To cite this paper:

Zhang Q, Reilly RG. (2020). What regions of Chinese characters are crucial for recognition? A web-based study. *Journal of Chinese Writing Systems*. doi:10.1177/2513850220950020

What Regions of Chinese Characters are Crucial for Recognition? A Web-based Study

Qi Zhang

School of Applied Language and Intercultural Studies, Dublin City University, Ireland

Ronan Reilly

Department of Computer Science, Maynooth University, Ireland

Abstract

The study involves approximately 250 Chinese university students from eight institutions to determine what parts of a representative sample of Chinese characters are crucial to their correct identification. One-hundred and two simplified characters were presented to participants using a web-based experimental platform. The characters were partially obscured using a gaussian blurring technique. The direction of maximum blur could either be from top to bottom, bottom to top, left to right, or right to left. Participants were asked to identify the blurred character and type its pinyin. Overall, participants correctly identified 88% of characters. The effects of all forms of blurring on correct recognition were correlated with character structures. Phonetic radicals seem to be more sensitive to the blurring than semantic radicals, while the radical transparency and radical frequency also plays a role in the recognition accuracy. The blurring conditions that impacted most significantly on correct recognition were top to bottom and bottom to top, which caused, respectively, the upper and lower parts of the character to be obscured.

Keywords: Chinese characters, recognition, blurring

What Regions of Chinese Characters are Crucial for Recognition? A Web-based Study

The uniqueness of the Chinese writing system has attracted the attention of scholars in various fields interested in exploring its acquisition and processing. The low-level aspects of reading Chinese are, of necessity, different from those of alphabetic scripts given its logographic origin. For alphabetic scripts, a number of orthographic units such as letters can be mapped onto phonemes and recombined to form written words. For this reason, learning to read an alphabetic writing system can strengthen the phonological representations of words. Reading proficiency can therefore be improved through success in establishing the phonological connections to orthography (Tan et al., 2005). However, the logographic writing system of Chinese does not provide systematic and reliable grapheme-phoneme correspondences (Shu et al., 2003; Xu et al., 2013). Even though native children have plenty of exposure to Chinese and have developed phonology-to-semantics links before formal schooling, it takes time to establish the connection between phonology and orthography (Li et al., 2002, 2016). Thus, it has been a tradition for native Chinese-speaking children to invest a great deal of time and energy in learning characters (Lam, 2011). Tan et al. (2005) suggested that, through extensive copying and writing characters, learners can consolidate knowledge of motor programming of character strokes and stroke patterns and eventually enhance their reading ability.

There has been a number of studies investigating the processing of Chinese characters in order to understand the acquisition and reading of Chinese language (Chen and Shu, 2001; Mo et al., 2015; Perfetti et al., 2005; Taft, 2006). Similar to the graphic symbols of letters and words used in research on alphabetic language reading, the basic orthographic units in Chinese reading research have been variously assumed to be strokes, stroke patterns, radicals, and whole characters. Results have been ambivalent; some researchers have found evidence to support the hypothesis that readers process Chinese characters in an analytic manner (Chen and Yeh, 2015; Kuo et al., 2014) or in a holistic way (Mo et al., 2015). Others propose that expert readers process characters less holistically than novices (Hsiao and Cottrell, 2009; Tso et al., 2012). Liu and Hsiao (2012) further pointed out that simplified characters may require more analytic processing than traditional characters.

The discussion of analytic and holistic processing of Chinese characters may contribute to the analyses of the data collected by using a gaussian blurring technique in the current study. The type of character structure, visual complexity, radical transparency and radical frequency are factors considered. Using a web platform, the study collected a large quantity of data from approximately 250 participants regarding their recognition of around 100 characters, which will offer insights into identifying the critical parts of these characters for successful recognition.

The Structure and Processing of Chinese Characters

The most significant difference between the Chinese and alphabetic systems is the linear arrangement of letters versus the two-dimensional configuration of characters for Chinese. Chinese characters ‘are packed into a square configuration, possessing a high, nonlinear visual complexity’ (Tan et al., 2005: 5). The orthographic composition of a Chinese character usually consists of strokes and radicals. Several strokes function as building blocks to construct a radical, and one or more radicals are used to form a character. The number of strokes for most Chinese characters ranges from 6 to 13 (Anderson et al., 2013). There are generally two kinds of Chinese characters: integral and compound (Shen and Ke, 2007; Wang et al., 2003). The former are composed using one radical only, while the latter often consist of two radicals including a semantic and a phonetic one. In other words, certain radicals can also be individual characters. The clear majority (approximately 80%) of characters in modern Chinese are semantic-phonetic compound characters with semantic radical on the left or the top of a character and the phonetic radical on the right or the bottom (Kuo et al., 2014; Shu et al., 2003). For example, the compound character 妈 (pronounced as mā, meaning mother) consists of the semantic radical 女 (meaning female) on the left and the phonetic radical 马 (pronounced as mǎ) on the right. As shown here, the semantic radical only offers information regarding the meaning of the character, whereas the phonetic radical provides information about pronunciation of the character.

Some radicals are independent and can be individual characters, such as the two examples above, 女 and 马. However, there are also radicals that cannot appear alone and have to be used as a part of a compound character. For instance, the radical 讠 often connotes “to speak" and has to be bound with another component to form a character, e.g. 记 (jì, meaning language) with the bound semantic radical 讠 on the left and the phonetic radical 己 (pronounced as jǐ) on the right. Interestingly, some radicals can be further divided into subcomponents. Subcomponents, also called chunks, are basic stroke sequences that do not have meaning or pronunciation but can be used to compose radicals and characters (Anderson et al., 2013; Cao, Rickles, et al., 2013; Shu et al., 2003). For example, the character 格 (meaning grid, pronounced as gé) is composed of the left semantic radical 木 (meaning wood) and the right phonetic radical 各 (pronounced as gè). However, the radical 各 can be further deconstructed into chunks 夂 and 口. Furthermore, Chinese characters also went through a simplification process (DeFrancis, 1984). In an effort to reduce illiteracy, the Chinese government published a list of simplified characters in 1956. For instance, the traditional character 頭 (tóu, meaning head) was simplified into 头. After this simplification, the phonetic radical positioned on the left of this character (i.e. 豆 pronounced as dòu) has been lost.

There is a growing body of research focusing on learning to read Chinese characters by native (e.g. Perfetti and Tan, 1998; Shu et al., 2003; Taft and Zhu, 1997; Tong and McBride, 2014) and non-native speakers (Cao, Vu, et al., 2013; Chang et al., 2014; Wang et al., 2003) involving behavioural and neuroimaging studies. They can be generally categorised into two kinds. One investigates the writing-on-reading effect in the Chinese writing system (e.g. Zhang and Reilly, 2015). Guan et al. (2011) pointed out that Chinese orthography ‘involves the coupling of writing related visual and motor systems’. This coupling may help establish the spatial configuration of strokes and radicals, which along with a temporal sequence of motor movements associated with stroke composition, completely defines the shape of the character (see also Cao, Vu, et al., 2013). Therefore, significant spatial analysis and highly organised motor activity is involved in writing a Chinese character (Tan et al., 2005), which in turn contribute to the reading of the character.

