

17 The Chinese character in psycholinguistic  
research: form, structure, and the reader

*Douglas N. Honorof and Laurie Feldman*

**Introduction**

Theories of lexical access and orthographic processing must accommodate an obvious difference between alphabetic/syllabic systems and logographic systems: the componential units of the latter encode semantic information. Although linguistic history has known several logographic and hybrid-logographic systems (e.g. Ancient Egyptian, Sumerian cuneiform, Mayan, Sino-Vietnamese, and Sino-Japanese), written Chinese stands alone as the only system that is both primarily logographic and the chief written system of a living language.<sup>1</sup> Despite a “phonographic tendency,” novel Chinese characters are not freely formed from existing components; Chinese is not “spelled.” Furthermore, the existing phonographic representation is incomplete; phonological information is encoded only loosely at the level of the component, only in some characters, and in phonological units that span a greater distance than the phoneme. The looseness of the phonological mapping arises in part because phonographic Chinese characters represent the multiple phonologies of the speakers who standardized them thousands of years ago. Nevertheless, present-day readers of Chinese manage to cope with the extreme fossilization of its ancient written forms across centuries of phonological change. Therefore theories of reading and word recognition must account for the apparent success of the Chinese reader. In the present chapter we discuss the foregoing issues and their implications for research on the reading of Chinese.

\* During preparation of the present chapter, the first and second author received support from NIH Grant DC-02717 and NIH Grant HD-01994, respectively. They thank Yi Xu for very helpful comments on an earlier draft.

<sup>1</sup> The term logogram (Du Ponceau, 1838: 110, cited in Martin, 1972) was popularized by Bloomfield (1933: 285–286) and Gelb (1963: 61), but is potentially misleading; Chinese does not encode words (Greek *logos*) or morphemes. Rather, it encodes syllables, typically hinting at meaning and sound, but it does so in a somewhat arbitrary mnemonic way. That is, several characters can be used to write a single syllable (with the consequence that Chinese has far more graphemes than syllables), but each of those “homophonous” characters may stand for several semantically unrelated items. Furthermore, some characters cannot be used except in combination with other characters, so the syllable-encoding grapheme is not always coextensive with the morpheme. See DeFrancis (1989: 89–121) and Xu (1992).

Specifically, we identify several issues that arise when one investigates the processing of Chinese characters, issues that are not necessarily obvious to the non-Chinese psycholinguist. First, we describe the Chinese character in linguistic terms in order to dissuade non-Chinese from equating single characters with words. Then we consider the implications of dialectal variation and bidialectalism for the processing of phonological aspects of Chinese characters. Here we note that a single set of (written) characters can capture more than one phonology, and that none of these phonologies is very close to the archaic Chinese phonologies for which the characters were devised over time. By describing the component structure of Chinese characters and implications for processing, we aim to prepare the non-Sinophone psycholinguist for access to the experimental literature on Chinese.

### Are Chinese characters like English words?

#### *Compounding of single character bound forms*

According to the Modern Chinese frequency dictionary (Beijing Language Institute, 1986), a reference work based on a 1.3 million-word corpus, 34 percent of words in Mainland Chinese texts are composed of a single character. The figure from Taiwan is most likely comparable. The modern Chinese language is highly polysyllabic (and especially disyllabic). Indeed we cannot know whether archaic dialects were less polysyllabic, but the written language of the classics is more monosyllabic than any known spoken variety of the language. That is, the older texts show a more nearly one-to-one mapping of morpheme, word, and character. However, we must caution non-Sinophone psycholinguists against equating character access for parts of two-character compounds with lexical access for words in languages such as English. In fact, where Standard Chinese (SC) is concerned, parts of many compounds tend to stand for single bound morphemes, not necessarily words. If the reader's lexicon is largely structured through exposure to the spoken language before he or she has learned the more monomorphemic written (especially, classical) language, it is unlikely that the reader of isolated bound forms always accesses them as words.

