Title
The Knowledge of a Chinese Reader Concerning how Chinese Orthography Represents Phonology

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN
1069-7977

Authors
Hue, Chih-Wei
Lo, Ming
Tsai, Fang-Zhi

Publication Date
2006

Peer reviewed

The Knowledge of a Chinese Reader Concerning how Chinese Orthography Represents Phonology

Chih-Wei Hue (Hue@ccms.ntu.edu.tw),
Department of Psychology of National Taiwan University,
No.1, Sec. 4, Roosevelt Road, Taipei, Taiwan, 106, ROC

Ming Lo (d93227101@ntu.edu.tw)
Department of Psychology of National Taiwan University

Fang-Zhi Tsai (r93227106@ntu.edu.tw)
Department of Psychology of National Taiwan University

Abstract

A phonetically aware Chinese reader is able to infer the pronunciation of an unknown character from its constituent components; especially the one occupies the right half of the character. According to the statistical model of language learning, such a position strategy is really a bias representation of how Chinese characters representing phonology. This research tested the two positions by collecting the way subjects pronounced two types of pseudo-characters, one with a high validity phonetic on the right and the other on the left. Although the results supported the statistical model, it also showed that the position strategy was a dominant heuristic subjects relied on to guess the pronunciations of unknown characters.

Introduction

A Chinese character is composed of line strokes, and thus does not carry the kind of grapheme-phoneme correspondence an alphabetic word usually possesses. It had been shown that because the Chinese orthography is not designed to represent phonology, without proper training of Chinese phonemes, a Chinese reader could not develop phonemic awareness (Cheung & Chen, 2004; Read, Zhang, Nie, & Ding, 1986). However, Shu, Anderson and Wu (2000) showed that a reader of simplified characters is able to develop the so called "phonetic awareness" as s/he learns to read increasing numbers of characters, and similar finding was also reported for readers of traditional characters (Hue, 2003).

Similar to a person with phonemic awareness, Shu et al. (2000) argued that a child who is phonetically aware has " …the insight into the principles that govern orthography-phonology relationships in Chinese …", and is able to form hypotheses to guide "…perceptual processing, strategies for learning and retrieving the pronunciations of characters, and … to forecast the pronunciations of unfamiliar characters….(p. 57)" They pointed out that the majority of the frequently used character are phonograms which are composed of two components, a phonetic located at the right half of a character and a radical located at the left (upper panel of Fig. 1). As a result, a phonetically aware reader is able to develop a position strategy to guess the pronunciation of an unknown character. S/he will use its right component to guess the character's pronunciation if the component is a pronounceable character itself. In the case that the component is not a pronounceable character, the reader will infer the pronunciation of the unknown character from its neighbors which also contain the component.

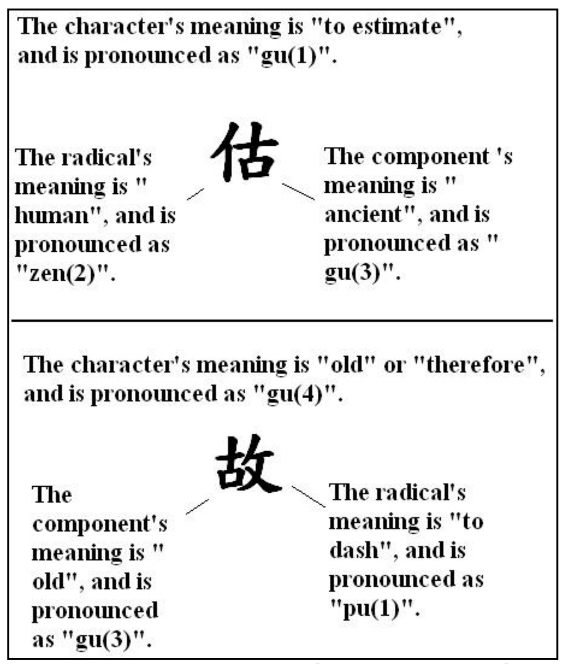

Figure 1: Two two-component characters, one with its right component as the phonetic and one with its left component as the phonetic.

