

• Sound Symbolism, ed. by Leanne Hinton,
Johanna Nichols, & John J. Ohala. Cambridge
10 U. Press, 1994

*An experimental investigation into phonetic symbolism as
it relates to Mandarin Chinese¹*

RANDY J. LAPOLLA

10.1. Introduction

10.1.1. General

About the time Plato was writing *Cratylus*, in which the protagonist argues that the relationship of sign to signified was one of object to imitation (Fowler 1977), that is, less than arbitrary, discussion of this question was also going on in China, especially among the Confucianists. The Chinese, in general, came to a different conclusion than Plato's *Cratylus*. The writer who spoke most directly to this question was Xun Zi (d. 221 BC) (Hong 1982). He felt that an object and its name had a totally arbitrary relationship, that no name was any more "suitable" than any other, that all names are "suitable" only as a result of convention and popular usage. Though there were some Chinese philosophers (from the Han Dynasty on) who attempted to find the *true* sound-meaning correspondences for words, going against Xun Zi's principle of non-suitability of names, the mainstream of Chinese linguistics to this day still does not really question the concept of total arbitrariness. Almost all of the general books on linguistics now in use in the People's Republic (e.g. Gao and Shi 1963; Ye and Xu 1981; Ma 1981) use the same quote from Xun Zi and another, from *Capital*, where Karl Marx says "The name of a thing is entirely external to its nature" (Fowkes 1977: 195).

In the West, although Saussure felt that "No one disputes the principle of the arbitrary nature of the [linguistic] sign" (1966: 68), there have been many since that time who have tried to show that the assignment of signifier to signified is not always completely arbitrary. Jespersen (1922) went to great lengths to show that among the languages he was familiar with there were definite correlations between certain types of sounds and certain categories of meaning, such as size, movement, feelings, and distances. Jespersen also felt that people have an instinctive feeling for what are the best sounds for the different meanings in each category.

10.1.2. Experimentation

Much of the debate between those who support the idea of sound symbolism and those who do not, and between those who support its universality and those who do not, has centered around methodology of experimentation.

Sapir (1929) was the first to try to prove the latter of Jespersen's hypotheses experimentally (particularly a feeling for "large" and "small" among English speakers), and was followed by Newman (1933), who tried to support Sapir's findings and expand on them by including the feeling for "bright" and "dark," and by examining the lexicon for sound-meaning correspondences. Though these two studies came under some criticism, the conclusions they felt their results led to set the tone for much of the discussion of sound symbolism that has taken place since that time. The five conclusions they drew were that

- (1) Phonetic elements tend to be rigidly patterned on a non-linguistic symbolic scale . . .
- (2) The factor of age . . . has little effect on the subjective scale . . .
- (3) The basis of phonetic symbolism is fundamentally objective . . .
- (4) Diverse types of phonetic patterns are formed by unlike symbolisms² . . .
- (5) These symbolic judgements are not produced by linguistic associations (Newman 1933: 75).

Since that time there have been some experimental investigations that have claimed to have disproved the theory of sound symbolism (Bentley and Varon 1933; Maltzman *et al.* 1956; Brackbill and Little 1957), but these results can often be explained as being due to important procedural problems (Brown and Nuttall 1959; Weiss 1963).

Starting with Brown *et al.* in 1955, a ten-year debate on the pros and cons of various types of methodology of experimentation and evaluation of results ran through the *Journal of Abnormal and Social Psychology* and later the *Psychological Bulletin* (Brackbill and Little; Brown and Nuttall 1959; Miron 1961; Weiss 1963; Taylor 1963; Weiss 1964; Johnson *et al.* 1964; Taylor and Taylor 1965), but no agreement on the best method was reached.

At least ten experiments have been done on sound symbolism using either the Chinese language (Mandarin) or Chinese subjects. (Sapir 1929;³ Brown *et al.* 1955; Brackbill and Little 1957; Brown and Nuttall 1959; Weiss 1963; Huang *et al.* 1968; Tsien-Lee 1969; Klank *et al.* 1971; Lester 1973, 1974.) Of these, nine had results supporting the sound-symbolism hypothesis, and only one (Brackbill and Little 1957) had results which did not support the hypothesis. With all of these experiments I believe there were problems of methodology and procedure serious enough to warrant a new set of experiments.