#### *Parsing of characters into words*

The task for the reader of Chinese two-syllable words and the English reader of polymorphemic words is also different in another way: where English polymorphemic words are nearly always run together (or in the case of compounds, often hyphenated), Chinese is never parsed into words with spaces or hyphens. The English reader need not group English morphemes into words online; this is done for him or her (with graphically misparsed exceptions *hot dog*, *paper cut*,

etc.). The hyphenation or running together of English morphemes reveals word status (and often reflects the stress pattern). In the case of Chinese, however, limiting experimental stimuli to only two-character words makes the parsing problem go away because the reader expects only two-character words or non-words. Nevertheless, limiting the stimuli in this way does not make the Chinese two-character words analogous to English polymorphemic words. Chinese two-character words remain words made up of potentially bound forms.

In the following sections, we consider the internal structure of the individual characters that are separable (that is, not bound) or that can make up compound words.

### Classification of Chinese characters

Crude but conventionalized forms of the basic building blocks of today's Chinese characters emerged at least as early as the the Shang Dynasty (~1700–1027 BC), as attested by divinatory inscriptions on the ventral shells of turtles and on animal bones. These writings are known as the "Oracle Bone Inscriptions" (甲骨文), and were very likely constrained in shape by the demands of the media on which they were carved (Boltz, 1986). By the time of the Bronzeware Script (金文) of the Zhou dynasty (~1122–256 BC), the logograms had begun to more closely resemble modern-day Chinese characters. While it appears that these early Chinese characters evolved out of simple pictograms, there are only a few pictographic Chinese characters still in use (for example 木 meaning "wood" from Old Chinese for "tree"), as well as a slightly larger number of ideograms representing more abstract concepts in graphic form (for example, 上 and 下 for "up" and "down," respectively). Today, most characters are assembled from more than one conventionalized multistroke component and the components tend to be less transparently iconic.

#### *Visual similarity*

In priming experiments, researchers often match primes with respect to visual similarity and then manipulate the presence or absence of a shared component. Chinese typography makes no distinction between cases. Thus, as in studies of Hebrew reading and processing, for example, manipulations of overall similarity in Chinese normally entail changes in font size or an alternation between typefaces. Raw differences in visual complexity among characters extend far beyond font and size parameters and are obvious to the fluent reader of Chinese, but to the naïve outside observer, Chinese characters may all look rather alike. As we shall see, if one knows something about the structure of a character, the differences are easier to detect. In the following sections, we delve deeper into the nature of those differences in visual complexity among characters.

*Stroke count*

Even the simplest characters can themselves be broken down into conventionalized “brush” strokes (indeed, each with its own conventional name; see figure 17.1). Some studies have focused on whether stroke count can influence character recognition (see Flores d’Arcais, 1992; Leong, Cheng & Mulcahy, 1987; Taft & Zhu, 1997; Tan & Peng, 1991). Arguably more important to studies of word recognition, real characters can differ by as few as one stroke. Indeed, experimenters often transform real characters into pseudocharacters by distorting as few as one or two strokes rather than by creating an entirely new component. However, it is unclear whether a difference of one stroke has the same effect on relatively simple and relatively complex characters.

Although some characters consist of only one (multistroke) component, most are internally structured of more components (for example, 穩 meaning “stable” with five, which is decomposed for the nonreader of Chinese in figure 17.2). Two characters with roughly the same number of strokes may appear to be roughly equally complex to the non-Chinese reader, but one may have an added layer of complexity introduced by its more complex internal structure. Even counting the number of components may not tell the entire story; some full-fledged components can, in turn, be subdivided into additional nested components. For example, the character 響, meaning “sound,” is composed of a component (on top) and a component (on the bottom), but the top component is in turn composed of three independent components (left-to-right) and the bottom component is composed of two independent components (top-to-bottom).