Although Hsiao and Shillcock (2005) did not provide the statistics concerning how the various parts of a Chinese character representing phonology, they mentioned that the position of a phonetic is not fixed in a Chinese character, whether or not the character is a two-component phonogram (lower panel of Fig. 1). To test Hsiao and Shillcock's argument, the characters contained in the frequency norms prepared by the Chinese Knowledge Information Processing

Group (the CKIPG character frequency norm ,1993) and the characters contained in the 12 volumes of Elementary Chinese used in the elementary schools in Taiwan were analyzed concerning how phonology is represented in characters. As shown in Table 1, in the characters used by the first to the sixth graders and the general public, there were 79%, 85%, 87%, 89%, 90%, 91%, and 93% two-component characters, respectively. Among the characters with their two components left and right arranged, there were 38%, 39%, 42%, 44%, 46%, 48% and 53% of the characters with a right component providing useful phonetic hints. Moreover, we found that in all the character populations, there are about 7% of the left-and-right arranged characters with a left component providing cues to the pronunciation of the character.

A number of research have demonstrated that the statistical nature of linguistic input can be acquired by a learner through repetitive encountering of the language (Aslin, Saffran, & Newport, 1998; Saffran, Aslin , & Newport, 1996). According to this line of research, a Chinese reader's knowledge (or meta-linguistic awareness) concerning how characters represent phonology should reflect the statistical distribution of how the various parts of a character represent phonology. Obviously, this prediction was not confirmed by the previous research on how Chinese readers guess the pronunciations of unknown characters (Hue, 2003; Shu, Anderson & Wu, 2000).There are two possible explanations to account for the research results reported by Shu et al. and by Hue. The first possibility is that the statistical distribution concerning how phonology is represented in characters is different for the characters used by children and the characters used by adults. It is possible that the majority of the characters used by children are phonograms, and in these phonograms, it is the right components providing phonetic cues, and the left component does not. In such a reading environment, a child is likely to develop a position strategy, and the strategy is kept while the child grows up. This explanation has to be discounted because we have just showed in Table 1 that less than 20% of the characters used in the first two years in school use their right components as phonetics. Among the left-and-right arranged characters, the ratio is less than 40%, and there are about 8% of them with a left component as phonetic.

The second possible explanation is that the participants of studies by Shu et al. (2000) and Hue (2003) used a position strategy to infer an unfamiliar or a pseudo-character's pronunciation because the stimulus characters used in these studies possessed the same orthographical structure, i.e., a radical on the left and a phonetic on the right. Although all the radicals are pronounceable, very few people know how to pronounce them. Thus, in these studies, the only phonological information a stimulus item provided for a participant to make response was the pronunciation of its right component. This explanation was tested in the present research. By and large, this research followed Shu et al. (2000) and Hue's (2003) methods to collect how college and elementary school students pronouncing pseudo- and real characters. However, to avoid the possible response

Table 1: Analysis of how phonology is represented in the Chinese characters used by elementary school students and educated adults.

			Left and right arranged characters				Up and down arranged characters			
Population of the characters used	Number of characters	Number of characters composed of two components	Number of characters with their components left and right arranged	Number of characters with a phonetic locating at the right side of a character	Number of characters with a phonetic locating at the left side of a character	Number of characters with both components as phonetics	Number of characters with their components up-and-down arranged	Number of characters with a phonetic locating at the upper half of a character	Number of characters with a phonetic locating at the lower half of a character	Number of characters with both components as phonetics
Adults	5656	5285	3436	1814	233	63	1218	209	347	21
Sixth graders	2687	2438	1477	704	100	23	657	107	154	10
Fifth graders	2306	2082	1251	575	85	19	567	91	130	7
Fourth graders	1814	1619	958	422	67	14	446	70	98	6
Third graders	1322	1156	677	287	47	8	323	46	71	3
Second graders	896	760	448	174	32	2	210	22	42	0
First graders	399	314	175	66	15	2	91	14	7	0

biases induced by the stimuli used in those research, either the right or the left component of a pseudo-character used in this research is pronounceable.