Another kind of research examines character recognition through encoding a character's components – regardless of whether the component refers to strokes, radicals or chunks. Indeed, previous research revealed the importance of various training methods, such as handwriting (Guan et al., 2011), visual chunking (Chang et al., 2014; Chen et al., 2013), or radical knowledge application (Shen and Ke, 2007; Xu et al., 2014), in assisting character learning by learners of Chinese as a foreign language (Henceforth CFL). For instance, Xu et al. (2013) designed an animation showing how a character was written stroke-by-stroke and played the animation to learners, which was shown to contribute positively to the recognition of the character by CFL learners. In a study of learning strategies, Grenfell and Harris (2015) also reported that the individual components of each character might be a key for CFL learners to help memorise new characters.

The study reported in this paper is the first phase of a larger project that aims to investigate the recognition of characters by native speakers in order to compare such information with findings of future research among CFL learners. Therefore, before we look into effective ways of recognising Chinese characters, it is necessary to understand how characters are processed. There have been different proposals in processing Chinese characters based on the analysis of Chinese character structure. The analytic hypothesis proposes that readers need to identify a component of Chinese character (e.g. a chunk or radical) to recognise it (Anderson et al., 2013). On the other hand, the holistic hypothesis suggests that 'each character itself is an orthographic processing unit' (Chen and Yeh, 2015: 48). As expert readers can decompose and recombine components of characters, they can attend to character parts and process characters analytically (Kuo et al., 2014; Shu and Anderson, 1997; Tso et al., 2012). In comparison, since novice learners of Chinese have not been trained to deconstruct and then reconstruct individual features into components, they would be unable to selectively attend to character parts and consequently be unlikely to reduce the information load through ignoring irrelevant parts of a character. They are therefore more likely to process characters holistically (Hsiao and Cottrell, 2009). However, Hsiao and Cottrell (2009: 7) also emphasised that 'whether holistic processing is employed depends on the stimulus features and the task typically performed on the stimuli'.

Further analyses indeed revealed that different processing skills might be required because of different stimulus features. Probably due to high similarity between simplified characters, readers who grew up using the simplified script (e.g., adults from mainland China) appeared to employ a less holistic approach to process simplified characters than Hong Kong adult readers of traditional script (Liu and Hsiao, 2012). When characters are put in words and text being tested in normal reading, Tao and Healy (2002) indicated that high-frequency words might be processed holistically while low-frequency words are processed analytically due to their unfamiliarity. Interestingly, this familiarity effect was observed only at the level of words rather than characters (ibid.).

The basic units involved in the analytic processing of Chinese characters are also different. Several studies highlighted the importance of radicals as the mediator for orthographic processing of Chinese characters (e.g. Chen and Yeh, 2015; Ding et al., 2004; Taft and Zhu, 1997). Contributing effects on character processing can be observed in radical frequency (Taft and Zhu, 1997), radical or chunk familiarity (Anderson et al., 2013), radical regularity (Shu and Anderson, 1997). The following table summarises the different aspects of two ways of processing Chinese characters based on previous research.

Table 1. Differences between holistic processing and analytic processing

	Holistic processing	Analytic processing
Unit	An individual character	a component of a character
Reader	Novice reader	Expert reader
Version of Chinese	Traditional Chinese	Simplified Chinese
Frequency effect	Frequency of character occurrence	Frequency of radical occurrence
Other effect	-	Familiarity of radical; regularity of radical[a]
Note.	[a]The regularity of radical is called the 'transparency of radical' in the current study. Its definition can be found in the next section.	

However, the study of Tso et al. (2012) did not focus on the usual character elements of previous literature. Using the centre of a character as the division point, they masked either left, right, top or bottom of the character. Then the participants were asked to attend to the unmasked half (either top or bottom for top-bottom characters, or left or right for left-right characters) of each character. In this case, the unit for readers to decompose and recombine in order to process Chinese characters is a perceptually defined component rather than linguistically defined concept such as radical. In reality, all parts of a character are processed in parallel. However, it is possible that readers learn to allocate attention to certain regions of a character for more efficient recognition. This region of a character is more vaguely defined than a radical. In addition, previous research suggested that the spatial density of Chinese characters – the space between different strokes at horizontal and vertical levels – can be used to account for effects of character structure and stroke number in visual recognition of Chinese characters ((Zeng et al., 2001).[1] Blurring rather than masking part of a character may encourage the reading of Chinese characters in a natural way. Based on this assumption, the current study chose to blur regions of characters in order to explore those that are more crucial to its recognition.

Research Method

As presented in the above sections, several proposals have been made regarding the way in which Chinese characters are processed and many studies have investigated the write-to-read effect in Chinese language learning. However, there is a scarcity of research examining what regions of a character crucial for its recognition. For this reason, the current study concentrated specifically on this issue. Instead of following the orthographic units involved in the previous literature, the current study modified the design of Tso et al. (2012) and blurred rather than masked different regions of characters. A gaussian blurring technique was employed to obscure parts of characters. The direction of maximum blur could be from left to right, right to left, top to bottom, or bottom to top of the character. Moreover, the blurring technique did not privilege any particular a priori approach to character segmentation.

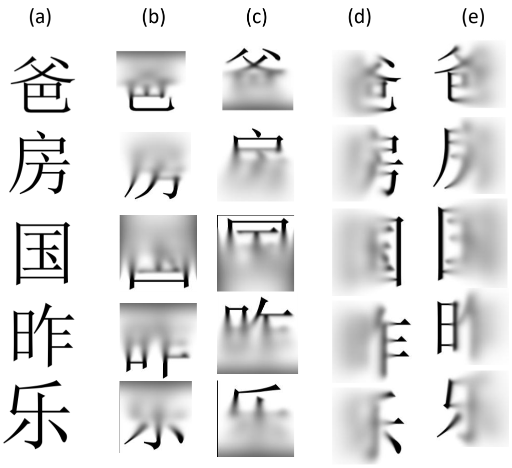

Figure 1. Example of five partially obscured characters (top-bottom structure; half-surround structure; surround structure; left-right structure; integral structure): (a) the original character; gaussian blurred versions from (b) top to bottom; (c) bottom to top; (d) left to right; and (e) right to left.

---

[1] Although the English translation of ‘hanzi tongtouxing’ can be ‘the space transparency of Chinese characters’ in the study of Zeng et al. (2001), it is translated into ‘the spatial density of Chinese characters’ in order to clearly differentiate it from other radical transparency in the current study.

Figure 1 above displays the four versions of a character following blurring from different directions. The blurring effect decreases towards the centre of character in each case. Note that the original character is shown here as reference, since only the blurred versions were shown to participants. In addition, participants only saw one blurred version of each character.

Chinese Characters Used in the Present Study

The current study selected 102 characters with an average frequency of 0.5 million per 300 million (Jun, 2010; see Appendix 1). The study reported here is the first phase of a project exploring effective ways of recognising Chinese characters among native and non-native Chinese speakers. Thus, the characters chosen for the study are those included in both the HSK Level 1 vocabulary list and EBCL (European Benchmarking Chinese Language) Proposed List of A1 characters. Those characters were also checked against the vocabulary list of New Practical Chinese Reader Volume 1 (Liu, 2015) to ensure that they were also included. New Practical Chinese Reader has been estimated to be used in approximately 2,000 universities around the world (Tinnefeld-Yeh, 2014) and has received the Award for Outstanding International Chinese Language Teaching Materials. In other words, all the characters chosen for the present study are from three different sources proposed for CFL beginners. They are required to be mastered by CFL beginner learners and consequently can be used in the next phase of the project. By examining the recognition of these characters by skilled readers, we hope in the future to compare the findings of the current study with character recognition among CFL learners. The 102 characters can be divided into five types. Around 29% have left-right structure, 18.6% top-bottom structure, 33% integral structure, 16.6% half-surround structure and 1% surround.

