a significantly higher error rate and a trend

---

<sup>56</sup> Shu et al. 2003:39 contains a sophisticated discussions of measures of phonetic consistency, which they define as “the degree of congruence in the pronunciations of the characters within a family,” where a family is a set of characters sharing the same phonetic element.



Can a logographic script be meaningfully simplified and remain logographic?

for longer response times. These results are interpreted to indicate that while educated native Chinese speakers have full use of both semantic and phonetic paths to character decoding, there is a slight predisposition to semantic decoding strategies over phonetic ones indicating that the semantic path is the default means of character recognition.

This predisposition does not reflect an inherent human bias in favoring semantic components over phonetic components; it may instead reflect adaptation to the contingent realities of the modern Chinese script, in which more consistency and regularity is found in the semantic components than the phonetic (Williams & Bever 2010:593):

Phonetic radicals in Chinese characters are unreliable indicators of pronunciation. Fan et al. (1984) estimated that only 26.3% of all semantic-phonetic compounds have a phonetic radical that is a reliable indicator of pronunciation. Additionally, when frequency is taken into account, the percentage of semantic-phonetic combinations that are pronounced identically to their phonetic portions falls further to a mere 18.5% (Zhu, 1987, cited in Hoosain, 1991). Hoosain (1991) noted that “the phonetic cuing function of phonetics is not rule governed, and the pronunciation of the phonetic itself, after all, has to be learned individually. This is quite distinct from the situation with the representation of sound by letters of the alphabet” (p. 11). In contrast, variable rates of accuracy from 65% (Fan, 1986, cited in Hoosain, 1991) to 100% (Jin, 1985, cited in Hoosain, 1991) have been found for specific semantic radicals, and most, if not all, semantic radicals are significantly more reliable than the 26% for the phonetic radicals. While the reliability of semantic radicals as predictors of semantic grouping varies from character to character, 100% of dictionary entries under semantic radicals such as 鱼 [/yu/: fish] and 鸟 [/niao/: bird] fit their respective categories. Additionally, when considering a lexical access model, the smaller corpus of semantic radicals (approximately 200) versus phonetic radicals (roughly 800 according to Taylor & Taylor, 1983), would suggest that lexical searches utilizing the smaller number of semantic

radicals would be inherently more efficient than searches based upon the much larger group of phonetic radicals.<sup>57</sup>

Similar conclusions are reached by Shu et al. (2003:40-44)<sup>58</sup>, who also observe that student learners are able to make use of semantic components earlier and more consistently than phonetic components. This is almost certainly because it is only after exposure to a larger number of lower-frequency characters that the patterned relationships of phonetic components starts to become apparent to script users, and because the lower number of distinct semantic elements means that students are exposed to larger semantic “families” at an earlier age. While orthographic depth is normally defined in relation to phonological representation, it may be useful when dealing with a logographic script of the Chinese type

---

<sup>57</sup> See the bibliography at the end of Williams & Bevers 2010 for full references to the works they cite.

<sup>58</sup> Shu et al. 2003:46: “So, does written Chinese have an orderly structure from which metalinguistically aware children can be expected to extract useful information? Or, is Chinese better described as a language [sic, should be “orthography”] that children must learn through repeated exposure and memorization? Although research on learning to read Chinese is still in its infancy, and only a handful of studies have been done, available evidence points clearly to the conclusion that written Chinese has a logic that young children can understand and use. The average effect size for phonetic regularity in four recent studies of Chinese children’s reading is 1.05 .... Similarly, the average effect size for semantic transparency in two recent studies is .92 .... These are large effects—large enough to be of more than theoretical interest .... At the same time, the robust effects of character frequency and familiarity indicate limits on the amount of information children are able to glean from characters. The average effect size of familiarity or frequency in four studies of Chinese children’s reading is 2.11 (Chan & Siegel, 2001; Ho & Bryant, 1997; Shu & Anderson, 1997; Shu, Anderson, et al., 2000). This is twice as large as the effect of regularity or transparency in roughly the same set of studies. Thus, the conclusion is complicated. Yes, it is important for Chinese children to use the logic of the writing system. However, there is no way for them to escape from repeated practice if they are to become skilled readers. Compared with the Western children speaking alphabetic languages studied by Goswami and her colleagues, the task Chinese children face in learning to read is more like the one facing English and French children than the one facing German, Greek, or Spanish children.”

