Read Writ (2010) 23:589–605
DOI 10.1007/s11145-010-9228-0

Chinese character decoding: a semantic bias?

Clay Williams · Thomas Bever

Published online: 6 April 2010

Abstract The effects of semantic and phonetic radicals on Chinese character decoding were examined. Our results suggest that semantic and phonetic radicals 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 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.

Keywords Chinese character processing · Chinese reading · Dual route · Semantic activation · Phonological activation

C. Williams (✉)
Communications 302, University of Arizona, PO Box 210025, Tucson, AZ 85721, USA
e-mail: clayw@email.arizona.edu

T. Bever
Departments of Linguistics, Psychology, Cognitive Science & Neuroscience, University of Arizona, Tucson, AZ, USA

Introduction

Reading Chinese characters

The Chinese character system presents an interesting conundrum to traditional letter-recognition-based models of word recognition. The fact that Chinese does not have letters, but rather employs a complex set of characters, each possessing a meaning and a pronunciation, suggests that the initial decoding process for the Chinese reader may vary significantly from that of readers of alphabetic languages. The nature of the Chinese character system offers reading specialists a unique opportunity to directly compare semantic and phonological activation in reading—a comparison which is not easy to manipulate or to measure in most languages written in alphabetic scripts. In alphabetic–phonemic languages, the systematic mapping of sound to symbol makes phonological activation a relatively reliable means of word recognition compared with semantic recognition strategies. Indeed, semantic and orthographic correlation in alphabetic systems is largely arbitrary (e.g., *light*, *bright*, and *sight* seem to overlap both semantically and orthographically, but *tight* has no immediately-intuitive connection). In the Chinese writing system, semantic (as well as phonological) information may be embedded within the character itself. This supports the possibility of a dual route to lexical recognition from visual presentation of a character: the first being indirect through recognition of the word's phonology, and the other being direct access between orthography and semantic category (Zhou & Marslen-Wilson, 2000). Such a dual route to character recognition would have immediate benefits if decoding is conducted in a search model such as that described by Forster (1976): it would allow for a simultaneous, parallel searches based on different aspects of a character, thus minimizing search time. Such a parallel search route has been hypothesized to be a constant across languages, and computer models have given credence to its feasibility as a search model (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Coltheart & Rastle, 1994). In a homonym-dense language such as Chinese, putting semantic constraints on a phonetic search would be particularly useful. However, the absence of grapheme-phoneme correspondences in Chinese script makes some modification of the non-lexical route necessary in order to explain Chinese character processing, as sublexical phonological processing *has been found* in Chinese character decoding. The embedded phonetic component found in many characters has been found to activate pronunciation (Perfetti & Tan, 1998; Tan, Hoosain, & Siok, 1996; Zhou & Marslen-Wilson, 2000; Shen & Forster, 1999; Zhou, Marslen-Wilson, Taft, & Shu, 1999), and thus, characters whose phonetic components are accurate indications of pronunciation will be named faster and more accurately than characters with irregular phonetic components. However, as Chinese characters are not assembled from phonemes, and the phonetic correspondence of character pronunciation with the phonetic component is so low, the non-lexical route as originally conceived cannot adequately explain how characters are read. Shu et al. (2005) note that the model fails to explain oral reading of Chinese characters because:

> …the model assumes that lexical representations are not needed to read aloud. Instead, the subsymbolic units of the script are used to generate a verbal output for words and nonwords. Given that oral reading in Chinese is likely to involve contact with lexical representations as well as sublexical units, it is not clear how their theoretical position would explain reading in Chinese (p. 314).

Fortunately, Siedenberg and McClelland (1989) have proposed a language-specific adaptation of the dual-route model, built upon by Plaut et al. (1996), and later adapted by Weekes et al. (1997) that will accommodate Chinese script processing. This 'triangle' model contains three levels of representation—semantic, orthographic, and phonological—all linked via two bi-directional pathways: the semantic pathway and the non-semantic pathway. While the semantic pathway is basically a renaming of the lexical route, the nonsemantic pathway varies from the nonlexical route in that it allows for phonological representation at both the character *and* the sublexical level.

The suggestion of a dual route to accessing the meaning of Chinese characters, however, begs the question of whether the two routes are equal in importance, or whether one route is privileged over the other. In alphabetic languages, we can surmise that both semantic and phonetic search models exist—at least in theory—but the reliable phonologically-based organization of the alphabetic scripts obviously predisposes the reader to phonetic search patterns. Chinese writing, with a large proportion of characters having both phonological and semantic information directly embedded in the character, could potentially allow for either search strategy. There is evidence of separate search patterns, depending upon task type (see below "Priming studies"), but there is much debate as to whether there is a *default* reading strategy that tips towards semantic or phonetic interpretation.