The compound characters involved in the research have been further annotated regarding their character structure, visual complexity, radical transparency and radical frequency:

(1) Chinese characters are generally grouped into integral and compound structures. The latter one can be further divided into four types according to the position of radicals in a character: top-bottom (including top-middle-bottom), left-right (including left-middle-right), half-surround and surround structures. Based on this classification, all characters used in the research can therefore be categorised and coded into five structures: integral, top-bottom, left-right, half-surround and surround.

(2) Radical transparency refers to the extent a semantic or phonetic radical can provide reliable clues to a character's meaning and sound, respectively.[2] High radical transparency might contribute to the success in recognising characters. The more reliable information a phonetic radical can offer to the sound of a character in which it appears, the more likely readers can infer the pronunciation of the character. However, if such a phonetic radical is obscured, it would immediately cause recognition difficulty for the whole character. Similarly, a semantic radical might also offer a reliable clue to the meaning of a character. Therefore, a semantic radical that is not blurred might assist readers to recognise the character through triggering its meaning. Following the transparency classification in the study of Shu et al. (2003), the current study employed three levels of radical transparency: transparent, semi-transparent and opaque.[3] Characters that cannot be categorised in this way are grouped under the label of 'unclassified'.

(3) Visual complexity of a character is usually measured by the number of strokes within the character. The more complex a character is, the more often readers need to employ analytic skills to process it in order to recognise the character and consequently less blurring would positively impact on successful recognition. On the contrary, a character having few strokes might need to be processed holistically. In this case, the partially obscured condition might impact more heavily on the reading of characters with fewer strokes than those with more strokes. The number of strokes of each character was based on information provided in Rui (2000).

(4) Radical frequency refers to the frequency of occurrence of a radical being used to form a compound character. In other words, radical frequency indicates how often this radical is involved in various characters and is based on the study of Han (1993). The position of a radical in a character will be significantly correlated with the accuracy of recognising characters. If a radical is positioned in a place in a character where it frequently appears, the blurring of this radical is unlikely to affect the recognition of the character.

Participants

---

[2] Previous research may term it differently. For example, the term 'transparency' is specifically for phonetic radicals and 'regularity' is used for semantic radicals (Shu et al. 2003).

[3] Shu et al. (2003: 34) used 'phonetic regularity' to refer to the extent a phonetic radical is reliably linked with the character's pronunciation. Consequently, 'regular', 'semiregular' and 'irregular' were used, which is equivalent to the three tiers outlined here (i.e. transparent, semi-transparent and opaque).

The participants in the current study were native Chinese speakers who were already fluent in the language. The research collected data from 275 undergraduate students from eight different universities across China. Of the 275, 173 (67%) completed the experiment. Among this latter group of participants, 68 were male and 105 female with an average age of 21.1 (median = 21 years). Almost 20% of them were from ethnic minorities whose first language may not have been standard Mandarin Chinese and so might speak Chinese with their own accents. For example, ethnic minority students from Southwest may speak Southwest Mandarin which features the ancient entering tone being pronounced using the rising tone (Zhang and Yang, 2017). However, these students will have needed to achieve native-level proficiency in Mandarin Chinese, since they will have had to use it to pass either the university entrance exam or the MHK (Minzu Hanyu Shuiping Kaoshi, commonly referred to as the Test of Chinese Proficiency for Minorities; Ministry of Education of the PRC, 2010, 2015). Moreover, statistical tests confirmed that the participant's ethnic background did not have a significant effect on the correct identification of characters ($\chi^2 = 7.5$, df = 68, p = 1.0).

Experimental Procedure

The research was conducted online over a month in 2016 through a web-based survey (PsyToolkit; Stoet, 2010). The survey was divided into two parts. The first collected demographic information from each participant, including their age, gender, major, year of study, university, place of birth and ethnicity. The novelty of a web-based design allows the collection of data from a relatively large and diverse sample of readers in terms of geographical location and ethnic background. A snapshot of previous studies of character recognition shows that the majority recruited small and relatively homogeneous samples, such as Anderson et al. (2013; N=50), Chen and Yeh (2015; N=48), Hsiao and Cottrell (2009; N=32), Liu and Hsiao (2012; N=48), Taft and Zhu (1997; N=20) and Tao and Healy (2002; N=40). Even for research with larger samples, all participants of Tso et al. (2012; N=217) came from one elementary school in Hong Kong and those of Shu and Anderson (1997; N=220) were based in one elementary school in Beijing. The web-based design of the current study allowed us to collect data from a sample with significantly greater diversity: 276 students from eight universities across China.

In the second part of survey, each of the 102 experimental characters was presented until a response was given and the button 'next' was clicked by the participant. The participants, therefore, could have as much time as they wished to recognise each character. They were instructed to give written response in pinyin, but without any tone indicator. This instruction was purely to save time and reduce any technical requirement when the participants were filling out the survey online, since specific software is usually needed to input proper pinyin characters with appropriate tone indicators. Moreover, it was felt that allowing participants to just type letters without diacritics reduced their task demands and allowed them concentrate on recognition. Although 16 homophones were included in the current study, none had similar orthography and consequently did not confound the scoring. For example, all five characters (i.e. 是, 识, 师, 时, 事) have exactly the same onset and finals and they only differ in the tones. It is extremely unlikely that a participant put down a correct pinyin for a blurred character while he/she meant to indicate another character. Besides, since the web-based study chose to blur regions of characters rather than blurring linguistically defined components (e.g. a phonetic or semantic radical), it did not focus on response accuracy cued by semantic or even sentential context. Instead, the responses collected in the current study were simply used to demonstrate whether or not a character had been correctly recognised. Therefore, the mode of response or the length of time that the participants viewed each blurred character is unlikely to have had a significant impact on response accuracy.

The experiment was piloted with four participants with similar demographic backgrounds to those in the subsequent data collection. Their attention was drawn to the homophones after the pilot. None of them suggest any confusion or misidentification due to the homophones. A few instructions were rephrased in order to be as clear as possible. Details of birthplace and ethnicity were also collected based on feedback following the pilot. During the pilot, five characters were discovered to be polyphones, so different spellings were considered as equivalent identifications for these.

Participant responses were scored for correctness. Most of the scoring was done automatically and a small percentage needed to be done by hand because of formatting issues with some participants’ input.

Results

The following sections will present the results relating to the effects on recognition performance of blurring direction, character structure type, radical transparency, character complexity, and radical frequency. Due to the relatively small number of surround and half-surround characters (17 and 2, respectively), the analyses below will focus on three character structures: left-right, top-bottom and integral characters. All of the analyses below used linear mixed models (Baayen et al., 2008) implemented on the R language platform (R Core Team, 2013).

Overall Recognition

As can be seen in Table 2, the most significant impact on correct recognition resulted from top-to-bottom blurring (z=-11.9, p<0.0001), with the next being bottom-to-top blurring (z=-6.32, p<0.0001).