Can a logographic script be meaningfully simplified and remain logographic?

to think about semantic orthographic depth, i.e. the degree to which the semantics of a represented morpheme fits with the semantic element in a character. Both Williams & Bever 2010 and Shu et al. 2003 make a persuasive case that in this sense the Chinese script is quite shallow.

What about differences in the processing of the traditional and simplified Chinese character scripts? The few psycholinguistic studies that explore this question show interesting results. As Hsiao & Cotrell (2009) have shown, skilled Chinese readers perceive characters more analytically and less holistically than novice Chinese readers; in other words, they tend to perceive Chinese characters in terms of their structural components.<sup>59</sup> There is a further tendency toward analytic perception in users of the simplified script. In a recent study that uncovered this trend, Liu & Hsiao (2012:689) conclude:

This effect may be because processing simplified characters generally requires more analytic processing due to higher visual similarity among characters compared with traditional characters .... This speculation is consistent with the recent finding that simplified Chinese readers have better visual skills than traditional Chinese readers ....

In other words, the increased similarity of sets of characters like 龍 ® 龙 *lóng* ‘dragon’ and 發 ® 发 *fā* ‘emit’ (discussed in section 5 above) may force readers of the simplified script to rely less on holistic processing because a higher number of characters would be confusable by this processing method. And, indeed, this too is substantiated evidentially. McBride-Chang et al. 2005:109 note “Chen and Yuen (1991) did find some differences in visual processing in their study of children aged 7 to 9.3. Specifically, children from China were more likely to make visual errors in character recognition than were children from Hong Kong. This difference

---

<sup>59</sup> This is hardly a surprising result given the discussion in sections 5 and 6.

in error patterns was attributed to differences in script across groups. Chen and Yuen (1991) argued that because the number of strokes is fewer in the simplified script, distinguishing among characters may be more difficult in beginning reading.”

What the studies don't tell us, as far as I can tell, is whether holistic or analytic processing overall is faster or more efficient when reading. Clearly, it seems that readers naturally develop whichever processing technique is most effective in coping with the features of the script they are learning. But I am not aware of any studies which attempt to simply measure, comparatively, reading speeds of the same text in traditional and simplified forms. What does seem clear is that by creating even more and fragmented “phonetic families” of characters (cf. the traditional “囊” family that becomes fragmented into a “上” and a “囊” family), any advantage conveyed by increased analytic ability may be outweighed in student learners by a greater degree of phonetic inconsistency and thus of orthographic depth.

McBride-Chang et al. 2005:123 hypothesize that

If children are exposed to a simplified script literacy environment, they might make greater use of visual skills in learning about this environment. If children exposed to the simplified script are prone more to visual errors because the characters written in this script have fewer features and are, therefore, more difficult to distinguish, they may gradually acquire more reliance on visual cues to discriminate print .... The traditional script, because it contains more visual features, may be easier to discriminate initially .... In addition, the phonetics and semantic radicals in this script may be more regular than in the simplified one, promoting sound- or meaning-based strategy use earlier than in the simplified script.

Can a logographic script be meaningfully simplified and remain logographic?

## **10. What would a truly “simplified” logographic Chinese script look like?**

Having reviewed some of the relevant psycholinguistic literature, we are now in a position to conduct a thought experiment: what might an idealized simplified character system look like?

First, it would increase the accuracy and consistency of the role of the phonetic elements. A sensible approach would be to select, from among the current set of 800 phonetics, approximately 400, one for each basic syllable of Modern Standard Mandarin (discounting tone).

