Yao-Chung Tsao & Tsai-Guey Wang

Information Distribution in Chinese Characters

Chinese passages were mutilated either in the right, left, upper, or lower halves and presented to native speakers to read. In Experiment 1 passages were read from left to right; while in Experiment 2 from top to bottom. Time taken to read them and errors made were analyzed. Both measures showed that in both experiments the upper halves of characters are easier to read than the lower half, and right halves easier than left. Regression analysis method was used to examine effects of seven independent variables on reading accuracy of the characters. Among them, phonetic cue, symmetry, and number of strokes in the presented half were found to be significant factors.

In this study we are interested in the relative speed and accuracy when reading different halves of Chinese logographs. As early as 1879 Javal, based on the eye's fixation points during reading, concluded that it is advantageous to read the upper half of a line of alphabetic printing (Huey, 1980/1968, pp. 99). However, Huey did not think the Javal's experiments were definitive. With an experimentally sound approach, Kolers (1969) found that the right halves of text written in Roman letters were easier to read. He concluded that the right halves must also carry more information. However, the results with texts written in Hebrew letters are different. Shimron and Navon (1980) presented mutilated English or Hebrew paragraphs to native English readers or native Hebrew readers, respectively. They found that for English the upper half was easier to read than the lower; the reverse was true for Hebrew. They concluded that the lower half of Hebrew letters is more informative. Using seven-character sentences from Chinese poetry as test materials, Chou (1930) conducted a similar study and found the left half and the upper half are easier to read. However, his reported data do not provide an adequate basis for the statistical evaluation of differences between conditions.

Three factors intrigued us and prompted this study. Chinese logographs occupy roughly an equal square space, with the number of strokes[1] appearing in that space ranging from 1 to 33 (we found the mean to be 10 based on our informal observations of samples of non-simplified characters). This imposes a wide range of stroke densities among different characters. First, we wanted to know how the number of strokes in half of a character affects the inference made about an entire character. Secondly, of the six categories of the Chinese characters, only phonograms contain information about pronunciation (see Wang, 1981, for more information). Typically, a phonogram consists

Visible Language, XVII 4 (Autumn 1983), pp. 357-364.
Yao-chung Tsao's address: Dept. of Psychology, Rensselaer Polytechnic Institute, Troy, NY 12181.

of two or more components. From the components contained in a phonogram, a reader gets *a hint* as to which semantic category a phonogram may be related to as well as *a hint* as to how it sounds. For example, the left portion of the phonogram for the word *river*  (pronounced /her/) signifies *water*, while the right portion provides a phonetic hint, namely, a word which rhymes with /ker/. Although phonograms account for approximately 80% of the Chinese characters, we have observed (based on a random sample of 240 characters from the text material used in the present study) that only 20% of the characters in these passages of modern Chinese carry useful phonological information for reading in Mandarin. These components appear predominantly on the right side of the characters. The fact that the phonetic component appears predominantly on the right side of some, but not all characters, makes it tempting to believe that the right half of a Chinese character should be easier to read than the left half. Thirdly, modern Chinese texts are printed in two typical orientations: one is in vertical columns, the other in horizontal lines. In the former, readers read downward then leftward; in the latter, from left to right and downward line by line in a manner similar to that of English. A contemporary Chinese reader, especially with a college education, usually has experience in reading text in both orientations. We thought it would be interesting to see whether the reading direction would influence the pattern of experimental results. For this reason two experiments were conducted. In Experiment 1 the materials were printed horizontally; while in Experiment 2 they were printed vertically. To study these, we used connected Chinese discourse rather than poetry which was used by Chou. We adopted the rationale and procedures used by Kolers (1969) and Shimron & Navon (1980).

Experiment 1

Method

Subjects. Twenty Chinese readers, with an undergraduate degree completed in Taiwan, served as subjects. Their ages ranged from 25 to 30 years. They had normal or corrected-to-normal vision. Five were women.

Materials. Five passages of 300 characters each were selected from a book on music composition (Lee, 1978) printed in left to right orientation. The content of these passages was believed to be unfamiliar to the subjects. Under Condition W, the characters were presented in their entirety; under Condition R, only the right halves of the lines of characters were presented; while under Condition L only the left halves could be seen. Similarly under Condition U, only the upper halves were visible, and under Condition D,

only the lower halves were visible. See Appendix A. Five other passages of 100 characters each were chosen from the same book. A different one of each of these served as practice material for each of the five experimental conditions.