10.1.3 *Tonal morphology*10.1.3.1 *Monosyllabic*

A distinctive feature of a large group of languages, including many Southeast Asian languages and several Chinese dialects in the southeastern part of China, is the use of tone change to mark morphological or semantic change. Though this use of tone is not always limited to marking the hypocoristic or diminutive,⁴ this is one of its more common uses. This is true of Lahu, Yi (Lolo),⁵ and the Hui, Wu (that of southern Zhejiang), and Yue dialect groups of Chinese (Hirata 1983). In the Canton dialect (one of the Yue dialects), for example, which is one of the most representative and best attested of the Chinese dialects using tone change to mark the hypocoristic, a change to a high level or high rising tone is used to mark the diminutive or familiar.⁶

In standard Mandarin Chinese the hypocoristic of nouns is marked only with the affixation of a retroflex suffix,⁷ with no change in the original tone. Even so, if there is a universal tendency to associate high frequency with "smallness" (the diminutive in general), then Mandarin speakers also should be sensitive to tone differences such as those in Cantonese.⁸

10.1.3.2 *Disyllabic*

A second use of the change to a high tone in Cantonese is to change the intensity of reduplicated adjectives. This can be done to different degrees depending on which syllable the changed tone is assigned to. Lengthening of the vowel is also used for extreme intensification. For example, [hong²¹] "red" can be modified to [ho:ng³⁵ hong²¹] "very red" or [hong²¹ hong³⁵ tei³⁵] "a little red."⁹ (Yue-Hashimoto 1972: 95; Whitaker 1955–1956: 31) The first type of tone change is also obligatory on the first syllable of a three-syllable adjective formed by reduplication of the first syllable of a disyllabic adjective for the purpose of intensification, as in [søn³⁵ søn²¹ pan³⁵] "very good tempered" from [søn²¹ pan³⁵] "good tempered." The second type of tone change is possible for onomatopoeic (either in the metaphysical or real sense) phrases with an initial reduplicated syllable, as in [lap³³ lap³⁵ lyn³³] "disorderly"¹⁰ (Yue-Hashimoto 1972: 95).

The change of the first syllable to a high tone is obligatory when reduplication is used for extreme intensification, and can apply to syllables in all of the nine tones of Cantonese because of the lengthening of the syllables. The tone change in the second type of reduplication, though, is optional, and can only apply to syllables with low tones (Yue-Hashimoto 1972: 99). Because with the tone change the resulting tone pattern (i.e. low-high) is the same as that for familiarity, it is possible that the changed form is less intense or serious, as with the phrase [t'an²¹ t'an³⁵ tsan⁴⁴] "shivering," because the change puts it in the field of the hypocoristic. It is

also the case, in Thai and the Hainan form of the Southern Min dialect of Chinese, that modification of the first syllable of a reduplicated adjective by raising the tone and often lengthening the vowel is used for the purpose of intensification. In Thai, there is [dii⁵⁵ dii³³] "very good," from [dii³³] "good" (Li 1977: 8; Haas 1946: 128–30), and in Hainan Chinese there is [bui⁵⁵ bui³³] "very fat" from [bui³³] "fat" (Woon 1979: 87). According to Lien Chinfa (personal communication), this is also the case in mainland Southern Min.

What is significant in all these languages is that for extreme intensification, modification must be of the first syllable and the resulting tone pattern is high-low.¹¹ In Cantonese and in Thai, other types of adverbial compounds or marked disyllabic words will be marked by modification of the second syllable, so generally have a low-high tone pattern. In Cantonese this is also the case for marking familiarity, especially family relation terms, such as [pa²² pa⁵⁵] "daddy" or [mui²² mui³⁵] "sis" (Gao 1980: 22–23).

In Mandarin, the tone on the second token of a reduplicated adjective changes to a high level (55) tone, but this change of tone is not usually thought to be semantically significant. Although this is probably the case, I wanted to see if native Mandarin speakers would also be sensitive to the tone pattern-to-meaning correspondences in Cantonese.