Chinese character composition

Chinese characters can be separated into four classes of symbols: pictographs, indicatives, ideographs, and semantic-phonetic compounds. This last category comprises the vast bulk of the language—roughly 81% (Chen, Allport, & Marshall, 1996). These characters are formed by joining together a character with a related meaning (the semantic element or "radical") and another character (the "phonetic" element) to indicate its pronunciation. For example, 氵water + 木/mù/ = 沐/m*ù*/ "to wash one's hair." The radical (semantic root) portion of the character is usually located either above or to the left-hand side of the character. It is used for: (1) identifying semantic elements (e.g., in the character 媽[*/ma/*: mother], the semantic radical, located on the left-hand-side, is 女[/nu/: girl]; and (2) looking up entries in dictionaries. The phonological radical is usually located below or to the left of semantic elements (but exceptions do occur). The reliability of the phonetic radicals is highly variable—some characters, such as 媽/ma/—shown above, possess true indications to their pronunciation (馬[/ma/: horse]), whereas other characters' pronunciation may differ considerably, depending upon combination with various semantic elements, i.e., 工/gong/: 紅/hong/, 江/jiang/, 杠/gang/, 扛/kang/. This compositional structure of the characters requires Chinese readers to develop reading strategies quite different from those of English readers (or readers of other alphabetic scripts). Native Chinese speakers tend to learn their reading and writing skills through rote, word-by-word memorization, and frequent repetition (Chan, 1999), and exhibit a large reliance on visual information in word decoding strategies (Chikamatsu, 1996). Chinese-speakers also focus on semantic recognition of characters, as opposed to phonological analysis (Pine, Huang, & Song, 2003). Shu and Anderson (1997) showed that learners of Chinese made extensive use of knowledge of characters' semantic radicals for determining semantic information. Literate readers also made use of semantic radicals for recognizing less commonly used characters.

Highly literate speakers can also make use of phonological information sometimes embedded in characters; however, Shu and Anderson (1997) found that native speakers would not apply this strategy consistently until around the sixth grade, by which time most average children would meet government standards of basic literacy.

Priming studies: evidence of dual routes to character decoding

Chinese word-recognition should be primed via presentation of semantically related stimuli if there is a direct route between orthography and semantic information. Some studies have indeed supported this hypothesis (e.g., Zhou & Marslen-Wilson, 2000). Feldman and Siok (1999) found that character recognition was significantly facilitated by semantically related primes with the same radical, as compared to primes that were semantically unrelated but had the same radical, or primes that were semantically related but had a different radical. Ding et al. (2004), likewise, found significant facilitation for primes that shared the radical with the target. However, they only found priming when the test character component was in the same spatial position within both the prime and the target (e.g., 皑 would not prime 柏, even though both contain the submorphemic unit 白).

By contrast, Perfetti and Tan (1998) made a strong claim based on carefully titrated word-naming tasks that phonological activation precedes semantic activation in Chinese reading, and that Chinese semantic information is only accessible through phonological access. This account, however, has come under some attack: Zhou and Marslen-Wilson (2000) have found strong semantic priming effects in lexical decision tasks, whereas phonological priming effects were only attained for prime durations of 200 ms or more (i.e., only with visible primes). Shen and Forster (1999), likewise, found that phonological priming in Chinese is task-dependent. Tasks that emphasize phonological access show phonological priming, semantic tasks elicit semantic priming: the word naming tests conducted by Perfetti and Tan (1998) may have inadvertently biased subjects towards a phonological mediation of character reading. However, this task-dependent nature of character reading is still up for debate, given the more recent claims of task-independence by Perfetti et al. (2005), and claims that phonological activation is instead dependent upon the structuring of graphemic units in the orthographic system.

In any case, it is implausible to completely discount the importance of semantic mediation of character-recognition. 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.

The following studies are directed at two main tasks: (1) presenting evidence for dual routes of Chinese character recognition by demonstrating that semantic and phonetic radicals can each affect recognition speed in appropriate tasks, semantic categorization, and homonym recognition, respectively; and (2) determine which route (semantic or phonetic) readers are more likely to employ for recognizing characters in a reading activity which does not particularly favor phonetic or semantic processing strategies (lexical decision).