Design. Each subject was tested individually under all five conditions. All subjects took the W Condition test first, then the other four. The sequence of the latter four tests was counterbalanced among subjects.

Procedures. Each subject was first informed of the nature of the task, namely to read the test material aloud in Chinese as rapidly and as accurately as possible. Before each test, the practice material of 100 characters was first presented in order to familiarize the subjects with the task. The subjects were encouraged to guess at any mutilated characters whose reading they were not sure of, or to skip them, if no guess was possible. A tape recorder was used to make a permanent record of the responses for later analyses. When the subject's testing was ended, the purpose of the study was explained in greater detail.

Results

The total reading time in seconds and number of errors per section of material were measured and counted. Total reading time was taken to be the interval between the first and last utterances. The number of characters skipped *and* misread combined constituted the number of errors. Table I shows a parallel effect for reading time and number of errors. Condition W is the easiest condition, followed by Conditions U, R, L, and D, in that order. Thus, the right halves of Chinese characters are read more rapidly and with fewer errors than the left. Similarly the upper halves are more easily and correctly read than the lower.

Table I. Means and Standard Deviations of Reading Times and Error Rates in Experiment 1*

	Conditions				
	W	U	R	L	D
Reading Times (sec)					
Mean	74.4	119.8	145.2	214.5	221.8
SD	9.3	30.5	41.3	109.6	105.6
Error Rates					
Mean	.01	.09	.13	.25	.51
SD	--	.03	.07	.07	.11

*300 Characters per passage.

A one-way ANOVA conducted on the reading times shows that the overall effect of variation in conditions was significant, $F(4,76) = 27.92, p < .0001$. Dunnett's Test involving a control mean (Kirk, 1968) was used to compare differences between any two of the five. It was found that all the differences in reading times are significant with the exception of the differences between Conditions U and R, and between Conditions L and D. The one-way ANOVA performed on the reading errors (with Condition W excluded because the error rate for the Condition was less than 1%) shows a similar pattern to that of the reading times, $F(3,57) = 194.54, p < .0001$. Duncan's test (Kirk, 1968) was used to test the *a posteriori* paired comparisons. The results indicate that all are significant at either the $p < .01$ or the $p < .05$ level.

Regression analysis was performed to study the effect of inherent factors of the Chinese logographs on the error rate of reading the mutilated characters. Reading time was not used for analysis because information on how much time a subject spent on an individual character was not available. Seven variables were chosen for this purpose. These are: (1) number of strokes in a visible half of a character; (2) whether the visible half of the character contained the semantic cue; (3) whether the visible half contained the phonetic cue; (4) whether the visible half was symmetrical, or identical to the other half; (5) whether a character was a function word; (6) whether the character was presented in the context of a short phrase, and finally; (7) whether any stroke of a character was severed as a result of the mutilation.

For this analysis 240 characters were randomly selected from the test materials, 60 from each condition. For each of these 240 characters, the number of subjects who identified the character correctly was counted. The proportion of subjects who identified each character was then calculated and used as the dependent variable. The BMDP stepwise regression method was then applied to the data. The following regression model was obtained:

Percent correct response $= 0.656 + 0.187x + 0.110y + 0.019z$ where

x: whether there is a phonetic cue
y: whether the two halves are identical
z: the number of strokes in the half character

Thus only three independent variables of the seven examined contribute significantly to the subject's correct verbal identification of a character. The analysis showed that these three account for only 21% of the total variance.

Experiment 2

The purpose of Experiment 2 was to study the effects of reading direction on reading half-character connected passages; if the information distribution found in Experiment 1 is independent of reading direction the same effects should be found when passages are presented vertically.

Method

Subjects: Twelve subjects were tested, five of them having served in Experiment 1.

Materials: Five passages of 150 characters each, half the length of those in Experiment 1, were chosen from a book on Chinese literature studies (Watson, 1962/1969). These passages were physically mutilated in the same ways as were those of Experiment 1. Each was to be read downward and then column by column leftward. Five other passages, of 75 characters each, were also selected from the same book and served as practice material. The design and procedures were the same as in Experiment 1.