The new set of phonetic elements could, in isolated form, serve as a non-tonal syllabary, but in our theoretical writing system, they will not be used that way. Rather, for each set of homophonous morphemes in common use, a semantic component will be selected and compounded with the phonetic element. (See appendix 2 for an example of how this might be done.) The approximately 200 semantic components in use in the Chinese script already work quite well, but they could be consolidated and reduced somewhat, eliminating those that are especially rare or semantically opaque and consolidating those that have considerable semantic overlap.<sup>60</sup>

More importantly, positional variation could be eliminated, with semantic components consistently occurring in only the top or left positions.<sup>61</sup>

With this new system, the degree of consistency in phonetic elements would increase from roughly 38% (depending on how one measures this

---

<sup>60</sup> As an example, the three semantic elements 彳, 行, 走, and 辶 all of which are related to movement (such as walking, running, or traveling), could be consolidated. On the other hand, retaining some “synonymous” but visually distinct semantic elements might be helpful in preserving graphic distinctions that write homophonous morphemes in the same general semantic domain.

<sup>61</sup> Ho et al. 2003:869 conclude based on their studies that “knowledge of character structure and position of semantic radicals (but not of the position of phonetic radicals) is important for children learning to read Chinese words.”

feature; this figure is from Hsiao & Shillcock 2006:407) to approximately 100%. For those syllable types with an exceptionally large number of morphemes (such as *fu* and *shi*), such that there are homophonous morphemes that fall into the same semantic field and would therefore be written with identical compounds, two or more distinct phonetic elements could be employed, ideally based on tonal distinctions.

The selection and form of the phonetic elements would be based not on the goal of reducing stroke numbers, but on preserving graphic distinctiveness. Some especially high-frequency morphemes could still be written with unitary characters (such as neutral-tone grammatical particles including *de* 的 and *le* 了).

Such a writing system would be highly transparent; it would expand the options for cognitive processing, allowing students and skilled script users to employ phonetic elements and semantic elements for reading-related tasks such as character recognition and learning. Because such a system would impose fewer constraints related to irregularity, opacity, and visual confusion, readers could develop whichever processing strategies are most advantageous for practical tasks involved in literacy. And because the system would make use of familiar functional elements from the current script, it would presumably allow for a smoother transition for literate Mandarin speakers.

Having constructed an idealized simplified character system, we can now ask a follow-up question: might such a system have an advantage over a purely phonographic system such as a syllabary? While the answer at this point must be considered highly speculative, the available studies suggest that the answer is yes.

This is because of the valuable role that psycholinguistic studies suggest is played by semantic elements of characters, especially given the high degree of monosyllabic homophony found among the morphemes of Modern Standard Chinese. This would presumably confer an advantage on

Can a logographic script be meaningfully simplified and remain logographic?

readers of our new Chinese script that is not enjoyed by users of shallow, purely orthographic scripts like the Japanese *kana* syllabaries.

It would seem then that the answer to the question posed in the title of this paper—*Can a logographic script be meaningfully simplified and remain logographic?*—is yes. A logographic script can be meaningfully simplified—in a sense that is scientifically measurable—and remain logographic. (It might be preferable, however, to do away with the contentious term *simplified* and replace it with a more objective term like *regularized* or *improved*.) More surprisingly, there is reason to think that such a script might be more efficient than a purely phonographic script. One way to think about this is that such a script would not only be orthographically shallow, with a close relationship between orthographic form and pronunciation, but would also be semantically transparent in ways that studies have already shown is beneficial to readers. Such a script type would not necessarily have universal application, since the available studies are intimately tied up with certain typological features of Chinese, namely monosyllabicity and homophony of morphemes.

In this discussion I have glossed over an important consideration, and that is the fact that users of the Chinese script for writing Modern Standard Written Chinese speak many different Chinese languages and dialects, many of which have their own “dialectal” reading pronunciation of characters. A regularization of the kind proposed here necessarily requires that one pronunciation system be chosen as the basis for the selection of phonetic elements, and this inevitably will reduce the functionality of the script for speakers of other varieties of Chinese. The gains in functionality that I have outlined above would apply only to speakers of Mandarin; they would be offset to an unknown degree by a loss of pan-dialectic applicability of the script.