Experiment 1: semantic categorization

In each Chinese character, the semantic radical is usually combined with a phonetic radical to form a compound symbol. While the semantic radical often reliably indicates the semantic category of the character/word, there are also characters in which the semantic radical does not correctly correspond to the semantic category that it usually conveys. Thus, orthographically similar semantic radicals do not always indicate a similar semantic category. Conversely, many words lack the usual radical that indicates their semantic properties. In Table 1, we see an example of this with the radical 水 [*/shui/*: water]. This variable relationship between form and meaning allows us to ascertain whether semantic priming in Chinese is form-dependent. The following experiment was designed to ascertain whether presence of the semantic radical would facilitate or impede recognition of the correct semantic category of a character.

Table 1 Illustration of semantic radicals

	Semantic radical is a true representation of semantic category	Semantic radical is *not* a true representation of semantic category (i.e., orthographic relation only)	Character has semantic similarity to target, but lacks the semantic radical
Target semantic radical: 水 */shui/*: water (note: radical form of character is 氵)	汤 [/tang/: soup]	法 [/fa/: law]	雨 [/yu/: rain]

Method

Subjects

Thirty-six subjects—all native speakers of Mandarin from the People's Republic of China (PRC), participated in this experiment. All participants were students, currently-enrolled at the University of Arizona, at the time of the experiment. Both undergraduates and graduate students were used in this experiment. All subjects had normal or corrected-to-normal vision. Subjects were recruited via advertisement on a local Chinese-language online discussion forum, and all were compensated for their participation in this study.

Design and materials

The test consisted of 35 different semantic categories, each containing four single-character test words. Subjects were shown a semantic category (e.g., water, animal, etc.), and then presented with four characters, one at a time. Subjects were asked to determine quickly whether each character fit within the target semantic category or not. All categories and characters were presented and reaction times were measured using DMDX (Forster & Forster, 2003). Four character conditions were used: (1) S+R+ characters with semantic radicals that accurately indicated semantic category (i.e., "yes" response); (2) S−R+ characters with semantic radicals appropriate for the target category, but actually unrelated to the whole character's actual meaning (i.e., "no" response); (3) S+R− characters that fit the semantic category but do not possess the radical normally associated with said category (i.e., "yes" response); and (4) S−R− a negative control category, wherein the character has no relation—semantically or orthographically—with the target category. See Table 2 for an example of condition types using the semantic radical for *mammal/ animal*. Character frequency was matched between conditions 1 and 3, and between conditions 2 and 4 (see Table 2)—i.e., "yes" response characters were matched to each other and "no" responses were matched to each other. An additional 12 filler categories and targets were presented in order to prevent subjects from recognizing patterns. These filler categories, likewise, were followed by four targets, but they varied from having two yes/two no answers (i.e., they could have all yes, all no, three yes/one no, or one yes/three no answers). These filler categories were not

Table 2 Character frequency across conditions

	Condition 1	Condition 2	Condition 3	Condition 4
Character frequency mean	59,315	63,125	67,732	57,434
SD of character frequency	164,597	90,685	97,502	59,634

analyzed, but merely were there to prevent subjects from recognizing a two yes/two no format, and thus being able to guess the answer to the 4th presented target prior to presentation of stimuli.

Procedure

Subjects were shown a semantic category (e.g., water, wood, etc.), followed by four individual target words in random order which they were asked to categorize as belonging to that particular semantic category or not. The characters were generated in SimSun script, with target characters presented in a size 20 font on a 1024 × 768 pixel display area. Presentation of stimuli and recording of reaction time were controlled via DMDX (Forster & Forster, 2003). Presentation of categories was self-paced (i.e., the program would pause at each new category and await the subject command to proceed with target display), but the presentation of targets, once commenced, was automated. Targets would disappear following response or time out after 4 s. After each response, feedback on accuracy and reaction time appeared for roughly 1,600 ms (e.g. "correct: 790 ms"), and then the next target or category would be automatically displayed. While giving accuracy and RT feedback to subjects is not always necessary, it is the default setting to the DMDX platform. Additionally, it was considered a benefit in this study, as it helped to keep subjects from second-guessing certain semantic relations (i.e., straining to relate categories) by showing them that their initial reactions were usually correct. Instructions and four practice categories (16 semantic categorization decisions) were presented prior to test items (Table 3).