Table 2

Recognition Accuracy as a Function of Blurring Mode and Character Structure

Item	**estimate**	**SE estimate**	**z**	**p**
intercept	3.80	0.31	12.15	<0.0001
orient: top-mean	-2.34	0.20	-11.90	<0.0001
orient: bot-mean	-1.26	0.20	-6.33	<0.0001
orient: left-mean	1.61	0.34	4.74	<0.0001
struct: LR-Int	-1.59	0.37	-4.27	<0.0001
struct: TB-Int	-0.76	0.42	-1.82	0.07
top-mean x LR-Int	1.99	0.22	9.12	<0.0001
top-mean x TB-Int	0.58	0.24	2.39	0.02
left-mean x LR-Int	-1.62	0.35	-4.58	<0.0001
left-mean x TB-Int	-0.58	0.41	-1.44	0.15
bot-mean x LR-Int	1.61	0.23	7.10	<0.0001
bot-mean x TB-Int	0.72	0.25	2.82	0.005

Note. [a]This table provides the results of the linear mixed model of the form:

correct ~ orient * struct + (1 | character) + (1 | participant),

where *correct* indicates whether the character was recognised (1 or 0), *orient* indicates the direction of blurring and *struct* is the character structure. The latter two factors are contrast coded such that the reference category for *orient* is the overall mean (e.g., top-mean) and the reference for *struct* is the integral structure (e.g., LR-Int). The * notation is used to indicate that interaction terms between these two factors are to be included in the model. The final two terms of the model indicate the random effects to be included. In this case, we include random intercepts for participant and character. The table comprises the model terms, the estimated effect, its standard error, the z-score for the estimate and its associated probability.

The other two blurring conditions, left-to-right and right-to-left, did not significantly affect the recognition. Further data analyses below indicated the importance of the top part of a character in its visual recognition.

Effects of Character Structure

Using the integral characters as the reference type, recognition of the left-right structures was most negatively impacted by all forms of blurring (z=-4.26, p<0.0001). However, when the results were divided further, we can find that: for top-to-bottom blurring and bottom-to-top blurring, recognition of left-right structures was less affected than other structures, hence the higher correct recognition rates apparent in Figure 2. In the case of left-to-right blurring and right-to-left blurring, left-right structures were more affected, which makes sense given the alignment between blurring direction and this particular character structure. In this case, the effect of blurring direction on character recognition is correlated to the character structures, which is consistent with the effect of character structure on the recognition due to the spatial density of Chinese characters in the study of Zeng et al. (2001). Blurring from left or right reduces the spatial density of a character and consequently impacted significantly on the recognition of characters with a left-right structure. Similarly, blurring from top or bottom affected most on the recognition of characters with a top-bottom structure due to the decrease of the spatial density.

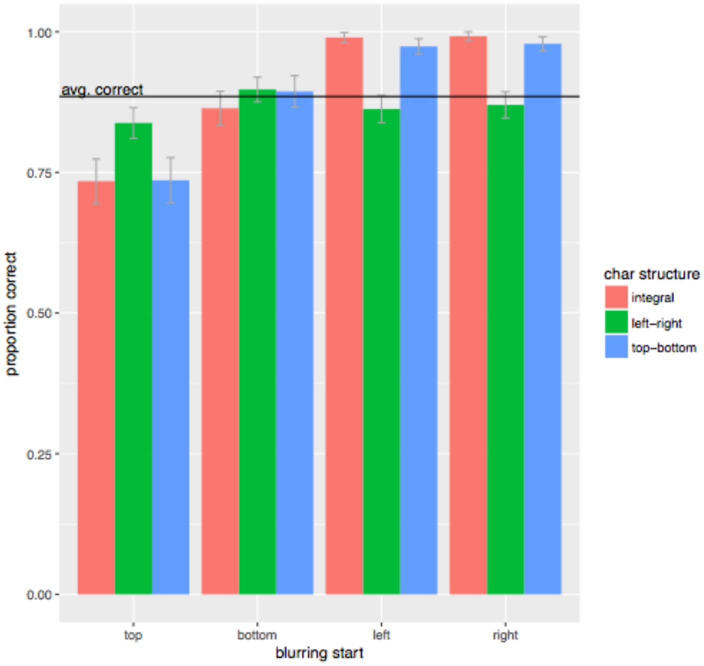

Figure 2. Recognition accuracy as a function of blurring direction and character type.

Note. ^{a}The horizontal line indicates the overall rate of correct responding, which is used as the reference value for comparing the effects of different blurring directions. The integral character is used as the reference category for testing the effects of different character types is.

With regard to the character structures themselves, overall, left-right structures were harder to identify than top-bottom structures, irrespective of blurring direction. Note that this result uses the integral characters as a baseline because of its simpler structure. Interestingly, the complexity of the characters as measured by their number of strokes did not have a statistically significant effect on recognition, nor did this factor interact significantly with either character structure or direction of blurring.

Effects of Radical Transparency

The hypothesis underlying this analysis is that the more transparent the radical, the more accurate the recognition of the character it comprises. In our analysis, this hypothesis was borne out for phonetic radicals ($z=2.62$; $p=0.009$), rather than semantic radicals (see Table 3). There was an overall significant linear effect on the correct recognition due to the transparency of the phonetic radical. From Figure 3 it can be seen that the linear effect is most pronounced for the bottom-to-top and right-to-left forms of blurring. This is probably due to the fact that a phonetic radical is usually positioned on the right or at the bottom of a semantic-phonetic compound character. Besides, the current study asked participants to write down the pronunciation of the character, which is also suggested by a phonetic radical, once they had recognised them. Therefore, when the blurring starts from where a phonetic radical is usually positioned, it would significantly affect the recognition of the sound of this character and consequently the character itself in the current study. In this case, the radical transparency became prominent in assisting in the character recognition.

Table 3

Recognition Accuracy as a Function of Radical Transparency

Item	estimate	SE estimate	z	p	
(Intercept)	2.578	0.193	13.33	<0.0001	***
phonetic linear (phon.L)	0.794	0.300	2.64	0.008	**
phonetic quadratic (phon.Q)	0.201	0.326	0.62	0.538	
semantic linear (sem.L)	0.672	0.324	2.07	0.038	
semantic quadratic (sem.Q)	0.104	0.303	0.34	0.731	
orient.top-mean	-0.588	0.129	-4.55	<0.0001	***
orient.bot-mean	0.495	0.169	2.93	0.003	***
orient.left-mean	0.052	0.154	0.34	0.734	
phon.L:orient.top-mean	-0.798	0.208	-3.83	<0.001	***
sem.Q:phon.L: orient.top-mean	-1.200	0.348	-3.45	<0.001	***
sem.Q:phon.Q: orient.left-mean	-1.285	0.426	-3.018	<0.01	**

Note. [a]The table represents the results of fitting the following model to the correct response data:

*correct ~ orient * phon * sem + (1 | character) + (1 | participant),*

where *orient* encodes the direction of blurring, using the overall mean as a reference value, *phon* is the level of phonetic transparency (low, medium, high) encoded as an ordered numerical value from 1 to 3 indicating increasing levels of transparency, and *sem* is also encoded as an ordered variable on a three-point scale. Interactions between all three of these fixed effects are also included in the model as well as random effects for both participant and character. Since *phon* and *sem* are ordered variables, we can test for their linear and quadratic effects (i.e., phon.L, phon.Q). The table below provides estimates of the effect of each equation term on correct responding, the standard error of the estimate, and a z-value with its associated probability. Note that only estimates for significant interactions are displayed in the table. Non-significant ones are omitted for the sake of succinctness and clarity.