*Position-based distortion*

Evaluating visual complexity is not always a straightforward matter. Many forms vary in conventional ways according to their position within a character. Sometimes only a linear scaling is involved, but at other times component shapes may be morphed from their full-character forms in number of strokes, axial distortion, center of gravity, and obliqueness of their horizontal and near-horizontal strokes. Indeed, there may be several position-based variants for such functionally equivalent components. In some cases, the differences are so great that relatedness among the forms would be unrecognizable to the learner unless it were explicitly pointed out, so the connection between them and their full-character-in-isolation forms must be memorized. For example, 彡, known as the “three dot water,” appears on the left-hand side of a left-right character and indicates a water-related meaning for the character in which it appears, but takes the very different form 水 as a character in isolation. In other positions, the water “radical” (a component used by lexicographers indexically) has the form of the full character but with distorted dimensions as in 漿. Variants can be

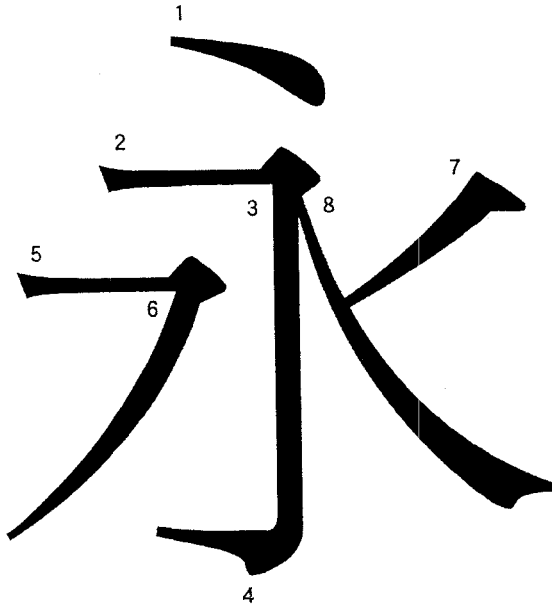


Figure 17.1 Brush stroke names. Students of Chinese calligraphy may improve by repeatedly tracing and copying the character 永 (SC: /jǒŋ3/ “forever”) and the eight basic stroke types it contains. Several of these stroke types (2–4 and 5–6) are written with distinct gestures yet without the brush leaving the paper. Thus the character 永 is looked up in the dictionary under a stroke count of five, not eight. Stroke types are named below in Chinese characters accompanied by Hanyu Pinyin (a system of romanization for Standard Chinese) and a rough translation of the name. Numbers correspond to the strokes indicated above. 1. 點 *diǎn* “dot.” 2. 橫 *héng* “horizontal” (drawn left-to-right). 3. 豎 *shù* “vertical” (drawn top-to-bottom). 4. 鉤 *gōu* “hook” (appended in one of several positions). 5. 挑 *tiǎo* (a.k.a. 提 *tí*) “rise” (drawn bottom-left-to-top-right, but here almost horizontal). 6. 撇 *piě* “fall” (drawn from top-right-to-bottom-left). 7. 短撇 *duǎnpiě* “short fall” (drawn from top-right-to-bottom-left). 8. 捺 *nà* “fall” (drawn from top-left-to-bottom-right).

even less visually similar; 邑 (the character) and 阝 (the component), for example, look entirely unrelated and share no strokes in common.

#### *Traditional versus simplified characters*

Traditionally, there are 540 components (部件 *bùjiàn*) that date back to Xu Shen’s AD 121 說文解字 (*Shuowenjiezi*) compendium. Of these *bùjiàn*, 214 部首 *bùshǒu* form a subset which, with minor adjustments, have been used