Results of experiment 1

Mean reaction times across the four priming conditions are shown in Table 4. There were strong effects associated with the kind of semantic radical discussed in detail

Table 3 Illustration of test conditions

Four-footed animal	Condition 1: semantically relevant and possesses associated radical (S+R+)	Condition 2: possesses associated radical but not semantically relevant (S−R+)	Condition 3: semantically relevant but does not possess associated radical (S+R−)	Prime condition 4: negative control (S−R−)
Associate radical	狼 [/lang/: wolf]	获 [/huo/: capture]	虎 [/hu/: tiger]	哥 [/ge/: older brother]

Table 4 Mean reaction times (ms) and error rates (percentage) across 4 conditions in Experiment 1

Conditions	RT (ms)	Error (%)	Facilitation of R+
Condition S+R+	749	12.9	+47 (vs. Cond. #3)
Condition S−R+	874	19.9	−58 (vs. Cond. #4)
Condition S+R−	796	20.9	–
Prime condition S−R−	816	7.0	–

below. As standard testing paradigms preclude comparison between positive and negative answers, for statistical purposes, Condition S+R+ was only analyzed with Condition S+R− and Condition S−R+ was only compared to Condition S−R−.

Conditions S+R+ and S+R− (comparison of effect of a semantically relevant radical)

A significantly faster response time was found when the S+R+ condition is compared to the S+R− condition, indicating that the presence of a semantically relevant radical facilitating recognition of the character's semantic grouping. $F_1(1, 31) = 29.07$, $P < 0.01$. Subject error rate also showed significance, $F_{\text{sub error}}(1, 31) = 41.32$, $P < 0.01$. Item analysis approaches significance, at $F_2(1, 68) = 3.96$, $P < 0.05$, but item errors do not: $F_{\text{item error}}(1, 68) = 2.35$, $P > 0.12$. Correlating the results with the frequency rate of the characters, $r(70) = -0.04$, $P > 0.37$, shows no real effect of character frequency compared to presence of the radical.

Conditions S−R+ and S−R− (comparison of effect of a non-semantically relevant radical)

The presence of a non-semantically relevant radical in the S−R+ condition slowed subject response time and increased errors when compared to the S−R− condition (control). $F_1(1, 31) = 13.02$, $P < 0.01$ and $F_{\text{sub error}}(1, 31) = 41.32$, $P < 0.01$. Subject error rate differences were also significant, $F_{\text{sub error}}(1, 31) = 41.32$, $P < 0.01$. Item analysis: $F_2(1, 68) = 6.46$, $P < 0.014$, $F_{\text{item error}}(1, 68) = 10.79$, $P < 0.01$. Correlating the results with the frequency rate of the characters, $r(70) = -0.06$, $P > 0.31$, indicates that these effects are from the presence of the radical and not due to frequency effects.

Discussion of experiment 1

These results support the hypothesis that there is a semantic route for Chinese character decoding. The presence of a relevant semantic radical facilitated semantic categorization—as long as the radical was a correct indicator of semantic category. In the S−R+ condition, where the entire character does not fit the target semantic category, despite possessing a related semantic radical, there was a clear pattern of impairment. Thus, the semantic radical can be a double-edged sword of sorts in

Table 5 Illustration of phonetic components in Chinese characters

	Phonetic component accurately represents pronunciation of character	Phonetic component *does not* accurately represent pronunciation of character	Character pronounced like target, but does not share any orthographic components
Target phonetic component: 工 /gong/	功 /gong/	红 /hong/	宫 /gong/

Chinese character decoding. The semantic radicals are helpful to character recognition only when they act as true semantic indicators. This finding complements the results from priming studies which have obtained semantic priming in Chinese reading (Feldman & Siok, 1999; Zhou & Marslen-Wilson, 2000) and supports the idea of a semantic path to character decoding, as described by Weekes et al. (1997).

Experiment 2: homonym recognition

Since Chinese characters allow for phonetic encoding, as well as semantic encoding, it is important to explore the effect that this phonetic radicals have on reading. While the phonetic radical has a relatively lower rate of correspondence with actual character pronunciation, there have been recent arguments that Chinese character decoding is primarily based upon phonetics. Perfetti and Tan's (1998) claims that phonological activation precedes semantic activation have triggered further study on the role of the phonetic component of compound Chinese characters, but subsequent studies have cast doubt on whether or not phonology takes a leading role in character decoding. Shen and Forster (1999) and Zhou and Marslen-Wilson (2000) indicated that no reliable phonological priming occurs in lexical decision tasks when the prime is presented for durations shorter than 200 ms. However, the fact that Shen and Forster did find priming effects from primes containing orthographically similar phonological radicals to the target indicates availability of the strategy of initiating the word search based upon the phonological radical. Given the variability in pronunciations of characters with the same phonetic component, this seems like an uncertain, but still plausible, strategy. Table 5 illustrates the phonetic component of characters. The following experiment was designed to probe whether accurate phonetic character components would facilitate whole-character recognition.