Results

Table II summarizes the average reading times and the average error rates for all conditions. Dunnett's test was again used to compare reading times between any two conditions. Generally speaking, the results were comparable to that of Experiment 1, i.e., upper halves were easier to read than lower halves ($p. < 01$) and right halves easier than left halves ($p. < 06$). Errors were also analyzed with the exclusion of Condition W. A pattern of results, similar to that found with reading times was again obtained, i.e., upper halves were easier to read than lower halves ($p. < 01$) and right halves easier than left halves ($p. < 01$). In other words, our results are qualitatively the same whether the text was printed horizontally or vertically.

General Discussion

This study indicates that the upper halves of Chinese characters are read more correctly and rapidly and hence must carry more information than the lower halves; similarly, the right halves are read better and contain more information than the left halves. These results are independent of reading direction. Although these results may appear to be similar to those of Kolers

Table II. Means and Standard Deviations of Reading Times and Error Rates in Experiment 2*

	Conditions				
	W	U	R	L	D
Reading Times (sec)					
Mean	35.0	73.3	68.9	90.3	115.3
SD	4.1	23.9	25.5	33.6	53.3
Error Rates					
Mean	.01	.07	.06	.17	.29
SD	--	.04	.05	.07	.09

*150 characters per passage.

with English text, we believe that the effects for Chinese and English arise for different reasons. For Chinese, we conclude that the phonetic cue, the symmetry of the two halves, and the number of strokes are the major factors. In the course of regression analysis in Experiment 1 we found that the average numbers of strokes in the halves of characters for Conditions U, R, L, and D are 6.20, 5.05, 5.50, and 4.35, respectively. The upper halves thus contain 42% more strokes than the lower and could therefore explain the better reading of the upper halves. Although there are fewer strokes in the right halves than in the left, the left are less informative because they are predominantly occupied by only a limited number of semantic *significs,* such as: water, tree, woman, etc. However, semantic *signific* were not found to be an important variable in our regression analysis. On the other hand, the right halves of 20% of the characters in our material provide phonetic hints and thus make the text easier to read. Our data agree with Chou's (1930) findings with respect to the upper-lower difference, but not the right-left difference.

The fact that our regression analysis showed a large value for intercept (namely, 0.656) and accounted for only 21% of the total variance suggests that there must be other important variables which are not taken into account by it. We suspect that one of these must be contextual, that is the semantic information from the context preceding the character. It should be noted that this study was done with connected discourse as the text material so that the results would be more ecologically valid. However, such material does not allow for good control of the contextual effect upon individual characters. In favor of the study is the statistical methodology. It identified the various variables related to the speed and accuracy of reading mutilated characters. It can thus serve as an aid to the further understanding of the information distribution of the Chinese writing system. The method may further provide help for the process of simplifying whole Chinese characters. Simplification of Chinese writing, namely to reduce the overall number of strokes per character, has been proposed by some researchers and policymakers in China for some years (see Wang, 1981). It would be ideal to be able to keep the more informative halves intact while simultaneously reducing the number of strokes in the remaining less informative halves with the result of a reduction in writing time.

Appendix A. Sample of Reading Materials

Condition U

Condition R

Condition L

Condition D

The authors appreciate the participation of the students from Taiwan who served as subjects.

1. Printed Chinese characters are composed of strokes. A stroke is conventionally defined as a trace on the paper as produced by pen movement. The departure of the pen from the paper surface completes a stroke. One or more strokes then constitute a radical. There are approximately 200 radicals. Radicals can be characters or components of characters.

References

Chou, S.K. Reading and legibility of Chinese characters: II. Reading half-characters. *Journal of Experimental Psychology,* 1930, *13,* 332-351.

Huey, E. B. The psychology and pedagogy of reading. Cambridge, Mass.: M.I.T. Press, 1968.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences.* CA: Brooks/Cole, 1968.

Kolers, P. A. Clues to a letter's recognition: Implications for the design of characters. *The Journal of Typographic Research* (now *Visible Language*), 1969, *3,* 145-168.

Lee, C.-B. The analyses of Chinese language from musical perspective. Taipei: Catholic Associations, 1978 (in Chinese).

Shimron, J., & Navon, D. The distribution of visual information in the vertical dimension of Roman and Hebrew letters. *Visible Language,* 1980, *14,* 5-17.

Wang, W. S.-Y. Language structure and optimal orthography. In O.J.L. Tzeng & H. Singer (Eds.) *Perception of print: Reading research in experimental psychology.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Watson, B. [*Early Chinese literature*] (C. Lo, Ed. and trans). Taipei: Hwakang Series, 1969. (Originally published, 1962.).