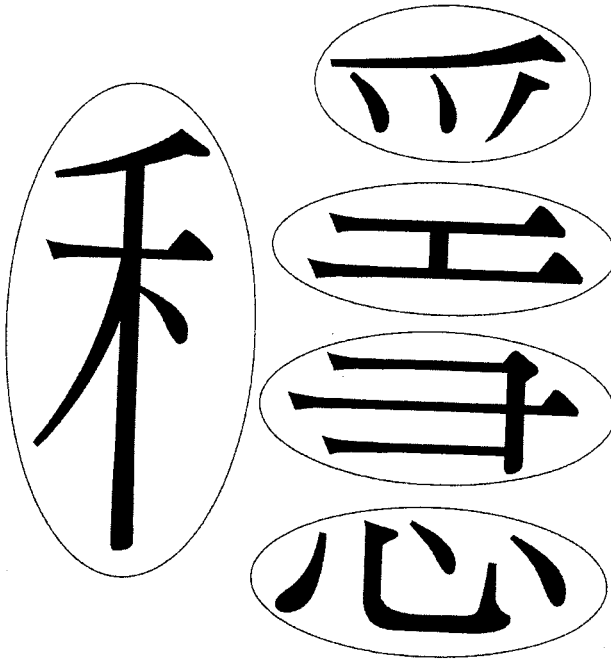


Figure 17.2 Circles indicate which strokes group together into a single component within the larger five-component character for 稊 (SC: /wɛn3/"stable").

indexically by many lexicographers since Mei Yingzuo's 字彙 (*Zihui*) of 1615. Chao tells us (1948: 64) that, of these 214 components, the twenty highest-frequency components index more than half of the characters. (See [www.unicode.org/charts/unihanrsindex.html](http://www.unicode.org/charts/unihanrsindex.html) for a list of indexical components with English translation.)

In Mainland China, a number of the *bùshǒu* (54) and a larger number of entire characters (over 2,200) have been simplified in visual complexity. Some of the simplified forms are based on 草書 (SC: /ts<sup>h</sup>aŋ3.ɕu1/), that is, "Draft Script," an ancient and extremely nonrepresentational form of cursive script, or on the more representational cursive, "Running Script" (行書, SC: /ɕiŋ2.ɕu1/). Most traditional and simplified characters are identical except for minor differences in typeface. However, in some cases, the simplification involves a reduction in the number of strokes that comprise a component, a reduction in the number of components, and/or a change in choice of component. For instance, the character 講 ("speak") is simplified as 讲. Here the so-called "speech radical" (言) appears on the left but is reduced to two strokes (讠), while the component on the right is actually a different (and simpler) component.

In general, because the number of strokes tends to be reduced from an average of six or seven to an average of four or five in simplified forms (Wang, 1971), some characters have become more similar to each other under simplification. The levels of simplification for a given character may be compounded if it contains multiple simplified components. In some cases, the simplifiers have even chosen to conflate two distinct characters into one (for example, 乾 "dry" and 幹 "to do" have both been reduced to 干). Because traditional components tend to look more like full-form-in-isolation characters to which they are related (where relevant), the reader who has mastered traditional characters may *ipso facto* have a slightly greater awareness of (historical) character-component relatedness, but the consequences of use of simplified versus traditional characters in psycholinguistic experiments have yet to be determined.

The choice of traditional versus simplified characters is a distinction available in Chinese but not in English. The use of one type of character or the other is dictated by the origins of the participant population: generally, traditional characters are used by readers from Taiwan (and perhaps still in Hong Kong), while simplified characters are used by readers from the mainland and some other East Asian countries with a substantial Chinese population such as Singapore and Malaysia. Until the twentieth century, most works were written top-to-bottom, then right-to-left, and even now there are presses that set type vertically. (There is a positive correlation between vertical typesetting and use of traditional characters.) Vertical presentation of multiple characters in psycholinguistic experiments is rare, and is said to be disadvantageous to the reader (Chang, 1942: 55).

### *Component order*

Visual complexity is not the only visual parameter that might help a reader process a Chinese character. Individual components within a character are written in a highly constrained order that is said to reduce smearing of ink (assuming right-hand-dominant writing practice). The actual rules are slightly complicated, but can be summarized as follows: characters are written top-left to bottom-right, with horizontal strokes coming before intersecting vertical strokes and center strokes before their embellishments, small strokes often being written at the very last. The constraints on the order in which characters are written raise intriguing questions about the order in which components are processed in reading, and about whether processing proceeds from left to right (see Taft & Zhu, 1995).