10.1.4 *The hypotheses*

The hypotheses we attempted to test in this investigation are similar to those tested by other researchers: (1) that English speakers should have better-than-chance-expectancy success at matching Chinese antonymic word pairs, thereby showing that Chinese exhibits sound-symbolic patterns which are perceptible to English monolingual speakers, and (2) that Chinese native speakers would have enough of a common innate sense of sound-meaning correspondences to give a common pattern of responses when asked to assign meanings to sounds.

10.1.5 *General methodology*

The method we used in the first experiment of the investigation was one of those that is generally most supported, matching antonymic pairs, being careful to eliminate or reduce any of those factors that have been called into question.¹² But for one part of the investigation using Chinese subjects (Experiment 2, Part a), we used a new method that gave the subjects much more freedom in making sound-meaning choices, while at the same time allowing us to analyze the choices into grave and non-grave segments, a key parameter of sound symbolism (Jakobson 1978).

10.2. Experiment 1

10.2.1. Method

10.2.1.1. Part a

The first part of the first experiment was designed to test the ability of native English speakers who had never been exposed to Chinese to match the correct word out of a pair of antonymic Chinese words and the written English translation of one of them. A significant number of correct answers would point to the existence of sound-symbolic patterns in the Chinese lexicon, and would also be evidence that English speakers are sensitive to them.

A list of 40 English words that are members of common antonymic pairs was given to each of the subjects.¹³ Beside each word were the numbers 1 and 2, and the word "neither."¹⁴ A tape recording of a list of antonymic pairs that were the Chinese equivalents of each English word and its commonly accepted antonym, spoken by a native speaker of Mandarin from Taipei,¹⁵ was played for the subjects. Each pair was spoken with two seconds between the members of the pair, and there was a five-second pause between pairs. The subjects were asked to circle 1 or 2, depending upon which word of each antonymic Chinese pair they thought had the same meaning as the English word given, or "neither," if they felt that neither of the words "felt right." They were told that some pairs might be similar in meaning to other pairs, but to judge each pair separately. The word list used, and the Chinese words¹⁶ spoken for each item, are given in table 10.1.

The Chinese words were spoken because giving a romanized transliteration of the Chinese words would not give linguistically naive subjects a good idea of what the sound was really like, especially the tone. This could also lead to the subjects matching the words because of orthographic similarities.

Only one English word was given for each pair, for three reasons. First, it simplified the answering process for the subject, comparing one word with two instead of two with two. Second, it reduced the possibility of the subject just matching up the vowels of one or both of the English words with that/those of the Chinese words, or basing his/her choice on the English-Chinese pair that was not being asked for.¹⁷ Third, it allowed us to use the other half of some of the English pairs later on in the test, with the same Chinese pair, either in the same order or not, depending on randomization. This third point was analyzed separately, and used as a kind of double check. Since the subjects were told that many pairs were possibly somewhat synonymous, they might not have noticed the same pair being given twice. If they did not notice, but got the correct answer on both, then it would reinforce the findings of the experiment; if they did not notice and had opposite results for the two times a pair was used, then any positive results would be cast in doubt. To find out if the subject noticed or not, the question was asked at

Table 10.1 *The antonymic pairs tested*

ugly	měi	chǒu	cold	lěng	rè
long	dǎn	cháng	shallow	shēn	qián
many	dūo	shǎo	concentrated	dàn	nóng
thick	báo	hòu	micro	wēi	jù
big	dà	xiǎo	friend	dí	yǒu
sharp	rùi	dùn	few	shǎo	dūo
bright	liàng	àn	small	xiǎo	dà
coarse	xì	cū	narrow	kuān	zhǎi
fat	pàng	shòu	hard	yìng	ruǎn
soft	ruǎn	yìng	skinny	shòu	pàng
heavy	zhòng	qīng	macro	jù	wēi
weak	qiáng	ruò	strong	ruò	qiáng
wide	zhǎi	kuān	hot	rè	lěng
tall	gāo	ǎi	diluted	nóng	dàn
fast	kuài	màn	thin	hòu	báo
drunk	xìng	zùi	short	ǎi	gāo
joy	xì	bēi	high	gāo	dī
dry	gān	shī	fine	cū	xì
low	dī	gāo	slow	màn	kuài
quiet	xuān	jìng	sorrow	bēi	xì

the end of the test: "Did you feel any of the separate pairs you heard on the tape was the same as any other pair, though possibly not with the same internal order? If so, which ones?"