Method

Subjects

The same participants from Experiment #1 above were used in the following study.

Table 6 Illustration of test conditions

	Condition 1: same pronunciation and same phonetic component (P+C+)	Condition 2: different pronunciation but same phonetic component (P−C+)	Condition 3: same pronunciation but not the same phonetic component (P+C−)	Prime condition 4: negative control —no relation (P −C−)
Character pair	安 氨 /an/ /an/	位 泣 /wei/ /qi/	丰 风 /feng/ /feng/	往 根 /wang/ /gen/

Design and materials

The test consisted of a total of 184 pairs of Chinese characters. All character pairs fit into one of the following four types of relationships (illustrated in Table 6): (1) P+C+ are pronounced the same and share the same phonetic component; (2) P−C+ are pronounced differently but share the same phonetic component; (3) P+C− are pronounced the same but have no orthographic components in common; and (4) P−C− are pronounced differently and have no orthographic components in common (control). In the interest of maintaining a stark phonological difference in the negative conditions, all effort was made to maintain difference between character pairs in onset, rhyme, and tone; however, in category 2, two character pairs still ended up having a shared rhyme and four character pairs shared the same tone. In the control condition (condition #4), three pairs shared the same tone. It is important to note here that the test conditions here sometimes extend beyond mere phonetic markings to character structure itself, as in condition 1 in Table 5, wherein the first character actually *becomes* the phonetic component of the second character. Still, this is a valid means of testing the condition as we are essentially testing awareness of how overlapping graphemic structure predicts phonology.

Procedure

Subjects were shown 184 different pairs of Chinese characters and were asked to indicate whether or not the characters were homonyms (differences in pronunciation were never restricted to tone—non-homonymic character pairs always varied in onset, rhyme, or both). The characters were generated in SimSun script at size 20 font. Presentation of stimuli, as well as recordings of reaction time was controlled via DMDX (Forster & Forster, 2003). The stimulus (i.e., pair of characters) was displayed until subjects responded or for a maximum of 4,000 ms. The character pairs were displayed simultaneously, and after each response, feedback specifying accuracy and reaction time appeared for 1,600 ms. Instructions and eight practice items were presented prior to test items.

Results of experiment 2

Mean reaction times across the four priming conditions are shown in Table 7.

Table 7 Mean reaction times (ms) and error rates (percentage) across four conditions in Experiment 1

Conditions	RT (ms)	Error (%)	Facilitation of R+
Condition P+C+	886	5.7	+24 (vs. Cond. #3)
Condition P−C+	1,059	13.1	−56 (vs. Cond. #4)
Condition P+C−	910	6.9	–
Prime condition P−C−	1,003	5.3	–

Conditions P+C+ and P+C− (comparison of presence of accurate phonetic component)

The presence of an accurate phonetic component showed a significant facilitatory effect $F_1(1, 35) = 4.65$, $P < 0.04$; $F_2(1, 90) = 1.54$, $P < 0.22$, but there was no significant effect on error rate: $F_{\text{sub error}}(1, 35) = 2.81$, $P > 0.10$, $F_{\text{item error}}(1, 90) = 0.31$, $P > 0.57$. Still, it seems apparent that the orthographic overlap of the shared phonetic component helped in identifying the two characters as homonyms.

Conditions P−C+ and P−C− (comparison of presence of inaccurate phonetic component)

There was a strong significant inhibitory effect from having a phonetic component embedded in a character which is not indicative of the whole character's pronunciation: $F_1(1, 35) = 42.35$, $P < 0.01$, $F_2(1, 90) = 17.73$, $P < 0.01$. This inhibitory effect corresponded with a rise in error rate as well: $F_{\text{sub error}}(1, 35) = 54.34$, $P < 0.01$, $F_{\text{item error}}(1, 90) = 11.50$, $P < 0.01$.

Discussion of experiment 2

These results complement those of the first experiment testing effects of semantic variables. Like the semantic radicals, phonetic components facilitate the relevant kind of processing, but only when accurate. When the phonetic component is an inaccurate indication of how to pronounce the character, there were slower response times, and larger error rates. If both semantic and phonetic information embedded in characters are useful for decoding only when accurate, then it stands to reason that a phonetic search strategy, dealing with over 800 phonetic radicals (Taylor & Taylor, 1983) and having a mere 18.5% (Zhu, 1987, cited in Hoosain, 1991) accuracy rate in the commonly-used character base would be significantly less efficient than semantic strategy composed of only about 200 radicals (Hoosain, 1991), whose accuracy ranges from 60 to 100%. Shen and Forster (1999) argued that Chinese decoding strategies may well be task-dependent, and that is certainly arguable in this case. A semantic categorization task would inherently bias test-takers towards a semantically based reading strategy, and a homonym recognition study would push one to use phonetic strategies.