### *Components and their functions*

There are many characters, each with its own meaning and pronunciation, that can also appear as components of more complex characters. When such components function to provide information about the meaning of the larger

character in which they appear, we refer to them as “separable semantic components”; *separable* because they also appear as characters themselves *separately* and *semantic* because they contribute to the general semantic field indicated by the full character. For example, the character, 女, meaning “woman,” appears as a semantic component on the left in 姥 (“maternal grandmother”), though it is more prolate in shape, that is, vertically elongated, as a component on the left of this character than when it appears as an independent character. Position-based distortion of component shapes has been discussed above. Some researchers refer to these components as 形符 (SC: /çiq̄ 2fu2/), that is “[shape-based] semantic radicals/classifiers/determinants,” because they encode meaning in order to help disambiguate homophones (or near-homophones).

Not all separable components are used for their semantic value; sometimes the components indicate something about the pronunciation of a larger character as a whole. For instance, such characters as 可 (SC: /kɤ3/) and 史 (SC: /ʂz3/) can appear as components, but when they do, they do not indicate meaning; rather, they suggest that the characters in which they appear are pronounced something like /kɤ3/ or /ʂz3/, respectively. Such “separable *phonological* components” derive their phonological cueing value from their pronunciation as characters in isolation. Some researchers refer to these components as 聲符 (SC: /ʂeq̄ 1fu2/), that is “phonetic radicals/classifiers/determinants,” because they use sound to help disambiguate synonyms.

Approximately 80–95 percent of characters (see Alleton, 1970: 33; Zhou, 1978; Zhu, 1988, cited in Feldman & Siok, 1999a) are made up of a combination of (at least) two components, one or more serving a semantic function (most often appearing on the left) and one serving a phonological function (most often appearing on the right). Such characters are called 形聲字 (SC: /çiq̄ 2 ʂeq̄ 1 tsz4/; also, *xiesheng* or *xiangsheng*), and are translated into English as [semantic-] *phonetic compounds* following the definition introduced by Xu Shen (121).<sup>2</sup> In the majority of cases, the semantic component is smaller in size than the phonological component (Martin, 1972: 85), though the implications for the reader of the size differential are unknown.

Although the individual semantic-phonetic compound character is compositional, it is more compositionally complex than a simple rebus which gives only (homo)phonic information on the basis of graphic forms. (A rebus might

<sup>2</sup> The other five origin-based categories (書) of character structure defined by Xu are: (1) pictographs (象形) such as 木 (archaic: “tree”) already mentioned; (2) simple ideographs (指事, 處事, 象事) such as 上 (“up”) and 下 (“down”), also already mentioned; (3) compound ideographs (會意, 象意), for example, 明 (“bright”) which is made up of 日 (“sun”) and 月 (“moon”) and represents a very rare type; (4) circularly interpreted [phonological] loans (轉注) in which a new character develops out of an older synonym (with a similar pronunciation) in order to distinguish the two as in 返 and 還 and (5) false loans (假借) such as 足 meaning “foot” being used for the homophone “enough.”



present the picture of an eye to indicate the English sound for the word “I”.) For example, the character 詭, meaning “a lie” and pronounced /xwɑŋ 3/ in SC, is composed of 言 and 亢, and the odd *bùshǒu* 艸. In this case, the component on the left has the meaning, “words,” and the component on the bottom-right suggests the SC pronunciation /xwɑŋ/. Characters such as 詭 are thus clearly compositional and it is fairly easy to see a transparent connection between the meaning and pronunciation of the character as a whole and the functional components 言 and 亢 which are fairly consistent across characters. Fairly is the operative word; transparency/opacity and consistency/inconsistency are always relative. The odd *bùshǒu* 艸 never appears in isolation, but is considered to have the same meaning as the character 草 (SC /ts<sup>h</sup>aŋ 3/), “grass.” The component is thus commonly known as the “top of the character 草.” In the example, the component 艸 adds little or nothing transparent to the modern meaning of the whole character and nothing to its pronunciation even though it is associated with a character that has a meaning and pronunciation of its own. In other characters, the “top of the character 草” may serve a semantic encoding function (e.g. 苗 “seedling,” 花 “flower”), but it never appears separately as an independent character. We refer to such components as “bound components” because they appear only as components, never separately as independent characters. There are also bound components that never encode semantic or phonological information, for example, the component 冫, which is known as “the cover for the character 包.”