10.2.1.2. Part b

For the second part of the first experiment, a second group of ten native English speakers was given the same test as in Part a, except that a second recording, using a different native Mandarin speaker from Taiwan,¹⁸ was made in which the tones were reversed on the words in those pairs whose members differed in tone. This was done to see if the tones of the words could have been instrumental in the choices made by the first group of subjects.¹⁹ This is of course assuming that the second group of English speakers would have roughly the same results as the first group if given the exact same test.

10.2.2. Results

The number of right and wrong answers²⁰ from Part a out of the total non-"neither" responses were calculated and are presented in line 1 of table 10.2. The

Table 10.2 Results of Experiment 1

Type	No. of responses	Right	Wrong	
Part a	359	190	169	($p < 0.30$)
Part b	337	187	150	($p < 0.05$)

Table 10.3 Comparison of items from Parts a and b

word	Part a	Part b	Part a result	Part b result
big	dà	dǎ	9/0	4/4
coarse	cū	cù	6/2	9/0
soft	ruǎn	ruàn	9/1	2/6
wide	kuān	kuǎn	1/8	6/1

subjects in general only noticed one or two repetitions, and were not sure which ones had been repeated, so it seems this did not interfere with the results. The number of correct answers was approximately the same for both halves of those antonymic pairs that were split up and tested separately.

The results for Part b, calculated the same way as for Part a, are given in line 2 of Table 10.2.

These results tell us that tone could have been an important criterion for the subjects' judgments.²¹ It seems here that the change in tone made it easier for the subjects to pick the correct Chinese word. This could mean either that the tones assigned to some words are not the ideal ones from an English speaker's point of view, or that something about the difference between the two recordings or the two groups of subjects influenced the results.

For some pairs the difference before and after the tone switch was quite startling. Table 10.3 contains some of the pairs that experienced the biggest change in response. The number to the left of the slash in each column is the number of right responses, and to the right is the number of wrong responses.

Some of the differences here match up with the results from Part a of Experiment 2 (see 10.3.2.1): as the high level tone is favored for the "small" category, the better scores for *coarse* and *wide* follow logically. The results for *big* also follow logically from our results showing a preference for the falling tone for the "big" category, but for *soft*, the effect is the opposite. The better score for *soft* before the change to a falling tone seems to contradict the tendency for the use of the falling tone for "big" category words, unless the "big" category is not a monolith, and can

be subdivided into several different semantic subgroupings, such as "soft," "warm," and "big." In this case, it seems the falling tone might be less suitable for those words connoting softness.

10.3. Experiment 2

10.3.1. Method

10.3.1.1. Part a

The second experiment was designed to test the subjective sensitivity of Chinese speakers to sound symbolism. In Part a, a list of 48 English words that are all members of antonymic evaluative pairs²² was printed on the answer sheets, with a blank line next to each word. A separate group of 50 nonsense syllables was taken from a list of all possible Chinese syllable types (from Zhong 1980) and printed in five columns of ten words each. They were chosen because they were simple CV syllables, with the vowels all monophthongs and the consonants all stops or fricatives, all of which could be easily classified as [\pm nasal] and/or [\pm grave]. No recording was made of the words.

The subjects were five native Mandarin-speaking graduate students from Taiwan. They were instructed: "Please say these words to yourself, then assign these words to the meanings given on the answer sheet, based only on how appropriate you feel they sound for the given meaning. Also assign tones to the words, based on the same criterion."

10.3.1.2. Part b

For Part b, a list of 25 Cantonese words or phrases was compiled from Chao 1947, Whitaker 1955–1956, Yue-Hashimoto 1972, and Kam 1977, from their lists of words and phrases that can be either made diminutive/familiar or, in the case of adjectives and adverbs, changed with regard to intensity by a change in the pronunciation to a high- or high-rising tone. Questions were then formed to contrast the semantic differences between the minimal pairs. Each question was then recorded (spoken by myself), followed by the relevant minimal pair, the order having been randomized, spoken by a native speaker of Cantonese from Hong Kong.²³ The subjects, all native Mandarin speakers from Beijing or Taiwan who had no knowledge of Cantonese, were told that they would hear a tape of Cantonese words, and that they should answer only "A" or "B" after hearing each pair, based on which one they thought was the best answer to the question asked.