Our current thought experiment is concerned with creating a Chinese script, based on the model of the existing scripts, that is more efficient and

consistent—by virtue in part of being more orthographically shallow—at least in terms of the spoken variety of Chinese it is based on. Such a script would not necessarily be a true innovation. There is reason to consider whether the Chinese script might not have been similar to our thought-experiment proposal early in its history, during the Old Chinese period in the first millennium BCE. There is some evidence that there was a nearly one-to-one relationship of phonetic element to base-syllable shape (\*CVC) in the Old Chinese period; and that it was only as a result of sound changes and language divergence in subsequent millennia that the orthographic depth of Chinese writing grew to the point where we see it today. One might even argue, then, that a regularized Chinese script would constitute a return of sorts to the script's origins.<sup>62</sup>

## **11. Conclusion - logography and the question of simplicity and complexity**

We have seen that, at a simple level, 20<sup>th</sup> century character simplification has reduced visual complexity in many individual graphs. But this is only obviously true when the metric for complexity is the number of strokes. This begs the question of how complexity of a writing system can or should be measured.<sup>63</sup>

---

<sup>62</sup> I recognized that this is a vastly oversimplified statement. In the early period of the Chinese script, before standardization, the choice of semantic element was highly fluid; and even early on, as sound changes occurred, the choice of phonetic element could also be quite fluid (see Boltz 1994). Nevertheless, the claim that the script was more orthographically shallow in the first millennium BCE than it is today is a valid one.

<sup>63</sup> Frost (1989:163) conceives of complexity and simplicity this way: "Complexity is generally related to the amount of effort needed for decoding a given word. One possible source of complexity that merits close examination is the lack of transparent correspondence between orthographic and phonologic subunits. Because the purpose of orthographic systems is the representation of phonology [sic -- is this in fact always true?], whether the skilled reader uses this information or not, the



Can a logographic script be meaningfully simplified and remain logographic?

Beyond number of strokes per character, which is clearly an inadequate measure, some other possibilities are:

- total number of distinctive components in the system
- average number of strokes in these distinctive components (perhaps weighted by frequency)
- frequency distribution of occurrence of components
- degree of correlation of components (both semantic and phonetic) with contemporary spoken language
- degree of redundancy and consistency of function of components
- degree of allography and degree of visual similarity across allographs

It is clear that, as with measures of complexity in other language systems (phonology, morphology, syntax), factors like systemic context, analogy, patterning, contrast, and the like must be taken into consideration. Psycholinguistic studies of the last 20 years have started to provide quantitative means for evaluating Chinese writing by these measures, and to connect those measures to practical tasks related to literacy. In this study I have attempted to apply these results not just to an evaluation of the existing scripts, but to the theoretical question of how Chinese writing might have been (or still could be) more meaningfully reformed. As I noted earlier, I believe this is more useful as a theoretical thought experiment than as a practical policy prescription.

I have also pointed out that the current state of psycholinguistic research is not sufficient to answer all of our questions, theoretical or practical, related to the functioning of Chinese. Issues of spoken language norms and of diachronic and cross-dialectal factors remain to be explored further. Moreover, there is a very real human dimension that cannot be ignored when considering issues of writing reform, a dimension that is in

---

relative directness and simplicity—the transparency—of this representation can be of major importance.”

many respects far more important than the theoretical linguistic dimension. Cultural context plays a role in determining complexity, because complexity that relies on inherited cultural knowledge will seem simpler—less complex—to the participants in that culture than will complexity introduced anew from outside the culture. In practical terms, this means that the older generation may not be able or willing to accept a simplified system that is more complex *given what has been learned before*, but which might be inherently less complex to a blank slate generation. The idealized logographic Chinese script I have described, if actually put into practice, would no doubt be highly disruptive. So while to my knowledge this is the first study which attempts to apply the results of the psycholinguistic literature on Chinese *not* to analysis of synchronic processing of scripts, but to forward-looking questions related to script modification, simplification, and typology, the results of this study are intended to be purely theoretical. They do not constitute advocacy for changes in Chinese writing systems, although it is conceivable that they might have practical applications some time in the future.