Method

Participants

Twenty-two college and 288 grade school students participated in this experiment. Among the grade school students, there were 93 sixth graders, 63 fourth and 132 second graders. The second graders were further divided into two groups, those were tested at the end of a first semester (48 children), and those were tested in at the end of a second semester (84 children). All the participants were residents of Taiwan, and were native speakers of Chinese.

Materials and Procedure

The components which were either radicals and characters used to construct the characters of the 7 characters populations were first identified. The phonetic validities of the components were then computed. The phonetic validity of a component was defined as the ratio between the number of characters in a targeting character population containing it and the number of characters that it provides useful phonetic hints. In this study, the phonetic validity of a component was defined as high if the ratio was higher than 0.6, and low if the ratio was lower than 0.3.

The test items were different for the five groups of participants, although the principles of selection (or construction) of the items were the same. (1) The test items were real- and pseudo-characters, and were all consisted of two left-and-right arranged components. The questionnaire for college students contained 24 real- and 30 pseudo-characters. The questionnaire for the grade school students contained 12 real- and 24 pseudo-characters. (2) For the grade school students, the real characters were selected from the character populations corresponding to their grades. A college student was estimated to know about 5150 characters (Hue, 2003), and thus the real characters in the questionnaire presented to the college students were selected from the high frequency characters contained in the CKIPG character frequency norm (1993) which contains 5656 characters. With this sampling method, the participants should know the pronunciations of the real characters contained in their questionnaires. (3) There were two types of pseudo-characters, one with a high validity phonetic as right component and the other with a high validity phonetic as left component. The phonetic validity of a component was computed separately for the five participant populations based on their corresponding character populations.

The participants were tested individually, and in the test, a questionnaire in which the test items were printed one item a line was presented to each participant. They were asked to write down the pronunciation for each item using the Mandarin Phonetic Alphabets, which were taught in the first month of the first grade in every elementary school in Taiwan. After all the items were responded, they were asked to review all the pronunciations they responded and to explain why they made these responses[1].

Results

Data from 6 six graders was excluded from analysis because they didn't finish the experiment. A participant's response to the pronunciation of an item was categorized into one of seven categories, according to her/his "reason of response" and the experimenters' best judgment. Because the purpose of the present study was to find out which component of a test item was used by the participants to infer the item's pronunciation, the following analyses were performed on the participants' responses to the pseudo-characters. Table 2 showed the means of the ratios as a function of Participant Group, Response Category and Pseudo-character Type.

The ratios that the participants used the right component of a pseudo-character to respond were analyzed separately from the ratios that they used the left component to respond. The data was analyzed using a mixed model of analysis of variance (ANOVA), with Participant Group (5 levels; 2nd-1, 2nd-2, 4th, 6th graders and college students) as the between subject variable and Pseudo-character Type (2 levels; high phonetic validity component on the right or on the left) as the within subject variable.

Analysis of the ratios of the response which was made based on the right component of a pseudo-character revealed significant effects of Participant Group ($F(4, 299) = 23.20$, Mse = 0.04, $p < .0001$) and Pseudo-character Type ($F(1, 299) = 875.4$, Mse = 0.04, $p < .0001$). In addition, the interaction of the two variables was also significant ($F(4, 299) = 26.70$, Mse = 0.04, $p < .0001$). These effects indicated that the participants were more likely to use the right component of a test item to infer the item's pronunciation if the item's right component was high in phonetic validity, and the tendency to use the right component of an item to respond enhanced for the participants with larger vocabulary size.