10.3.2. Results

10.3.2.1. Part a

The 48 English words used for this part of the experiment were divided into "small" category words and "big" category words. Out of the total 210 choices (initial and vowel) made by the subjects for the "big" category words, grave segments were chosen 151 times, much greater than chance probability ($p < 0.001$), but for the "small" category words, grave segments were chosen only 103 times, no better than chance probability. The number of grave initials and grave vowels chosen by the subjects for each category are given in table 10.4.

The results show a clear tendency on the part of the subjects to assign grave consonants and vowels to "big" category meanings, and not to the "small" category meanings. This agrees with the findings of Solomon 1959: 494, which showed that subjects' judgments of "heavy" or "high magnitude" relate to the lower octave bands, while judgments of "light" or "low magnitude" relate to the higher octave bands. It also correlates well with the findings of Huang *et al.* 1969, which showed that when asked to produce words meaning "large" or "small," subjects produced words that overwhelmingly favored the acute vowels for "small" and grave vowels for "large."

Two other measurements were done, one on the nasals and one on the tones. A count of the number and type of nasal used for each category is given in table 10.5.

From this table we can see that the subjects made a clear choice of the grave nasal for the "big" category, and the acute nasal for the "small" category.

The number of times each tone was used for each category is given in table 10.6. The numbers 1, 2, 3, 4, and 5 refer to the high level, high-rising, falling-rising, falling, and neutral tones respectively.

We can see at least two tendencies here. One is the preferred use of the high level tone for the "small" category meanings, and the other is the preferred use of the falling tone for the "big" category meanings.

It is interesting to note that the tendency on the part of the subjects to use the high-level and falling tones more than the other tones correlates well with the findings of Yue-Hashimoto 1980 and Li and Thompson 1977. These studies showed that the high-level and the falling are the first two tones to be mastered by a child learning to speak Mandarin as its native language. Tse (1977) has reported that for Cantonese speakers as well, the high-level tone is the first to be acquired. It would be hard to determine if this is because this tone is easier to learn and produce (the "Difficulty Hypothesis"²⁴ of Li and Thompson), or because it is compatible with the size and defenselessness of the child (see discussion of "frequency code" below).

Table 10.4 Analysis of grave segments

Type	No. of responses	Grave initial	Grave vowel
big	105	71	80
small	105	52	51

Table 10.5 Comparison of nasals

Type	No. of responses	No. of [m]	No. of [n]
big	23	17	6 ($p < 0.01$)
small	29	9	20 ($p < 0.01$)

Table 10.6 Results for each tone-type

	N	1(55)	2(35)	3(214)	4(51)	5(var.)
big	105	13.3	16.2	15.2	50.5	4.8 ($p < 0.001$)
small	105	35.2	16.2	16.2	20.9	11.4 ($p < 0.01$)

10.3.2.2. Part b

To analyze the results of this part of the experiment, the questions were divided into four types, depending on what was being tested: (a) the hypocoristic (i.e. size-to-sound and monosyllabic familiarity); (b) "onomatopoeic" intensification; (c) disyllabic familiarity; and (d) intensification of reduplicated adverbs. Out of the 150 responses to type (a) questions, 87 were the same as they would be for a native Cantonese speaker (the definition of "correct") ($p = 0.05$). If we break this down into the subtypes, we see that of the size-to-sound responses, only 51 out of 100 were correct. Of the 50 responses to the questions testing monosyllabic familiarity, 36 were correct ($p < 0.01$).

Fifteen responses out of the twenty for type (b) pointed to a feeling on the part of the subjects that the unmodified form was the more intense of the two ($p < 0.05$). As I pointed out in section 10.1.3.2, this is the same for some native Cantonese speakers. Twenty-one of the 40 responses were correct for type (c).

The results for type (d) were much more straightforward and strongly support the universality of the feeling for the high-low pattern as an extreme intensive. Thirty-four of the 40 responses were correct ($p < 0.001$).