A subset of the *bùshǒu*, both separable and bound, can serve semantic *or* phonological functions depending on the characters in which they occur. For example 火 (“fire,” SC /xwǒ 3/), a separable component that most often encodes semantic information, can alternatively be used to encode information about the pronunciation of a character in which it appears (伙 “companion,” SC /xwǒ 3/). We refer to these components, of which there are approximately forty, as *functionally bivalent*. In a few characters, such functionally bivalent components can even encode both functions at once (e.g. 米 “rice,” SC /mi 3/ appears as the bottom component in 糜 “gruel,” SC /mi 2/).

Clearly, the potential contributions of semantic and phonological components to the meaning and pronunciation of semantic–phonetic compounds are complex (in the present volume, see chapters by Chen et al.; Liu et al.; Peng & Jiang; Perfetti & Liu; Shu & Wu; Taft; Tan & Siok). In experimental studies that entail priming of a target by a prime that precedes it, when two characters share a component and that component serves the same *function* in both characters, we call this relationship F+. Facilitation in character decision tasks for pairs of words that share a semantic component is well documented (e.g. Feldman & Siok, 1999a, b; Zhou & Marslen-Wilson, 1999). Likewise, there is evidence for facilitation in character naming for pairs of words sharing a phonological component (Fang Horng & Tzeng, 1986; Feldman et al., 2004; Hue, 1992;

Seidenberg, 1985). When two characters share a component and that component serves a different function in each character, we term this relationship F–. We have evidence for F– inhibition in the naming task (Feldman et al., 2005).

There are many components that are themselves composed of nested sub-components appearing elsewhere as full components on their own. It is not known whether nested sub-components are decomposed by the reader in relation to their meanings and pronunciations, but clearly, at the larger full-character level, decomposition is a useful strategy given the number of characters in which components serve a specific encoding function.

### *Layout/position*

Most semantic–phonetic compound characters are composed of a semantic component on the left and a phonological component on the right, therefore position and function tend to be highly correlated. In our description, we have differentiated between components with respect to function, but it is not uncommon for researchers to ignore a component's ability to function semantically, phonologically, both or neither. For example, Taft and Zhu (1997) used a frequency count taken from the *Chinese Radical Position Frequency Dictionary* (Hunan Institute of Computer Science, 1984) to classify components as high or low in frequency. This count reflected the tendency of phonological, semantic, and other graphic components to enter into compound characters. They observed differences in character decision latencies as a function of frequency for components appearing on the right. Others have suggested, however, that the outcome may reflect the manner in which frequency was measured and the tendency within a Chinese compound for the position and the function of a component to be interdependent (Feldman & Siok, 1997).

In some cases, it is the identity of the component itself that leads to this correlation between position and function; for example, the semantic component 扌 (meaning “hand” or “actions related to hand”) only ever appears in this form on the left of a character (e.g. 打). Clearly, however, the correlation is not perfect. Looking across components, there are multiple positions in which components may appear: left, right, top, bottom, central, peripheral, in one quadrant, partially surrounding, etc., and even a single component may appear in more than one position irrespective of function. Thus, in some semantic–phonetic compounds, the semantic and phonological components appear reversed or in non-left-right positions. For example, the semantic component, 山 (meaning “mountain”), can appear on the top (崇), at the bottom (岱), or on the left (岐). Phonological components may deviate from their canonical right-hand positions less often than semantics do from their canonical left-hand positions, but even they can be found elsewhere (e.g. on the bottom as in 芸, on the top as in 警, enclosed within another component as in 圓, etc.). The correlation between position and function

