

1165

Reprinted from

Psychologia—An International Journal of Psychology in the Orient—

Vol. XLII, No. 4 December 1999

**ARE PHONETIC ELEMENTS IN CHINESE CHARACTERS DRAWN FROM
A SYLLABARY?**

Ignatius G. MATTINGLY and Pai-ling HSIAO

Haskins Laboratories and University of Connecticut, U.S.A.

Psychologia Society
Department of Educational Psychology
Faculty of Education
Kyoto University
Sakyo-ku, Kyoto 606-8501, Japan

ARE PHONETIC ELEMENTS IN CHINESE CHARACTERS DRAWN FROM A SYLLABARY?

Ignatius G. MATTINGLY and Pai-ling HSIAO

Haskins Laboratories and University of Connecticut, U.S.A.

The key process that insures the productivity of Chinese orthography is phonetic compounding. According to some scholars (e.g. Defrancis, 1989), the phonetic component of a phonetic compound is drawn from a finite, limited set of phonetic elements, i.e., a syllabary. Phonetic compounding, however, is a recursive process: A phonetic compound may serve as the phonetic component of another, more complex compound. This implies that there is no finite set of phonetics, and suggests an alternative account: The set of possible phonetics may be simply the entire inventory of Chinese characters. To test the syllabary account, we compared the times required, in a character verification task, to reject three different bogus character types: pseudocompounds legal on both accounts, in which the phonetic is an element occurring as the phonetic in genuine phonetic compounds; pseudocompounds illegal on both accounts, in which the phonetic position is occupied by the combining form of a semantic; and doubtful pseudocompounds, in which the phonetic is a freestanding character that never occurs as the phonetic in a genuine compound. Such a pseudocompound is illegal according to the syllabary account, but legal according to the alternative account. We found that the clearly illegal pseudocompounds were rejected quickly and the clearly legal ones rejected slowly, as both accounts would predict, while the doubtful ones were rejected slowly, as if they, too, were legal, supporting the alternative account rather than the syllabary account.

Key words: writing systems, logography, Chinese characters, phonetic compounds

Every practical writing system must be productive: that is to say, it must afford the writer some procedure for transcribing a word which is phonologically possible in his language, but for which the standard transcription is nonexistent or at any rate unknown to him. In the case of an alphabetic system, the writer with this problem can make use of the set of grapheme-to-phoneme correspondences to write the word. He may not arrive at the standard spelling, if there is one, but he will probably be understood by the reader. Of course, most standard spellings were originally devised by someone relying on these same correspondences to transcribe a newly-coined or newly-borrowed word. Because there is no inherent limit on the length of a written word, the power to generate an indefinitely long word — and hence an infinite number of words — by concatenating alphabetic signs from a finite set, is implicit in

This research was supported by grant HD 01994 from the National Institute of Child Health and Human Development to Haskins Laboratories. The authors are grateful to Weijia Ni for his translation of Zhou (1978).

Correspondence concerning this article may be addressed to Ignatius G. Mattingly, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, USA (E-mail: ignatius@uconnvm.uconn.edu.).

the productivity machinery of alphabetic writing systems. Analogous observations can be made for syllabary and moraic systems.

The productivity processes of these phonologically-based systems depend on the same fundamental orthographic structure — the correspondence between graphemic and phonological units — that must be mastered by the beginning reader and that enables skilled readers to analyze unfamiliar words and, we would argue, familiar words also. Thus, for these orthographic systems, at least, consideration of a system's productive processes leads directly to an understanding of its fundamental nature.

PHONETIC COMPOUNDING

What productivity machinery is afforded by the Chinese writing system? Chinese orthography, like the Sumerian, Egyptian, and Japanese orthographies, uses logograms, that is, signs that stand for morphemes. But unlike these other writing systems, in which logograms have a merely auxiliary role and productivity depends on phonological signs, Chinese is a true logographic system. Accordingly, it has what the other systems lack, a systematic process for generating new logograms, namely, phonetic compounding. As is well known, phonetic compounding consists in joining the *semantic*, a classifying sign appropriate to the meaning of the morpheme to be represented, to the *phonetic*, a sign that is already a character itself, having a sound that is similar or identical to the sound of the morpheme.