For native script users growing up in China today, arguments about script reform are largely irrelevant. In my experience, literate Chinese (whether in the Taiwan or the PRC) do not perceive their writing system to be particularly difficult, nor do they bemoan the time and effort taken to learn it. The notion that writing reform is essential to the success of China is so outdated as to seem faintly ridiculous. It is perhaps fair to say that the group most deeply and frustratingly affected by 20<sup>th</sup> century Chinese writing reform are academics, researchers, and intellectuals—and especially adult foreign learners of Chinese—most of whom now have to master two scripts in order to be fully competent in Chinese. There is, in my opinion, no practical need for any changes in the writing systems for Chinese, which are serving their current users quite well.

Can a logographic script be meaningfully simplified and remain logographic?

## Appendix 1

A passage from the story *Ā Q Zhèng Zhuàn* 阿Q正传 [*The True Story of Ah Q*] by Lǚ Xun, in both traditional and simplified scripts. Character forms that differ in the two scripts are highlighted.

然而要做這一篇速朽的文章，才下筆，便感到萬分的困難了。第一是文章的名目。孔子曰，“名不正則言不順”。這原是應該極注意的。傳的名目很繁多：列傳，自傳，內傳，外傳，別傳，家傳，小傳……，而可惜都不合。“列傳”麼，這一篇並非和許多闊人排在“正史”裡；“自傳”麼，我又並非就是阿Q。說是“外傳”，“內傳”在那裡呢？倘用“內傳”，阿Q又決不是神仙。“別傳”呢，阿Q實在未曾有大總統上諭宣付國史館立“本傳”——雖說英國正史上並無“博徒列傳”，而文豪迭更司也做過《博徒別傳》這一部書，但文豪則可，在我輩卻不可。其次是“家傳”，則我既不知與阿Q是否同宗，也未曾受他子孫的拜託；或“小傳”，則阿Q又更無別的“大傳”了。總而言之，這一篇也便是“本傳”，但從我的文章著想，因為文體卑下，是“引車賣漿者流”所用的話，所以不敢僭稱，便從不入三教九流的小說家所謂“閒話休題言歸正傳”這一句套話裡，取出“正傳”兩個字來，作為名目，即使與古人所撰《書法正傳》的“正傳”字面上很相混，也顧不得了。

然而要做這一篇速朽的文章，才下筆，便感到萬分的困難了。第一是文章的名目。孔子曰，“名不正則言不順”。這原是應該極注意的。傳的名目很繁多：列傳，自傳，內傳，外傳，別傳，家傳，小傳……，而可惜都不合。“列傳”么，這一篇并非和许多阔人排在“正史”里；“自传”么，我又并非就是阿Q。说是“外传”，“内传”在那里呢？倘用“内传”，阿Q又决不是神仙。“别传”呢，阿Q实在未曾有大总统上谕宣付国史馆立“本传”——虽说英国正史上并无“博徒列传”，而文豪迭更司也做过《博徒别传》这一部书，但文豪则可，在我辈却不可。其次是“家传”，则我既不知与阿Q是否同宗，也未曾受他子孙的拜托；或“小传”，则阿Q又更无别的“大传”了。总而言之，这一篇也便是“本传”，但从我的文章着想，因为文体卑下，是“引车卖浆者流”所用的话，所以不敢僭称，便从不入三教九流的小说家所谓“闲话休题言归正传”这一句套话里，取出“正传”两个字来，作为名目，即使与古人所撰《书法正传》的“正传”字面上很相混，也顾不得了。

## Appendix 2

An example of Modern Standard Mandarin morphemes pronounced *juan* and how they might be written in a regularized system.