is also slightly weakened by the functionally bivalent components (discussed above) as they can appear in the same position irrespective of function. For example, 禾 ("grain," SC /xɿ2/2) functions as a semantic component on the left in 種 ("plant," SC /tʒoŋ4/) but as a phonological component on the left in 科 ("division," SC /kʰɿ1/).

By implication, when two characters are paired in a primed character recognition experiment, for example, they may vary in layout (left–right versus top–bottom, etc.). When layout is shared, we term the relationship L+. Furthermore, when layout is shared, we may go on to manipulate the position in which the repeated component appears. When the repeated component appears in the same position in both characters, the relationship is said to be P+. Taft (this volume) has argued that facilitation in character recognition can only arise when a component preserves the same position (and therefore layout) in prime and target. In contrast, we have some evidence that the semantic component does produce facilitation when the position and layout vary between prime and target (see Feldman et al., 2004).

### *Semantic transparency and consistency*

*Transparency* is a measure of the contribution of a specific semantic component to the meaning of a character in which it occurs. Assessing the transparency of the relationship between a component and a character presupposes knowing the semantic field of the components itself. For components that can appear as characters in isolation on their own, the semantic field of the component is clear, even though it may have more than one meaning across characters, and even though those meanings need not be closely or obviously related. However, as discussed above, there are also bound components that are not themselves characters. Bound components must take their meanings from the characters in which they appear much as bound morphemes in English (e.g. *spect*) take their meanings from the words in which they appear (e.g. *respect*, *inspect*, etc.). We rely on the concept of semantic *consistency* to capture the relatedness in meaning among characters that share a semantic component. Just as with English, some instances of a component will be more related to the meaning of the whole item than others (e.g. *spect* and *respect* versus *spect* and *spectacular*; the latter pair is more related for many native speakers irrespective of etymology).

## **Phonological considerations**

### *Phonological transparency and consistency*

Transparency and consistency are also relevant measures of the phonological cueing function. When phonological components are separable, they take

their cueing functions from their pronunciation as characters in isolation. Some phonological components cue only one pronunciation in isolation and in larger characters (采, *cai3* in SC, for example), whereas others can vary in pronunciation from their form in isolation but nevertheless have a prototypical pronunciation. A character whose pronunciation does not closely match the pronunciation suggested by its phonological component is phonologically opaque (that is, irregular). Moreover, when phonological components are bound, they can only derive their pronunciations by association across the multiple characters in which they occur, and those pronunciations can be very inconsistent across characters. Consequently, many Chinese characters are comparable in phonological opacity to the worst cases in English (e.g. *cough*).

Even within a single dialect, phonological values for a single component can be very inconsistent, even for separable phonological components. Fan, Gao and Ao (1984) report that only 26.3 percent of semantic-phonetic compounds have a pronunciation identical to that of their phonological component in isolation; the majority of phonological components, on the other hand, may cue a variety of initials, finals (including any medials) or tones in the various characters in which they appear. In fact, it can sometimes be difficult to say with any degree of certainty which pronunciation is most strongly cued by a given phonological component, even in cases in which a component appears as a character itself in isolation with its own pronunciation. For example, the phonological component 是, which in isolation is the copula with the SC pronunciation /sz/, also occurs consistently as a phonological component in characters pronounced either /ti/ or /t<sup>h</sup>i/ (and rarely /sz/) in SC.