- (1) 木 + 風 = 楓 Karlgren (1923, No.36*)
 mu⁴ feng¹ feng¹
 'tree' 'wind' 'maple'

Thus, in (1), the character for the morpheme feng¹, 'maple' is produced by joining the 'tree' semantic to a phonetic that is the character for a morpheme pronounced feng¹, 'wind'. There are, of course, a number of other processes by which complex characters have been formed. But only phonetic compounding is genuinely systematic and of broad applicability. It is not surprising that this process accounts for over 80% of the characters in the inventory (Zhou, 1978).

The semantics used in phonetic compounds are a specific, finite set. The question to be considered here is whether this is true of the phonetics as well. The notion that the phonetics do constitute a specific limited subset of the total character inventory has a long history. Early Chinese scholars called them "the thousand mothers of sounds" (Wieger, 1915/1965, p. 108.) For descriptive or pedagogical purposes, several later scholars, from Marshman (1814, as cited in DeFrancis, 1989) to Karlgren (1923), have made inventories of phonetics and the compounds derived from them. But it is DeFrancis (1989) who has taken this idea furthest. He claims that;

* Items in Karlgren (1923) are cited by entry number.

the phonetic elements are syllabograms that comprise a sort of syllabary. It is, to be sure, an outsized, haphazard, inefficient and only partly reliable syllabary. Nevertheless it works ... (p. 107).

A similar proposal has been made by one of us (Mattingly, 1992).

DeFrancis invites us to consider the process of Chinese character formation as a matrix, with the semantics on the vertical axis, the phonetics on the horizontal axis, and the phonetic compounds in the cells. Because both the semantics and the phonetics are taken to be finite sets, the number of possible phonetic compounds is necessarily also finite. Thus, on DeFrancis's account, the productivity process of Chinese contrasts in this respect with those of systems based on phonological units, which, as has been remarked, have potentially infinite outputs.

There are certain obvious problems with the syllabary account, as DeFrancis acknowledges. In many cases, one syllable sign corresponds to several phonological syllables. In many other cases, several syllable signs correspond to one syllable. For still other syllables, there are no signs at all. While the number of distinct semantics is generally taken to be around 200, the number of distinct phonetics is harder to pin down; the larger the dictionary analyzed, the more there prove to be. Thus Soothill's pocket dictionary (cited in DeFrancis) lists 895, while Marshman's (1814) analysis of the much larger Kang Xi dictionary lists 3867. But on any account, the number of elements is far larger than in any other known syllabary. Finally, only a small fraction of the cells in the matrix are actually used: Using Marshman's figure for the number of phonetics, DeFrancis estimates the number of cells to be over 800,000, of which 25,000 at most are filled with actual characters.

RECURSION IN PHONETIC COMPOUNDS

Although none of these objections is fatal to the syllabary account, they hardly encourage it. There is, however, an aspect of phonetic compounding DeFrancis does not discuss that we think is fatal to this account, namely: Phonetic compounding is recursive, as Zhou (1978) and Boltz (1994) have pointed out. Consider the following derivations:

(2a) 虫 + 凡 = 風 Karlgren (1923, No.18)
 'insect' fan² feng¹
 'all' 'wind'

(2b) 木 + 風 = 楓 Repeated from (1)
 'tree' feng¹ feng¹
 'wind' 'maple'

What (2a) shows is that the phonetic 風 in (2b) is *itself* a phonetic compound (though a less obvious one than (2b) because the semantic is not obviously appropriate and the phonetic is less similar in sound). Thus, a phonetic compound may become the phonetic component of a more complex phonetic compound.