A *Chinese-English Dictionary* (Beijing: Wuyun Shuguan 1988) lists 11 simplified characters that represent 13 morphemes with the pronunciation *juan*. As part of the thought experiment outlined in section 10 of this paper, new graphs for these morphemes might be created as follows:

Current graph (simp.)	Pronunciation	Meaning	Proposal/ Semantic
涓	<i>juān</i>	tiny stream	㇀ + 卷 / water
捐	<i>juān</i>	donate	貝 + 卷 / valuables
娟	<i>juān</i>	graceful	女 + 卷 / woman
圈	<i>juān / juàn</i>	to pen / a pen	口 + 卷 / enclosure (圈)
鵲 (鵲)	<i>juān</i>	cuckoo	鳥 + 卷 / bird
鐫 (鐫)	<i>juān</i>	engrave	金 + 卷 / metal
捲 (卷)	<i>juǎn</i>	roll up	扌 + 卷 / hand (捲)
卷	<i>juàn</i>	book	+ 卷 / bamboo
倦	<i>juàn</i>	weary	亻 + 卷 / person (倦)
絹 (絹)	<i>juàn</i>	type of silk	纟 + 卷 / silk
雋 (隹)	<i>juàn</i>	meaningful	言 + 卷 / speech
眷	<i>juàn</i>	think fondly of	忄 + 卷 / heart

## REFERENCES

- Anonymous (ed.) (1958). *Reform of the Chinese written language*. Peking: Foreign Languages Press.
- Barnes, Dayle (1988-1989). A continuity of constraints on orthographic change: Chen Guangyao and character simplification. *Monumenta Serica* 38:135-166.

Can a logographic script be meaningfully simplified and remain logographic?

- Behr, Wolfgang (2010). In the interstices of representation: ludic writing and the locus of polysemy in the Chinese sign. In Alexander J. de Voogt and Irving L. Finkel (eds.) *The idea of writing: play and complexity*. Boston: Brill. 281-314.
- Boltz, William G. (1994). *The origin and early development of the Chinese writing system*. (American Oriental Series, 78.) New Haven: American Oriental Society.
- Chang, Su-chen (1976). Reduction in number of Chinese characters. *Journal of the Chinese Language Teachers Association* XI:3. 187-191.
- Chen, Guangyao 陈光垚 (1956). *Jianhua Hanzi ziti shuoming* 简化汉字字体说明 [Explanation of simplified Chinese character forms]. Beijing: Zhonghua Shuju.
- Chen, Mary J. & Joseph C.-K. Yuen (1991). Effects of pinyin and script type on verbal processing: Comparisons of China, Taiwan, and Hong Kong experience. *International Journal of Behavioral Development* 14:4. 429-448.
- Chen, Ping (1999). *Modern Chinese: history and sociolinguistics*. Cambridge, UK: Cambridge University Press.
- Chen, Y.P., D.A. Allport & J.C. Marshall (1996). What are the functional orthographic units in Chinese word recognition: The stroke or the stroke pattern? *The Quarterly Journal of Experimental Psychology* 49A:4. 1024-1043.
- Cheng, Chin-chuan (1975). Directions of Chinese character simplification. *Journal of Chinese Linguistics* 3:2/3. 213-220.
- Coblin, W. South (2007). *A handbook of 'Phags-Pa Chinese*. Honolulu: University of Hawai'i Press.
- DeFrancis, John & J. Marshall Unger (1994). Rejoinder to Geoffrey Sampson, "Chinese script and the diversity of writing systems". *Linguistics* 32. 549-554.
- DeFrancis, John (1984). *The Chinese language: fact and fantasy*. Honolulu: University of Hawai'i Press.
- Ellis, Nick C., Miwa Natsume, Katerina Stavropoulou, Lorenc Hoxhallari, Victor H.P. Van Daal, Nicoletta Polyzoe, Maria-Lousia Tsipa & Michalis Petalas (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly* 39:4. 438-468.
- Frost, Ram & Leonard Katz (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition* 17:3. 302-310.