There were also significant interactions between phonetic radical transparency and two of the blurring conditions: top-to-bottom blurring (z=-3.92, p<0.001 – an interaction with the linear term) and left-to-right blurring (z=4.20, p<0.001 – an interaction with the quadratic term). This is likely to be related again with the usual position of a phonetic radical being the right and bottom of a character. When blurring from the top or from the left, a phonetic radical is possibly visible and so would contribute to the recognition of the character. Therefore, a possible explanation for the interaction between blurring direction and phonetic radical transparency might be because the participant was asked to type the pinyin of the character, which made them more dependent on figuring out the sound of the character and consequently the identity of the phonetic radical.

Figure 3 below shows that the correct recognition rates for all three levels of phonetic radical transparency were below average. In other words, phonetic radical transparency helps most of the conditions except when blurring is from the top. As shown in the result of the overall recognition, the correct recognition was affected most by the top-to-bottom blurring. Even with the assistance from phonetic radical transparency, the importance of the top part of characters for recognition still exists.

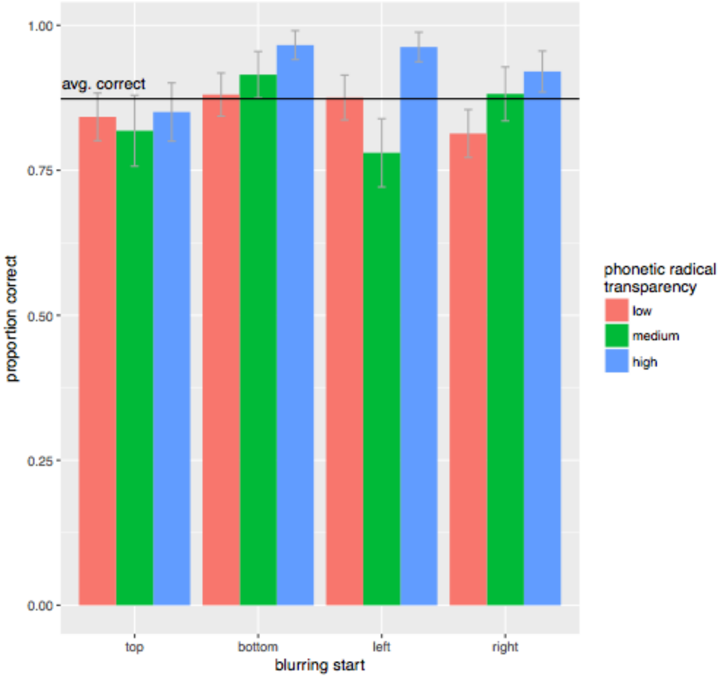

Figure 3. The recognition accuracy of blurred characters as a function of direction of blurring and phonetic radical transparency

Note. ^{a}Transparency is coded as an ordered value from low (1) to high (3) levels of transparency. The horizontal line indicates the overall rate of correct responding, which is used as the reference value for comparing the effects of different blurring directions.

Another interesting result is that there is a significant quadratic interaction with both phonetic and semantic transparency for blurring from left (see Figure 3 and 4), which suggests that the effect may not be dependent on the specific type of radical.

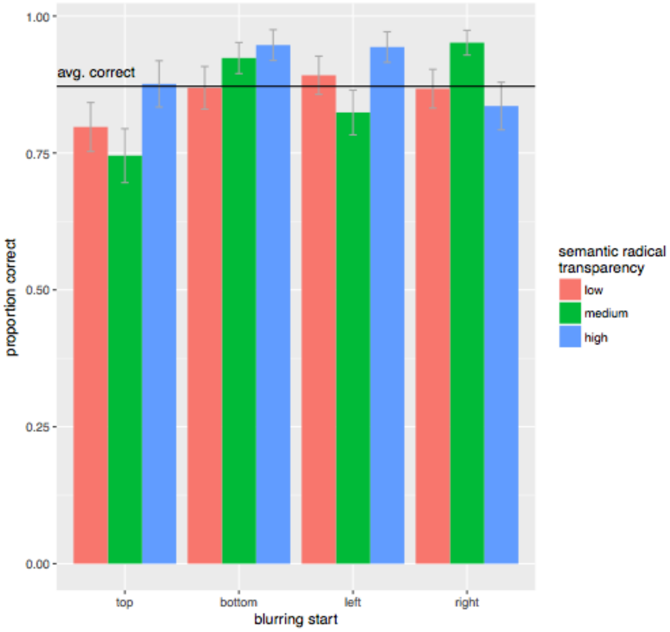

Figure 4. The recognition accuracy of blurred characters as a function of direction of blurring and semantic radical transparency

Note. ^{a}Transparency is coded as an ordered value from low (1) to high (3) levels of transparency. The horizontal line indicates the overall rate of correct responding, which is used as the reference value for comparing the effects of different blurring directions.

Effects of Visual Complexity

The following table shows that there are significant interactions between visual complexity as measured by number of strokes and two of the blurring conditions, (i.e., bottom to top and left to right).

Table 4

Recognition Accuracy as a Function of Blurring Direction and Character Complexity

Item	**estimate**	**SE estimate**	**z**	**p**	
(Intercept)	3.220	0.406	7.923	<0.0001	***
complexity	-0.087	0.057	-1.524	0.127	
orient.top-mean	-1.526	0.200	-7.625	<0.0001	***
orient.bot-mean	-1.562	0.222	-7.027	<0.0001	***
orient.left-mean	1.185	0.273	4.342	<0.0001	***
complexity: orient.top-mean	0.067	0.028	2.424	0.015	*
complexity: orient.bot-mean	0.224	0.032	6.920	<0.0001	***
complexity: orient.left-mean	-0.101	0.036	-2.787	0.005	**

Note. [a]The table represents the results of fitting the following model to the correct response data:

*correct ~ orient * complexity + (1 | character) + (1 | participant),*

where *orient* encodes the direction of blurring, using the overall mean as a reference value, and *complexity* uses the number of strokes as a complexity measure. Interactions between complexity and blurring direction are also included in the model as well as random effects for both participant and character. The table below provides estimates of the effect of each equation term on correct responding, the standard error of the estimate, and a z-value with its associated probability.

In Figure 5, blurring from the bottom (green line) varies as a function of complexity and differs significantly from the average pattern of the other blurring conditions. The more complex a character is, the more information that a reader could detect from the visible part of a character and consequently the less impact from blurring from the bottom on accurate recognition. In other words, complexity helps you overcome the effect of blurring from the bottom. However, the recognition accuracy and character complexity do not have the same linear effect in all blurring conditions.