Recursion is by no means a marginal phenomenon. Karlgren's *Analytic*

Dictionary of Chinese and Sino-Japanese (1923) is arranged in such a way as to make it easy to observe many such examples. Moreover, the recursion is not necessarily limited to two stages, as we see from (3a) (Karlgren, 1923; No. 42) and (3b-e) (Zhou, 1978, p. 175):

- (3a) 用 + 父 = 甫
 'use' fu⁴ 'father' fu³ 'name'
- (3b) 寸 + 甫 = 尊
 'inch' fu³ 'fu¹ 'spread'
- (3c) 氵 + 尊 = 溥
 'water' fu¹ pu³ 'great'
- (3d) 艹 + 溥 = 薄
 'grass' pu³ bao² 'thin'
- (3e) 石 + 薄 = 礪
 'rock' bao² bo² 'vast'

(3a)–(3e) show five stages of derivation. At each stage after the first, another semantic combines with the previously formed phonetic compound to yield a new compound. The recursive derivation of 礪 bo² could be more parsimoniously represented as in (4):

- (4) 礪 = 石 + [艹 + [氵 + [寸 + [用 + 父]]]]

There are relatively few such extreme examples, in practice, because of pressure to keep down the total number of strokes. Moreover, from time to time, the characters have been modified in a way that obscures some of the recursive structure. Thus, because of the simplification of 風 to 风, the derivation given in (2a) is not obvious in the Simplified Character set now used in the People's Republic of China (PRC).

For Zhou (1978), recursive derivation is one of several negative features of Chinese orthography. "Generations of phonetic radicals make the sound prone to change, hence weakening the phonetic representation function of phonetic radicals" (p. 176). For Boltz (1994), recursion is a significant process in the formative stage of Chinese writing, accounting for the many characters with more than two constituent elements. As for us, we believe that recursion is a synchronic as well as a diachronic process; the recursive structure of the characters is part of what the writer-reader knows about the structure of the Chinese writing system.

What the phenomenon of recursion implies is that in principle, every phonetic compound is a potential phonetic, and every phonetic can yield another generation of phonetic compounds. The number of potential phonetics in the supposed syllabary, and the number of possible phonetic compounds, are thus both infinite. The fact that there are an infinite number of possible phonetics is a fatal objection to the syllabary account, which necessarily assumes a finite set of syllable signs. A syllabary cannot be permitted to have an infinite number of elements.

AN ALTERNATIVE PROPOSAL

Because the syllabary account of phonetic compounding leads to these serious difficulties, we prefer an alternative account, namely: There *is* no special subset of potential phonetics. Rather, every freestanding Chinese character is a potential phonetic, available to be used in a phonetic compound. We can replace DeFrancis's vast but thinly populated matrix with a simple recursive formula for character generation

$$(5) S + C_n > C_{n+1},$$

where C stands for character, S, for semantic, and n , for order of complexity.

Notice that on this alternative account of character generation, the anomalies of the syllabary account are just what are to be expected. Some characters happen to have been used as phonetics relatively frequently, others, not at all. There is no tidy, one-for-one correspondence between phonetics and phonological syllables. As for the recursion, it is observed simply because any phonetic compound, just like any other character, is a potential phonetic. We see also that, in contrast with the finite output of DeFrancis's matrix, the output of the character generator is potentially infinite, just as the output of the productivity machinery of every other writing system is potentially infinite.

EXPERIMENT

To test this account of phonetic compounding, we devised an experiment to investigate whether, as the syllabary account would predict, bogus compounds whose "phonetics" were *not* drawn from the DeFrancis syllabary would be perceived by Chinese readers as impossible characters.

Stimuli:

Three sets of twenty bogus phonetic compounds were prepared for use in a character verification task; see Table 1 and the Appendix. Since we intended to use subjects educated in the PRC, these stimuli were based on the Simplified Characters. The same semantic components were used in all three sets; they were semantics whose combining forms are substantially different in appearance from the characters they derive from.

In each character in the first bogus set, called "Impossible," the position of the phonetic component is occupied by another such semantic combining form. Such formations are not possible in Chinese writing.

In the second set, called "Syllabary Pseudo," the phonetic position is occupied by a simple character that *does* appear as a phonetic in real Chinese phonetic compounds, but not with the particular semantic with which it is now associated. These items correspond to empty cells in DeFrancis's syllabary matrix.

In the third set, called "Nonsyllabary Pseudo," the phonetic position is occupied by one of the relatively small number of simple characters that *never* appear as phonetics in real compounds. Such a compound would be excluded from DeFrancis's matrix.

We also prepared a set of twenty real phonetic compounds, half of them of high frequency and half of them of low frequency.