Experiment 3: lexical decision task

The preceding studies and prior research show that Chinese readers can utilize both semantic and phonetic routes for lexical access, as best fits the lexical processing task. The remaining question, however, is which route dominates in a strategy-neutral task. Towards that end, we conducted a lexical decision task designed to measure whether participants predominantly used either the semantic radical or the phonetic component of compound characters in identifying words. Chinese characters have to be recognized with access to both kinds of radicals: a holistic lexical decision task does not logically impel recognition towards one radical or the other. To study this, we experimentally manipulated physical informativeness of individual radicals, making it possible to identify which part of the character is more critical for decoding. In this task, the characters presented for identification as actual or pseudo-characters were specially treated by blurring either the semantic radical or the phonetic component. The results were analyzed to determine whether one component would impede lexical decision time more than the other. To knowledge, this technique has not been employed previously for measuring the relative impact of character-internal components on whole-character recognition, but this study is, in effect, a higher-tech version of Liu's (1983) quadrant deletion study which concluded that quadrants removed from the upper, left-hand side of characters had the most detrimental effect on character recognition. Blurring techniques allow one to retain all character features; however, the information degradation measurably slows processing. It has been used previously in studies on Chinese and Japanese to study pace constants required for detection (e.g., Osaka, 1992) and component separation in two-character combinations (e.g., Huang, 1984, cited in Huang & Wang, 1992).

Method

Subjects

The same participants from Experiments #1 and #2 above were used in the following test. The order of presentation of this task, along with Experiments #1 and #2 (described above) was randomized. Subjects were allowed to take breaks between experiments, but all subjects chose to complete all three experiments within a 1-day period.

Design and materials

A total of 48 pseudo-characters and 50 true characters were used in this task. Pseudo characters were constructed by arranging the semantic radical and the phonetic component from existing characters into illegal combinations (see Table 6). Semantic radicals and phonetic components occupied their normal positions within the pseudo characters (e.g., a radical like 氵 "water" would only be normally seen

Fig. 1 Example of pseudo-character

Fig. 2 Examples of a pseudo character with blurred semantic radical (R) and blurred phonetic component (L)

on the left hand side—never on top or on the right)—and thus, readers would not know just from seeing one part of the character that it was an illegal character—however, the combination or semantic radical and phonetic component yielded a character that does not exist in the Chinese corpus (i.e., the characters violated no orthographic properties of Chinese, but nonetheless were pseudo-words). A total of 41 of the 48 characters had semantic radicals on the left side of the character, 6 semantic radicals occupied an upper position, and a single character had a right position radical (which was withdrawn from the data analysis over concerns of being a possible confound). Images were created using the GIMP GNU Image Manipulation Program (www.gimp.org) and stored as *.bmp files. True characters were likewise created as *.bmp files. All characters were then used to create two different blurred versions—one with a blurred semantic radical, and one with a blurred phonetic radical (see Fig. 1 for an example). The GIMP software blur feature "Gaussian Blur" level 7 was used to attain the desired amount of high frequency filtering. Such blurring would impede recognition of that character part, and require a higher amount of focus on that part. In effect, the blurring was designed to delay the initiation of semantic and phonological search patterns, respectively, and the resultant time difference between blurring would allow us to recognize whether lexical searches were initiated from the blurred or un-blurred portion. All characters were created at approximately a 20pt. font in SimSun script (fitting on an 80 × 80 pixel background). The characters were presented with DMDX (Forster & Forster, 2003). Subjects were split into two groups, each receiving a different experiment script, according to standard counterbalancing procedure. In the first script, half of the characters were presented with blurred semantic radicals, and the rest were displayed with blurred phonetic components. All characters were presented in a randomized order. In the other script, the blurring effect was reversed, so that a character with a blurred semantic radical in the first script would now have a blurred phonetic component, and vice versa (Fig. 2).

Procedure

Participants were shown a total of 98 characters (48 pseudo-characters and 50 true characters) and were asked to indicate (by pressing either a 是 [YES] key or a 否

[NO] key) whether or not the given character was an existent character in the Chinese language. Each character was displayed for a maximum of 4,000 ms (or until subjects responded), and responses were followed by a 1,600 ms display of accuracy and speed of response before displaying the next test character. All presentation and reaction time was controlled as before with DMDX, and instructions were given along with practice items before testing began.