Proceedings of the SCRIPTA 2012, Seoul, Oct. 9~11, 2012

- Frost, Ram (1989). Orthography and phonology: the psychological reality of orthographic depth. *Haskins Laboratories Status Report on Speech Research* 1989: SR-99/100. 162-171.
- Handel, Zev (2009). Towards a comparative study of sinographic writing strategies in Korean, Japanese, and Vietnamese. *Scripta* 1. 89-125.
- Hannas, William C. (1997). *Asia's orthographic dilemma*. Honolulu: University of Hawai'i.
- Ho, Connie S.-H., Ting-Ting Ng & Wing-Kin Ng (2003). A "radical" approach to reading development in Chinese: The role of semantic radicals and phonetic radicals. *Journal of Literacy Research* 35:3. 849-878.
- Hsiao, Janet H. & Garrison W. Cottrell (2009). Not all visual expertise is holistic, but it may be leftist: The case of Chinese character recognition. *Psychological Science* 20:4. 455-463.
- Hsiao, Janet H. & Richard Shillcock (2006). Analysis of a Chinese phonetic compound database: Implications for orthographic processing. *Journal of Psycholinguistic Research* 35:5. 405-426.
- Lehmann, Winfred P. (ed.) (1975). *Language and Linguistics in the People's Republic of China*. Austin: University of Texas Press.
- Liu, Tianyin & Janet H. Hsiao (2012). The perception of simplified and traditional Chinese characters in the eye of simplified and traditional Chinese readers. *Proceedings of CogSci 2012* (<http://mindmodeling.org/cogsci2012/index.html>, accessed September 13, 2012.)
- Liu Fu 劉復 and Li Jiarui 李家瑞 (eds.) (1930). *Song-Yuan yilai suzi pu* 宋元以來俗字譜 [Table of vulgar character forms from the Song-Yuan period and later]. Beijing: Guoli Zhongyang Yanjiuyuan Lishi Yuyan Yanjiusuo.
- McBride-Chang, Catherine, Bonnie W-Y. Chow, Yiping Zhong, Stephen Burgess & William G. Hayward (2005). Chinese character acquisition and visual skills in two Chinese scripts. *Reading and Writing* 18:2. 99-128.
- Milsky, Constantin (1974). *Préparation de la réforme de l'écriture en République populaire de Chine 1949-1954*. The Hague: Mouton.
- Myers, James (2011). The psychological reality of formal regularities in Chinese characters. Paper presented at the 7th Conference of the European Association of Chinese Linguistics, Venice.

Can a logographic script be meaningfully simplified and remain logographic?

- Norman, Jerry (1980). Linguistics. In Leo A. Orleans & Carline Davidson (eds.) *Science in contemporary China*. Stanford, CA: Stanford University Press. 490-496.
- Norman, Jerry (1988). *Chinese*. New York: Cambridge University Press.
- Qiu Xigui 裘錫圭 (2000). *Chinese writing*. (Translated by Jerry Norman & Gil Mattos). Berkeley: The Society for the Study of Early China and the Institute of East Asian Studies, University of California.
- Ramsey, S. Robert (1987). *The languages of China*. Princeton University Press.
- Seybolt, Peter J. and Gregory K.-K. Chiang (eds.) (1979). *Language reform in China: documents and commentary*. White Plains, N.Y.: M.E. Sharpe.
- Shu, Hua, Xi Chen, Richard C. Anderson, Ningning Wu & Yue Xuan (2003). Properties of school Chinese: implications for learning to read. *Child Development* 74:1. 27-47.
- Sproat, Richard (2000). *A computational theory of writing systems*. Cambridge: Cambridge University Press.
- Unger, J. Marshall & John DeFrancis (1995). Logographic and semasiographic writing systems: a critique of Sampson's classification. In Insup Taylor & David Olson (eds.) *Scripts and literacy: reading and learning to read alphabets, syllabaries and characters*. Dordrecht: Kluwer. 45-58.
- Williams, Clay & Thomas Bever (2010). Chinese character decoding: A semantic bias? *Reading and Writing* 23:5. 589-605.
- Wu Yu-Chang (1958). Report on the current tasks of reforming the written language and the draft scheme for a Chinese phonetic alphabet. In Anonymous (1958). 30-54.
- Yi Xiwu 易熙吾 (1956). *Jianti ziyuan* 簡體字原 [Sources of simplified characters]. Beijing: Zhonghua shu ju.
- Yu, Xialong 于夏龙 (1978). Guanyu di'er ci hanzi jianhua gongzuo de yixie yijian 关于第二次汉字简化工作的一些意见 [Some views on the second simplification of Chinese characters]. *Zhongguo Yuwen* 1978:2. 127-129.
- Zhao, Shouhui and Richard B. Baldauf Jr. (eds.) (2008). *Planning Chinese characters: reaction, evolution, or revolution?* (Language Policy, 9.) Dordrecht: Springer.
- Zhao, Shouhui (2008). Chinese character modernization in the digital era: a historical perspective. In Robert B. Kaplan and Richard B. Baldauf Jr. (eds.)