It was expected that the impossible characters would be rejected more rapidly than the syllabary

Table 1. Four Stimulus Types Used in the Experiment

Character type	Character Structure			
	Compound		Semantic	Phonetic
Impossible	𠄎	=	革 'hide'	+ 𠄎 'move'
Syllabary Pseudo	𠄎	=	革 'hide'	+ 昆 kun ¹
Nonsyllabary Pseudo	𠄎	=	革 'hide'	+ 书 shu ¹
Real	物 wu ⁴	=	牛 'ox'	+ 勿 wu ⁴

pseudocharacters. If the syllabary account is correct, nonsyllabary pseudocharacters would also be rejected rapidly, for they, too, would be "impossible." But if the alternative account is correct, then nonsyllabary pseudocharacters would be rejected slowly, in about the same time as syllabary ones.

Procedure:

The 57 subjects, Chinese students from the PRC studying at the University of Connecticut and their spouses, were divided arbitrarily into three groups of 19. Each group saw the real compounds and one of the three sets of bogus compounds. This "mixed" design was chosen because it had been found in pilot work that a subject presented with more than one type of bogus character in the same test tended to accept, erroneously, the less obvious types. Each subject was tested individually and was presented with real and bogus compounds in random order on a computer monitor, one character every 10 seconds. The subject's task was to press the "Yes" key on the response box if he believed a character was real, otherwise the "No" key. Reaction times were measured and errors counted.

RESULTS

The results are shown in Table 2. Recall that each subject group saw *one* of the three types of bogus characters, but all three groups saw the same set of real characters.

For the error data, between-subjects analyses of variance were carried out for the responses to high and low frequency real characters and for the responses to the three

Table 2. Errors and Reaction Times by Stimulus Type

Character Type	Group		
	Impossible	Syllabary Pseudo	Nonsyllabary Pseudo
	Percent Errors		
Real High	6.8 (7.5)	10.0 (16.0)	6.8 (9.5)
Real Low	16.8 (18.6)	32.1 (19.9)	16.8 (18.0)
Bogus	1.6 (2.9)	19.5 (16.4)	12.4 (10.7)
	Reaction Times (msec)		
Real High	646 (152)	627 (111)	662 (147)
Real Low	752 (225)	756 (138)	731 (127)
Bogus	623 (163)	767 (157)	746 (118)

Note: Standard deviations are given in parentheses.

types of bogus characters. Real low frequency characters were recognized significantly less accurately than real high frequency characters by all three groups of subjects, $F(1, 54)=35.8$, $p<.0001$. The effect of subject group was just significant, $F(2, 54)=3.242$, $p<.05$, but a post-hoc Bonferroni/Dunn All Means analysis showed no significant differences between any two groups. There was no significant interaction between real character frequency and group. For the bogus characters, subject group (i.e., bogus character type) had a significant effect, $F(2, 54)=11.8$, $p<.0001$. A post-hoc All Means analysis showed that impossible bogus characters were recognized significantly more accurately than syllabary ($p<.0001$) or nonsyllabary ($p<.01$) bogus characters. The difference between syllabary and nonsyllabary characters was not significant.

Analyses of variance were also carried out for the reaction times for correct responses to real characters and to bogus characters. Reaction times were significantly shorter for high frequency real characters than for low frequency real characters, $F(1, 54)=39.5$, $p<.0001$. The effect of subject group was not significant. For the bogus characters, there was a significant effect of character type, $F(2, 54)=5.253$, $p<.01$. A post-hoc All Means analysis showed that impossible bogus characters were recognized significantly faster than syllabary ($p<.005$) or nonsyllabary ($p<.05$) bogus characters. The difference between syllabary and nonsyllabary characters was not significant.

DISCUSSION

The relatively high error rates for low-frequency real characters may be explained by the fact that many of the subjects are graduate students in the physical sciences and had probably read little Chinese in recent years. Some of them had great difficulty distinguishing syllabary and nonsyllabary bogus characters from real characters. On the other hand, they had much less difficulty distinguishing impossible characters from real characters. This difference is consistent with our alternative account, whereas the syllabary account would predict that the nonsyllabary characters would be as readily rejected as the impossible ones, being themselves impossible.