Results of experiment 3

There was a small but note-worthy impairment effect for blurred semantic radicals relative to blurred phonetic components. Both groups responded more slowly in the case of blurred semantic radicals. The differences were small: $Z_{\text{blurred phon}} = -0.04$ (av. 725 ms), $Z_{\text{blurred semantic}} = +0.03$ (av. 735 ms), for an average difference of 10 ms slower response when the semantic radical was blurred; but this effect approached significance: $F_1(1, 26) = 4.12$, $P < 0.05$. Subject error rate, however, is where an effect became clear. The subject error rate indicated that subjects made significantly more errors when the semantic radical was blurred (13.66% error rate) than when the phonetic component was blurred (10.2%): $F_{\text{sub error}}(1, 26) = 7.80$, $P < 0.01$. Item analysis showed no significant effects: $F_2(1, 39) = 2.3$, $P < 0.14$, $F_{\text{item error}}(1, 39) = 3.24$, $P < 0.08$.

Discussion of experiment 3

These results indicate a small but definite preference for using semantic information in a strategy-neutral reading task. This would suggest that, for the average Chinese reader, the semantic route to lexical access is dominant over the phonetic one. These results are particularly striking as one considers that 2/3 of the phonetic components (average stroke count: 7.6) used contained at least 2 more strokes than the accompanying semantic radical (average stroke count: 4.1), and thus the degradation from the blurring effect might well be more pronounced. Thus, in terms of total stroke-count, a character with a blurred semantic radical would thus usually contain a higher percentage of unblurred information than a character with blurred phonetic component. In other words, responses were slower and error rates higher with blurred semantic radicals even though the reader had more total unblurred strokes to analyze. These results affirm Peng's (1982) studies showing that covered or missing information in the top-left part of the character (where semantic radicals are much more likely to be) caused a significantly higher rate of character-misreading or inability to identify than any other character quadrant. One must note that Taft and Zhu (1997) have previously argued that the radicals in a compound character are processed serially from left-to-right, and thus one may be tempted to chalk the results up to blurring the left-hand portion of the character having a more significant inhibitory effect on character recognition than blurring the right-hand side. However, it is important to take into account that the slower mean RT when the semantic radical was blurred (compared to a blurred phonetic radical) was consistent for characters with radicals that were positioned on the top of the character (a total of 8 of the test characters), as well as characters with radicals that frequently appear on the right-hand side (90 test characters).

If phonological decoding schemes were the default mode of all languages, as Perfetti et al. (1992) have argued, we would expect that impairing the phonological route by blurring the phonetic component to have impeded recognition time more so than impeding the semantic route to recognition by blurring the radical. Instead we find a more pronounced delay in processing when the semantic information is degraded, from which we can infer a slight preference for semantic decoding strategies by Chinese readers.

It is further important to remember that the small differences between the blurred sides are to be expected. Chinese characters can only be identified holistically, so it would be impossible to properly identify a character without taking both the semantic and phonetic parts into consideration. However, the difference in error rate when the semantic radical is obscured should give one some pause before assigning both character parts equal weight in the decoding process.