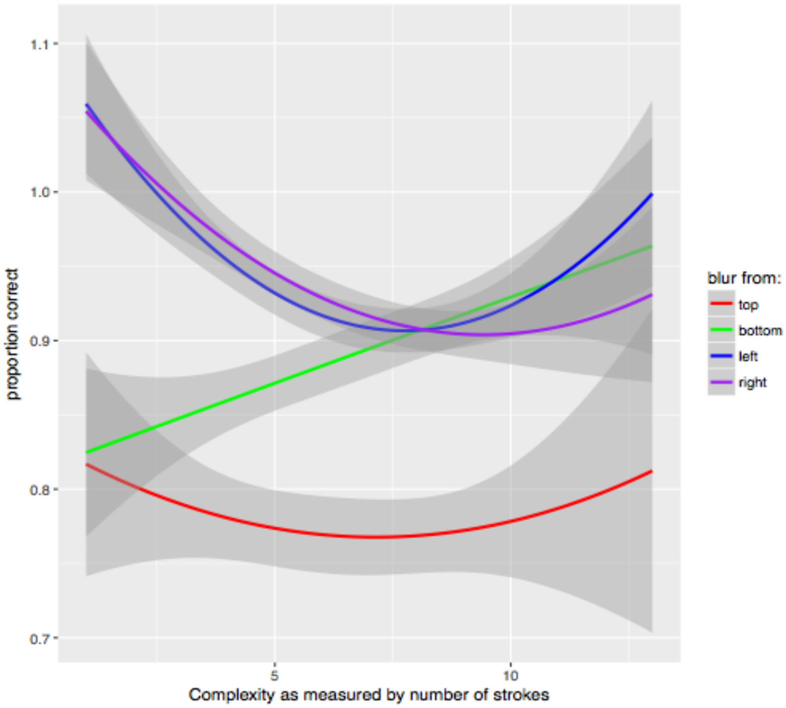

Figure 5. Recognition accuracy of blurred characters as a function of direction of blurring (from top, bottom, left, and right) and character complexity as measured by number of strokes.

Figure 5 also shows a general U-shape pattern for three other blurring conditions. When there are fewer strokes, correct recognition is quite good. As the current study used characters with a high frequency of occurrence in modern Chinese language, characters with fewer than five strokes would be frequently encountered in the reader's day-to-day experience. Familiarity with these simple characters might contribute to their correct recognition. Importantly, this kind of character has better spatial transparency and consequently more likely to be recognisable when blurred. On the other hand, a complex character, such as those with more than 10 strokes, could provide more information to the readers and consequently still assist in the visual recognition through the part that was not blurred. Therefore, complexity can also contribute to the correct recognition under some circumstance. This effect of stroke number is evident in previous research (e.g. Zeng et al., 2001) which shows that characters with 4-6 strokes were relatively easier to be identified than those with 9-10 strokes.

There seems to be a dip of the recognition only when a character has an intermediate level of complexity (see Figure 5). It would be intriguing to conduct further research in order to scrutinise whether the correct character recognition indeed depends on two levels of character complexity, i.e. one with stroke number 1-5 or above 10, the other with stroke number 5-10. Interestingly, although the blurring from top also demonstrates a U-shape in Figure 5 (red line), the correct recognition in this condition is significantly lower than in other conditions. So the top part of a character seems to be crucial for recognition.

Effects of Radical Frequency

While we controlled for the frequency of occurrence of each of the 102 characters selected for the study (Jun, 2010), it is still possible that the frequency of individual radicals might have some effect on task performance. Therefore, we analysed whether radical frequency (Han, 1993) might have an impact on recognition. Table 5 below gives the results of the analysis. The frequency of the semantic radical had a marginally significant effect on recognition performance ($z=-1.985$, $p<0.05$) and there was a highly significant interaction between direction of blur and semantic radical frequency ($z=3.545$, $p<0.001$).

Table 5

Recognition Accuracy as a Function of Radical Frequency and Blurring Direction

Item	**estimate**	**SE estimate**	**z**	**p**	
(Intercept)	3.385	0.503	6.735	<0.0001	***
log(phonF)	0.014	0.090	1.150	0.881	
orient.top-mean	-1.937	0.389	-4.985	<0.0001	***
orient.bot-mean	-0.291	0.451	-0.645	0.519	
orient.left-mean	1.423	0.627	2.270	0.023	*
log(semF)	-0.328	0.165	-1.985	0.047	*
log(phonF): orient.top-mean	-0.085	0.069	-1.236	0.216	
log(phonF): orient.bot-mean	-0.074	0.078	-0.950	0.342	
log(phonF): orient.left-mean	-0.166	0.072	-2.298	0.22	
orient.top-mean:log(semF)	0.445	0.126	3.545	0.0004	***
orient.bot-mean:log(semF)	0.181	0.143	1.266	0.206	
orient.left-mean:log(semF)	-0.278	0.184	-1.513	0.130	

Note. [a]The table represents the results of fitting the following model to the correct response data:

correct ~ orient * log(phonF) + orient * log(semF) + (1 | character) + (1 | participant),

where _orient_ encodes the direction of blurring, using the overall mean as a reference value, _phonF_ is the frequency of occurrence of the phonetic radical, and _semF_ is the corresponding value for the semantic radical. Interactions between radical frequency and blurring direction are also included in the model as well as random effects for both participant and character. The table below provides estimates of the effect of each equation term on correct responding, the standard error of the estimate, and a z-value with its associated probability.

The nature of the interaction is shown in Figure 6, below. As can be seen, the overall trend is for greater frequency of the semantic radical to be associated with poorer recognition performance, this is likely because the more frequent the radical the less diagnostic it is of a particular character. The one exception to this pattern is the top-down blurring condition, where overall recognition performance is impacted more dramatically by the top-down condition. Again, this result indicates the importance of the top part of a character in the recognition accuracy.

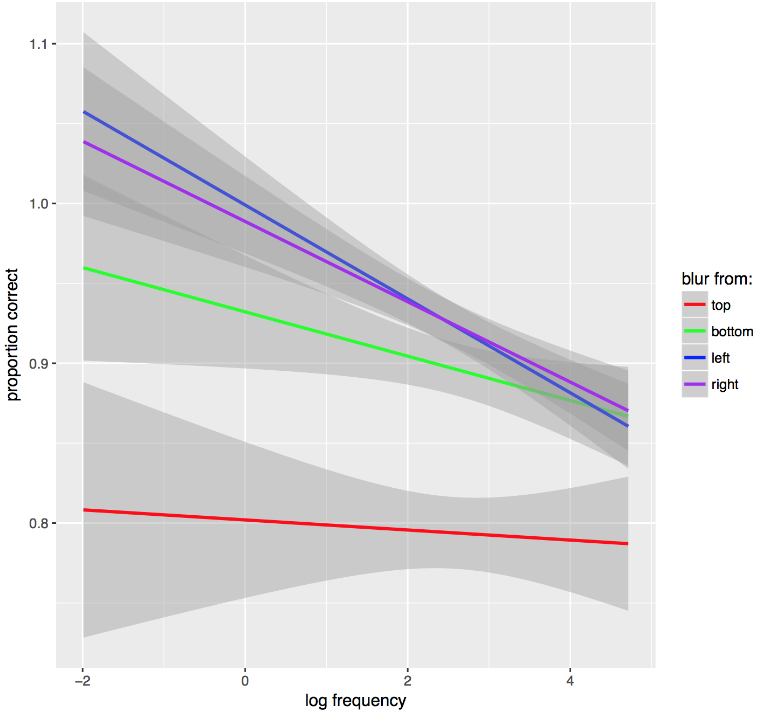

Figure 6. Recognition accuracy of blurred characters as a function of direction of blurring (from top, bottom, left, and right) and log frequency of the semantic radical.