Analysis of the ratios of the response based on the left component of a pseudo-character revealed a significant effect of Pseudo-character Type ($F(1, 299) = 772.97$, MSe = 0.02, $p < .0001$). In addition, the interaction between the two independent variables ($F(4, 299) = 21.43$, MSe = 0.02, $p < .0001$) was also significant. Results of the analysis of simple main effects indicated that compared with younger participants, older participants were more likely to use the left component of a pseudo-character to infer the character's pronunciation if the left component was high in phonetic validity ($F(4, 299) = 10.99$, MSe = 0.02, $p < .0001$), and they were less likely to use the left component of a character to infer character pronunciation if the right component was high in phonetic validity ($F(4, 299) = 7.83$, MSe = 0.02, $p < .0001$).

[1] The 2nd graders' answers to "reason of response" were written down by the experimenter.

Table 2: Means of the ratios as a function of response categories, subject groups and pseudo-character types.

Response types	Position of the high validity phonetic	Second graders		Fourth graders		Sixth graders		Adults	
		Right	Left	Right	Left	Right	Left	Right	Left
Using right component	Using it to infer character pronunciation	0.37	0.16	0.55	0.19	0.54	0.15	0.71	0.13
Using right component	Using other characters containing it to infer character pronunciation	0.19	0.20	0.09	0.19	0.19	0.24	0.18	0.27
Using left component	Using it to infer character pronunciation	0.09	0.29	0.08	0.26	0.03	0.35	0.01	0.35
Using left component	Using other characters containing it to infer character pronunciation	0.04	0.05	0.05	0.07	0.04	0.08	0.02	0.07
Using a character with visual shape similar test item to infer pronunciation		0.05	0.03	0.06	0.02	0.04	0.02	0.03	0.05
No responses		0.02	0.03	0.06	0.09	0.00	0.00	0.00	0.00
Others		0.24	0.24	0.10	0.17	0.09	0.13	0.05	0.10

Discussion

An analysis of the characters used by children and adults showed that in a considerable number of characters, the components of a character provide cues to the pronunciation of the character. The purpose of the present research was to investigate how a reader of Chinese represents the kind of correspondence, and how s/he uses this knowledge to guess the pronunciation of an unknown character. In particular, two positions, the statistical model and the right position strategy, concerning the issue were tested in this research. We showed that with proper design of the stimuli, the participants of this research used not only the right component of a pseudo-character to infer the character's pronunciation, but also its left component. That is, the results provided support for the statistical model of language learning. Moreover, the results that the two groups of second graders were able to use character components to infer character pronunciation indicated that the statistical nature of how phonological information is represented in characters is acquired fairly early by a Chinese reader. However, the significant interaction effects between Participant Group and Pseudo-character Type indicated that as a reader knows more characters, s/he will develop a right position heuristic to infer character pronunciation.

References

Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of conditional statistics by 8-month-old infants. *Psychological Science, 9,* 321-324.

Cheung, H., & Chen, H.C. (2004). Early orthographic experience modifies both phonological awareness and on-line speech processing. *Language and Cognitive Processes, 19,* 1-28.

Chinese Knowledge Information Processing Group (1993). *Corpus-based frequency count of characters in journal Chinese : Corpus-based research series no.1.* Taipei, Taiwan: Academia Sinica Institute of Information Science.

DeFrancis, J. (1984). *The Chinese language: Fact and fantasy*. Honolulu, HI: University of Hawaii Press.

Hue, C.W. (2003). Number of characters a college student knows. *Journal of Chinese Linguistics, 31,* 300-339.

Janet Hui-wen Hsiao, & Richard Shillcock. (2005). Differences of Split and Non-split Architectures Emerged from Modelling Chinese Character Pronunciation. *Proceedings of the Twenty Seventh Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Read, C, Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic spelling. *Cognition, 24,* 31-44.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science, 274,* 1926–1928.

Shu, H., Anderson, R. C., & Wu, N. (2000). Phonetic awareness: Knowledge of orthography–phonology relationships in the character acquisition of Chinese children. *Journal of Educational Psychology, 92,* 56–62.