Proceedings of the SCRIPTA 2012, Seoul, Oct. 9~11, 2012

*Language planning and policy in Asia, Vol. 1: Japan, Nepal and Taiwan and Chinese characters.* NY: Multilingual Matters. 38-101.

Zhou, En-Lai (1958). Current tasks of reforming the written language. In Anonymous (1958). 7-29.



Can a logographic script be meaningfully simplified and remain logographic?

## Can a logographic script be meaningfully simplified and remain logographic?

Lessons from the 20th century Chinese writing  
reform informed by recent psycholinguistic  
research

Jeon Kwangjin

Sungkyunkwan University, KOREA

It is an honor to be designated the moderator for this praiseworthy thesis, which encompasses a total of 43 pages. I would also like to give thanks to (NAME OF SOCIETY) for giving me the chance to read such a remarkable paper in advance. I only regret that my ability to comprehend English is rather low and I worried whether I could grasp the text's true meaning.

In his paper, Professor Zev Handel prepared three goals in order for the topic of whether there is a way to simplify the Chinese script while retaining its basic logographic character. The first goal is the description and analysis of the character-internal problem resulting from simplification, the second the assessment of the practicality of the results of simplification, and the third is new research and suggestions about the theoretical methods regarding simplification.

The most creative part of the paper's broad-scoped contents is the new simplification method via phonetic component described on page 40 of appendix 2. I would like to ask a few questions regarding this method.

(1) According to your paper, the total number of phonetic components written in hanzi in Modern Chinese is said to be about 800 (page 35). This number is the result of whose research? The number of permissible syllables in Modern Chinese is said to amount to 418 according to Oh, Jong-Chae (1992, 133). If we were to make new hanzi with identical syllables representing identical phonetic values and adding each different semantic component, the 800 phonetic components would decrease to 400. But have you considered how much confusion this would cause if this method were actually used?

(2) If we take Modern Mandarin Chinese as the standard, according to Dong Gon (1998, 167) the highest number of hanzi representing identical syllables are as follows; yi (177), ji (163), yu (139), li (133), xi (130), and zhi (128). One idea could be to simplify these 177 characters representing the same phonetic component, but don't the losses outweigh the gains? (As we say in Korean, "Aren't we breaking the jar trying to catch the mouse?")

(3) Early last century in China, there was a claim (Jeok Geon-ung) to establish some 454 'syllabic hanzi' of very simple shape. Also in March 1953 according to Chairman Mao Zedong's instruction, the Committee for the Reformation of Chinese Characters enacted legislation to keep only 1469 'Simplified Chinese Characters' and discard all others (Chart of 1469 Simplified Characters). But it caused many unexpected problems and consequently was repealed. What are your thoughts on this precedent?

(4) As you, Professor Handel have indicated among simplified characters used in Modern Chinese, there is a contradiction as seen in the characters 賈 ⇒ 賈, 價 ⇒ 價, 環 ⇒ 環, 懷 ⇒ 懷. Two principles can be said to be the cause of these defects, the first of which, *sulibujak*, loosely translates as "to follow tradition without innovation". The second principle

Can a logographic script be meaningfully simplified and remain logographic?  
is *yakjeongsokseong*, which means that even incorrect things can become standard through habit. Do you suggest that these principles be discarded?

Lastly I would like to give one recommendation. The paper's creative claim, i.e. description focusing on originality, a premise reported to specialists and not to laypersons. However, the removal of both part two (Chinese character structure and features) and part three (Chinese character simplification and history) would improve the level of the paper. How do you feel about the suggested removal?

Besides what I have already said there is much I would like to ask, but in the interest of time I will finish here. Thank you.

Proceedings of the SCRIPTA 2012, Seoul, Oct. 9~11, 2012