Discussion

There are three main findings from the current study. First, the top and bottom parts of a character seem to play a more important role in character recognition since recognition performance is impacted more dramatically overall by the top-to-bottom blurring condition. In other words, the top and bottom regions of a character are likely to be crucial for its recognition. This may be because in normal Chinese reading, skilled readers are used to the visibility of the left and right parts of characters being reduced by lateral masking from adjacent characters. In contrast, there is usually good spacing between lines in a text, so the top and bottom of each line – a few characters comprises a line of text – are generally clear and visible to readers. Consequently, the information at the top and bottom of characters will be the most visible parts of the character in continuous reading and consequently be more valuable in character recognition.

The second finding is that phonetic radicals seem to be more sensitive to blurring than semantic ones. This is reflected in the result that the transparency of phonetic radical significantly influences the character recognition in comparison with that of semantic radical. This effect is more prominent in the conditions of characters being blurred from right and from the bottom, two positions where a phonetic radical is usually located. Since the present study required the participants to type the pinyin of the character, it may have led to more emphasis on identifying the phonetic radicals, which offer clues to pronunciation and thus to the character's pinyin form. However, the source of the significant quadratic transparency effect found for both semantic and phonetic radicals in the blurred-from-left condition is unclear and will require a more detailed analysis in an eye tracking study. As mentioned earlier, the response in pinyin was chosen in order to balance the technical difficulties and task demands, especially in a web-based study like this. Despite possible effects of pinyin response being minimised in the current study, it might be worthwhile for future research to ask participants to type characters rather than pinyin, in order to examine if a different design leads to any variations in the findings.

The third finding relates to the debate concerning holistic or analytic nature of character processing. The character complexity effects seem to provide some support for the holistic processing of complex characters. The reason might be that a complex character may offer more information to readers even when part of it is obscured. Therefore, proficient readers can employ their holistic skills to recognize it. This finding is in line with the result of Liu and Hsiao (2012), which suggests that readers of the more visually complex traditional characters tended to process them holistically. On the other hand, the semantic radical frequency effect in Table 4 shows evidence of a more analytic processing approach. This might be because the survey nature of the study allowed participants to take their time to reflect before giving their responses. In addition, the current study was built upon partially blurred characters. This design forced participants to employ their analytical skills to some extent as the whole character was not available. Further research, involving an eye tracking paradigm are needed to explore the factors involved in readers' immediate processing of Chinese characters in depth.

Conclusion and Future Research

The current study employed partially blurred characters as experimental material, which has several advantages over the partially masked characters used in Tso et al. (2012). Our technique does not assume *a priori* that the components for character recognition are linguistic units such as radicals. Moreover, blurring rather than masking a part of a character can encourage readers to identify characters using their low spatial frequency features, thus allowing the study to take place under more realistic conditions. The overall finding of the current study suggests that the top of a character seems to be crucial in character recognition, while other factors such as type and frequency of radical also impact on recognition accuracy. Further studies are planned, particularly involving the use of an eye-tracker to focus on the fine temporal details of the processing of Chinese characters in reading and to further explore some of the phenomena identified in this preliminary study.

References

Anderson RC, Ku Y-M, Li W, et al. (2013) Learning to see the patterns in Chinese characters. *Scientific Studies of Reading* 17(1): 41–56. DOI: 10.1080/10888438.2012.689789.

Baayen RH, Davidson DJ and Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). Special Issue: Emerging Data Analysis: 390–412. DOI: 10.1016/j.jml.2007.12.005.

Cao F, Rickles B, Vu M, et al. (2013) Early stage visual-orthographic processes predict long-term retention of word form and meaning: A visual encoding training study. *Journal of Neurolinguistics* 26(4): 440–461. DOI: 10.1016/j.jneuroling.2013.01.003.

Cao F, Vu M, Lung Chan DH, et al. (2013) Writing affects the brain network of reading in Chinese: A functional magnetic resonance imaging study. *Human Brain Mapping* 34(7): 1670–1684. DOI: 10.1002/hbm.22017.

Chang L-Y, Xu Y, Perfetti CA, et al. (2014) Supporting orthographic learning at the beginning stage of learning to read Chinese as a second language. *International Journal of Disability, Development and Education* 61(3): 288–305.

Chen H-C and Shu H (2001) Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin & Review* 8(3): 511–518. DOI: 10.3758/BF03196186.

Chen H-C, Hsu C-C, Chang L-Y, et al. (2013) Using a radical-driven character e-learning platform to increase learner knowledge of Chinese characters. *Language Learning & Technology* 17(1): 89–106.

Chen Y-C and Yeh S-L (2015) Binding radicals in Chinese character recognition: Evidence from repetition blindness. *Journal of Memory and Language* 78: 47–63. DOI: 10.1016/j.jml.2014.10.002.

DeFrancis J (1984) *The Chinese Language: Fact and Fantasy*. Honolulu: University of Hawaii Press. Available at: https://uhpress.hawaii.edu/title/the-chinese-language-fact-and-fantasy/ (accessed 26 March 2019).

Ding G, Peng D and Taft M (2004) The nature of the mental representation of radicals in Chinese: A priming study. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 30(2): 530–539. DOI: 10.1037/0278-7393.30.2.530.

Grenfell M and Harris V (2015) Memorisation strategies and the adolescent learner of Mandarin Chinese as a foreign language. *Linguistics and Education* 31: 1–13. DOI: 10.1016/j.linged.2015.04.002.

Guan CQ, Liu Y, Chan DHL, et al. (2011) Writing strengthens orthography and alphabetic-coding strengthens phonology in learning to read Chinese. *Journal of Educational Psychology* 103(3): 509–522. DOI: 10.1037/a0023730.

Han B (1993) *Frequency effect of character components and radical combination on the recognition of Chinese characters [Hanzi shibie zhong bujian he bujian zuhe de pinlu xiaoying]*. Ph.D. Thesis. Institute of Psychology, Academia Sinica, Beijing.

Hsiao JH and Cottrell GW (2009) Not all visual expertise is holistic, but it may be leftist: The case of Chinese character recognition. *Psychological Science* 20(4): 455–463. DOI: 10.1111/j.1467-9280.2009.02315.x.

Jun D (2010) Chinese Text Computing. Available at: http://lingua.mtsu.edu/chinese-computing/ (accessed 26 March 2019).

Kuo L-J, Li Y, Sadoski M, et al. (2014) Acquisition of Chinese characters: The effects of character properties and individual differences among learners. *Contemporary Educational Psychology* 39(4): 287–300. DOI: 10.1016/j.cedpsych.2014.07.001.

Lam HC (2011) A critical analysis of the various ways of teaching Chinese characters. *Electronic Journal of Foreign Language Teaching* 8: 57–70.

Li H, Zhang J, Ehri L, et al. (2016) The role of orthography in oral vocabulary learning in Chinese children. *Reading and Writing* 29(7): 1363–1381. DOI: 10.1007/s11145-016-9641-0.