The reaction time data provide further support for the alternative account. The impossible characters were rejected very rapidly, indeed in slightly less than the time required to accept a real high-frequency real character. On the other hand, the syllabary pseudocharacters took longer to reject than the impossible characters, and about the same time as required to accept a low-frequency real character. Since, on either account, syllabary characters are possible characters, some nontrivial lexical processing must have been required before they could be rejected. Most importantly, the nonsyllabary pseudocharacters took longer to reject than the impossible characters, again implying some lexical processing. This nonsyllabary reaction time, in fact, was not significantly different from the syllabary reaction time. Apparently both the nonsyllabary pseudocharacters and the syllabary ones are possible characters containing possible phonetics. Our results thus discourage the view that there is a

psychologically real subset of phonetics; rather, all characters are potential phonetics, and the supposed syllabary has no demonstrable role in character recognition.

CONCLUSIONS

While conforming to general principles that hold for all practical orthographies, the Chinese writing system is unique in two important ways.

The first is the linguistic motivation for the infinite output of productivity processes. The infinity of possible written words in alphabetic, moraic or syllabary systems responds to the fact that in the languages these systems transcribe, there is in principle no limit to the number of phonemes or syllables in a spoken word, and so no limit to the number of different words. Thus, for these systems, to be productive means being able, in principle, to transcribe a word of infinite length. Such systems, however, provide no protection against homography, and if the language is prone to homophony, other devices — using logograms as determiners, for instance — may be necessary.

The infinity of possible Chinese characters, in contrast, is a response to the ubiquitous homophony resulting from the limitation of morphemes in Chinese languages, for the most part, to a length of one syllable with very restricted phonological structure. For Chinese writing, to be productive means being able, in principle, to transcribe unambiguously an infinite number of homophonous morphemes for any given syllable. On the other hand, just because Chinese words are made up of monosyllabic morphemes, the Chinese productivity process need not be concerned about the length of words.

Secondly, Chinese writing system is logographic, and thus its productivity process, phonetic compounding, depends on morphemes, rather than on phonological units. Given the close relation, in other writing systems, of productivity to reading and reading acquisition, we should not be surprised to find that for Chinese, morphemes are the basis of these other processes also. This does not mean, it should be emphasized, that the Chinese reader is exclusively occupied with meaning and has no concern with phonology. On the contrary, because both the meanings and the phonological values of morphemes play a crucial role in the productivity process, we would expect to find that the same holds true for reading and reading acquisition.

REFERENCES

- Boltz, W. G. 1994. *The origin and early development of the Chinese writing system*. New Haven: American Oriental Society.
- DeFrancis, J. 1989. *Visible speech/The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.
- Karlgren, B. 1923. *Analytic dictionary of Chinese and Sino-Japanese*. Paris: Librairie Orientaliste Paul Geuthner.
- Mattingly, I. G. 1992. Linguistic awareness and orthographic form. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 11-26). Amsterdam: North-Holland.

- Wieger, L. 1965. *Chinese characters/Their origin, etymology, history, classification and signification. A thorough study from Chinese documents* (L. Davrou, Trans.). 2nd ed. New York: Paragon Book Reprint Corp. (Original work published in 1915).
- Zhou, Y.-G. 1978. Xiandai Hanzizhong shengpangde biao Yin gongneng wenti (On the question of the phonetic representation function of phonetic radicals in modern Chinese). *Zhongguo Yuwen*, 146, 172-177. (In Chinese.)

(Manuscript received September 1, 1998; Revision accepted October 26, 1998)

Appendix

Characters used as stimuli

Bogus Characters

Syllabary	Nonsyllabary	Impossible
射	射	射
控	孫	孛
伦	帽	帽
酩	耐	耐
藕	輪	輪
峽	峴	射
砲	砲	扩
鞞	鞞	鞞
緯	緯	扩
缺	蟻	射
酌	酌	酌
呖	眼	眩
孩	旆	旆
瓠	燭	燭
饪	矯	矯
膵	膵	乱
購	疑	賤
彌	彈	羽
錡	錡	射
珪	璣	珪

Real Characters

High Frequency				
殖	物	粉	娘	阶
种	现	经	地	样
Low Frequency				
峽	喞	蒺	駢	轆
钜	釉	肫	蹀	聆