General discussion

Chinese script is significantly different from alphabetic scripts as semantic information is specifically embedded into characters, thus making a semantic route to meaning much more viable than in alphabetic languages. The existence of twin routes of lexical access—phonetic and semantic—is hypothesized to be a universal constant across languages and scripts, but some have argued that the dominance of phonetic interpretation is a language universal, and not merely restricted to alphabets and syllabaries which, understandably, predispose readers to using phonetic strategies of access over semantic ones, and thus, one would argue, Chinese lexical access must be principally phonologically mediated (Perfetti & Tan, 1998; Zhou et al., 1999). Indeed, there is strong evidence of phonological mediation in Chinese lexical access, but, as asserted by Shen and Forster (1999), such is likely to be task-dependent. The fact that these studies claiming stronger effects for phonological activation are largely naming studies should have raised some flags. Here, it has been demonstrated that evidence for both semantic and phonological processing strategies can be found if one is looking specifically for such, but in a task devoid of any particular advantage for either processing scheme, subject's error rates spiked significantly when their semantic input was impaired. Further strengthening the task-dependent nature of the findings, no significant effects were found for order of presentation of the experiments—i.e., the subjects performed similarly no matter what order the experiments were conducted in. Such strongly suggests that the semantic route is, for most people at least, the default or primary means of lexical access. Phonological access may be a parallel process that is either simply slower in most cases, or it may only be activated if the semantic search route hits a problem or lags (e.g., in the case of lower frequency characters, where, indeed, one finds a much higher rate of phonetic accuracy in phonetic components). Such would make intuitive sense, as the corpus of possible phonetic components is four times larger than that of semantic radicals, and the accuracy of semantic information also far outpaces that of phonetics, and thus would make for a more efficient search strategy. This is not to assign more importance to semantic radicals than to phonetic radicals, as indeed, the combination of the two must be recognized before whole character recognition can be realized; however, these results do indicate that the *search* strategies themselves may not be used equally. Such a conclusion could have a large impact on our view of Chinese processing, and especially upon the area of Chinese literacy acquisition pedagogy. While much more study is needed before making drastic changes, if characters are indeed principally accessed via information obtained from the semantic radical, one could make some recommendations of changes to traditional Chinese language pedagogy: (1) learners should be taught semantic access strategies explicitly (i.e., teach them how to identify the radical, and differentiate characters with radicals that do not correlate with character meaning); and (2) arrange the character acquisition order to more closely correspond to radical groupings.

Acknowledgments The authors would like to thank Sunjing Ji and Xuetao Xu for their valuable contributions to this study.

References

Chan, S. (1999). The Chinese learner: A question of style. *Education and Training, 41*, 294–304.

Chen, Y. P., Allport, D. A., & Marshall, J. C. (1996). What are the functional orthographic units in Chinese word recognition: The stroke or the stroke pattern? *The Quarterly Journal of Experimental Psychology, A49*, 1024–1043.

Chikamatsu, N. (1996). The effects of L1 orthography on L2 word recognition. *Studies in Second Language Acquisition, 18*, 403–432.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1197–1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Ding, G., Peng, D., & 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*, 530–539.

Fan, K. Y., Gao, J. Y., & Ao, X. P. (1984). Pronunciation principles of the Chinese character and alphabetic writing scripts. *Chinese character reform, 3*, 19–22. Beijing: National Commission of Chinese Character Reform (In Chinese).

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language, 40*, 559–576.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116–124.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Elbaum Associates, Inc.

Huang, J. T., & Wang, M. Y. (1992). From unit to Gestalt: Perceptual dynamics in recognizing Chinese characters. In H. C. Chen & O. J. L. Tzeng (Eds.), *Language and processing in Chinese* (pp. 3–35). Amsterdam, North Holland: Elsevier.

Liu, I. M. (1983). Cueing function of fragments of Chinese characters in reading. *Acta Psychologica Taiwanica, 25*, 85–90.

Osaka, N. (1992). Size of saccade and fixation duration of eye movements during reading: Psychophysics of Japanese text processing. *Journal of the Optical Society of America A: Optics, Image Science and Vision, 9*, 5–13.

Peng, R. X. (1982). A preliminary report on statistical analysis of the structure of Chinese characters. *Acta Psychologica Sinica, 14*, 385–390.
Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112*, 43–59.
Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 24*, 101–118.
Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a universal phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227–248). North Holland: Elsevier.
Pine, N., Huang, P., & Song, H. R. (2003). Decoding strategies used by Chinese primary school children. *Journal of Literacy Research, 35*, 777–812.
Plaut, D. C., McClelland, J. L., Siedenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.
Shen, D., & Forster, K. (1999). Masked phonological priming in reading Chinese words depends on the task. *Language and Cognitive Processes, 14*, 429–459.
Shu, H., & Anderson, R. C. (1997). Role of radical awareness in the character and word acquisition of Chinese children. *Reading Research Quarterly, 32*, 78–89.
Shu, H., Meng, X. Z., Chen, X., Luan, H., & Cao, F. (2005). The subtypes of developmental dyslexia in Chinese: Evidence from three cases. *Dyslexia, 11*, 311–329.
Siedenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.
Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*, 761–775.
Tan, L. H., Hoosain, R., & Siok, W. W. T. (1996). Activation of phonological codes before access to character meaning in written Chinese. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*, 865–882.
Taylor, I., & Taylor, M. M. (1983). *The psychology of reading*. New York: Academic Press.
Weekes, B. S., Chen, M. J., & Yin, W.-G. (1997). Anomia without dyslexia in Chinese. *Neurocase, 3*, 51–60.
Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology, 26*, 1245–1265.
Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology in reading Chinese compound words. *Language and Cognitive Processes, 14*, 525–565.