Li W, Anderson RC, Nagy W, et al. (2002) Facets of metalinguistic awareness that contribute to Chinese literacy. In: Li WL, Gaffney JS, and Packard JL (eds)

Chinese Children's Reading Acquisition. Dordreeht, the Netherlands: Kluwer Academic Publishers, pp. 87–106.

Liu T and Hsiao J (2012) The perception of simplified and traditional Chinese characters in the eye of simplified and traditional Chinese readers. *Journal of Vision* 12(9): 533–533. DOI: 10.1167/12.9.533.

Liu X (2015) *New Practical Chinese Reader Vol.1 – Textbook*. 3rd Edition. Beijing Language & Culture University Press, China.

Ministry of Education of the PRC (2010) Notice of the Ministry of Education of PRC regarding piloting MSH in relevant provinces [Jiaoyubu guanyu zai youguan shengqu shixing zhongguo shaoshu minzu hanyu shuiping dengji kaoshi de tongzhi]. Available at: http://www.moe.gov.cn/s78/A18/yys_left/moe_805/s3139/201001/t20100127_78874.html (accessed 18 May 2020).

Ministry of Education of the PRC (2015) *2015 Chinese Education Yearbook*. Beijing: People's Education Press.

Mo C, Yu M, Seger C, et al. (2015) Holistic neural coding of Chinese character forms in bilateral ventral visual system – ScienceDirect. *Brain and Language* 141: 28–34. DOI: https://doi.org/10.1016/j.bandl.2014.11.008.

Perfetti CA and Tan LH (1998) The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24(1): 101–118.

Perfetti CA, Liu Y and Tan LH (2005) The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review* 112(1): 43–59.

R Core Team (2013) R: The R Project for Statistical Computing. Available at: https://www.r-project.org/ (accessed 2 April 2019).

Rui S (2000) *Dictionary for Standard Stroke Orders of Common Chinese Characters [Changyong Hanzi Bishun Guifan Zidian]*. Beijing: Zhongguo Shuji Chubanshe [China Books Publishing House].

Shen HH and Ke C (2007) Radical awareness and word acquisition among nonnative learners of Chinese. *The Modern Language Journal* 91(1): 97–111. DOI: 10.1111/j.1540-4781.2007.00511.x.

Shu H and Anderson RC (1997) Role of radical awareness in the character and word acquisition of Chinese children. *Reading Research Quarterly* 32(1): 78–89.

Shu H, Chen X, Anderson RC, et al. (2003) Properties of school Chinese: Implications for learning to read. *Child Development* 74(1): 27–47. DOI: 10.1111/1467-

8624.00519.

Stoet G (2010) PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42(4): 1096–1104. DOI: 10.3758/BRM.42.4.1096.

Taft M (2006) Processing of characters by native Chinese readers. In: Li P, Tan LH, Bates E, et al. (eds) *Handbook of East Asian Psycholinguistics*. Cambridge: Cambridge University Press, pp. 237–294.

Taft M and Zhu XI (1997) Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*: 761–775.

Tan LH, Spinks JA, Eden GF, et al. (2005) Reading depends on writing, in Chinese. *Proceedings of the National Academy of Sciences* 102(24): 8781–8785. DOI: 10.1073/pnas.0503523102.

Tao L and Healy AF (2002) The unitization effect in reading Chinese and English text. *Scientific Studies of Reading* 6(2): 167–197. DOI: 10.1207/S1532799XSSR0602_03.

Tinnefeld-Yeh Y-L (2014) Taking a leaf from each other’s book: A comparative study of vocabulary in Chinese textbooks from China, the UK and Germany. In: Tinnefeld T (ed.) *Fremdsprachenunterricht Im Spannungsfeld Zwischen Sprachwissen Und Sprachkönnen*. Saarbrücken: htw saar, pp. 239–258.

Tong X and McBride C (2014) Chinese Children’s Statistical Learning of Orthographic Regularities: Positional Constraints and Character Structure: Scientific Studies of Reading: Vol 18, No 4. *Scientific Studies of Reading* 18(4): 291–308. DOI: 10.1080/10888438.2014.884098.

Tso RVY, Au TK -f. and Hsiao JH -w. (2012) Writing facilitates learning to read in Chinese through reduction of holistic processing: A developmental study. *Journal of Vision* 12(9): 530–530. DOI: 10.1167/12.9.530.

Wang M, Perfetti CA and Liu Y (2003) Alphabetic readers quickly acquire orthographic structure in learning to read Chinese. *Scientific Studies of Reading* 7(2): 183–208. DOI: 10.1207/S1532799XSSR0702_4.

Xu Y, Chang L-Y, Zhang J, et al. (2013) Reading, writing, and animation in character learning in Chinese as a foreign language. *Foreign Language Annals* 46(3): 423–444. DOI: 10.1111/flan.12040.

Xu Y, Chang L-Y and Perfetti CA (2014) The effect of radical-based grouping in character learning in Chinese as a foreign language. *The Modern Language Journal* 98(3): 773–793. DOI: https://doi.org/10.1111/modl.12122.

Zeng J, Zhou X and Yu B (2001) Effects of construction way and stroke number in visual perception of transforming Chinese characters. *Acta Psychologica Sinica* 33(3): 204–208.

Zhang Q and Reilly RG (2015) Writing to read: The case of Chinese. *29th Pacific Asia Conference on Language, Information and Computation*: 345–354.

Zhang Q and Yang T (2017) Bilingual education for the Tujia: The case of Tujia minority schools in Xiangxi autonomous prefecture. *International Journal of Bilingual Education and Bilingualism*: 1–19. DOI: 10.1080/13670050.2017.1358694.

Appendix 1: List of 102 characters and their log$_{10}$ frequency of occurrence in a corpus of 258 million

(Adapted from Jun, 2010)

爸	包	北	病	不	茶
4.48	5.01	5.33	5.12	6.49	4.51
床	道	到	的	对	多
4.59	5.99	6.05	6.92	5.88	5.76
二	发	房	父	个	狗
5.74	5.80	5.08	5.10	6.14	4.40
哥	贵	过	国	还	果
4.82	4.85	5.84	6.04	5.71	5.44
和	候	喝	回	家	会
6.04	5.17	4.64	5.54	5.84	5.90
间	介	进	觉	今	可
5.54	4.68	5.72	5.20	5.41	5.95
课	快	来	儿	了	乐
4.43	5.13	6.16	5.48	6.41	5.05
里	两	零	卖	妹	门
5.85	5.60	4.34	4.64	4.60	5.52
米	们	面	母	南	呢
4.90	6.09	5.70	5.08	5.42	5.12
你	年	起	气	去	让
5.94	5.93	5.72	5.38	5.84	5.16
人	上	绍	是	识	师
6.40	6.14	4.49	6.48	5.15	5.32
时	事	手	说	四	他
6.01	5.86	5.54	6.05	5.53	6.26

听	头	玩	问	我	午
5.36	5.56	4.53	5.57	6.29	4.71
系	姓	习	休	西	学
5.32	4.76	4.83	4.65	5.56	5.72
羊	要	医	英	一	友
4.42	5.97	4.96	5.14	6.57	4.94
以	有	院	月	云	在
6.15	6.35	5.19	5.62	5.20	6.36
住	这	中	祝	再	昨
5.26	6.24	6.14	4.31	5.36	4.37

mean	5.46
SD	0.62
max	6.92
min	4.31