### *Synchronic and diachronic bidialectalism*

While the Chinese writing system continued to develop new semantic-phonetic compounds that encoded both semantic and phonological cues over the course of many dynasties, the logographic principle and most of the forms themselves have been very stable since the Seal Script (篆書, SC /tʂwan4 ʂu 1/) of the Qin Dynasty (circa 221–210 BC). For any language, the divergence of a millennia-old written system from many evolving spoken systems increases the complexity of the mapping between print and speech. This opacity can pose problems for the reader of an alphabetic script, but the challenges are potentially greater for the Chinese reader both because the writing system was fossilized so long ago and because it was only partially phonologically transparent early on.

Although many very old Chinese characters can be read aloud today in Standard Chinese, those same characters can be read aloud with a different pronunciation in other Sinitic dialects. In fact, Chinese characters are even used to write dialects whose pronunciations are diverse to the point of mutual unintelligibility such that they would be considered separate languages under different

circumstances (see Cheng, 1994). Across dialects, pronunciations for a given component may differ in the voicing/aspiration of the syllable onset, the presence/absence of a medial glide, the precise phonetic specification of the nuclear vowel, the presence/absence or identity of a syllable coda, the tone or a combination of the above. For example, according to the *Hanyu Fangyin Zihui* (Beijing Daxue, 1989: 149), abstracting across tones, the component 𠂇, which appears on the bottom of the character 窄 ("narrow"), suggests the SC pronunciation /tsuo4/, /tsa/, /tʂa/ or /tʂai/. The character as a whole is pronounced /tʂai/ in Beijing (Northern Mandarin), /tsɿ/ in Wuhan (Southwestern Mandarin), /tsia/ in Shuangfeng (Xiang/Southeastern Mandarin), /tsa? / in Fuzhou ([North]eastern Min) and Suzhou (Wu), /tset/ in Nanchang (Gan/Southeastern Mandarin), and /tʂak/ in Guangzhou (Yue).

Despite the use of the same characters with different pronunciations by readers and writers of divergent dialectal systems, the written form is tenacious and stable. While it is true that we now see dialects, (viz. Shanghaiese [Wu], and especially Cantonese [Yue]), in which a handful of characters have been devised to represent dialect-specific words that do not appear in a Northern dictionary, by and large, the ancient writing system remains in tact throughout China; this despite the centrally imposed writing simplifications of the previous century. Clearly, the dialect background of participants in psycholinguistic experiments (and thus the mapping between print and dialect) cannot be ignored because there is now near consensus that skilled readers access phonological information in the course of processing many Chinese written forms, at least for lower-frequency items (Perfetti, Zhang & Berent, 1992; Tan & Perfetti, 1997; Zhou & Marslen-Wilson, 1999). Some studies focus on homography among whole character forms usually without reference to tone (e.g. Perfetti & Tan, 1998; Seidenberg, 1985; Spinks et al., 2000; Xu, Pollatsek & Potter, 1999) while others concentrate on facilitation due to a phonological component that recurs in more than one character, in our terms,  $P_{\pm}$  (Zhang, Perfetti & Yang, 1999). For either type of phonological manipulation, the native pronunciation of the reader may influence the outcome. To complicate matters further, owing to exposure to multiple dialects and inconsistent local educational practices, a single reader may control multiple dialects, and give different dialect readings to the same characters when reading different types of literature.

In contrast to many languages where one can identify an ideal dialect in terms of the closeness of its mapping to the written language, the choice of ideal participant-dialect is less clear in the Chinese context because it has never been shown that the relation between phonological components and the characters in which they appear is more systematic in one modern dialect than in another. In fact, often, we are not really choosing between speakers of one or another dialect, but between speakers of one or another *set* of dialects because bidialectalism and diglossia are on the increase among ethnic Han Chinese

(Norman, 1988: chapter 10). Thus knowledge of more than one phonological system and the fossilization of ancient written forms across centuries of phonological innovation and divergence have conspired to produce a situation in which a single participant in a psycholinguistics experiment may have two (or more) options as to how to read aloud virtually all characters.

Bidialectalism is not limited to synchronic dialects or to phonological concerns. Chinese readers inherit knowledge about the lexicon of archaic and ancient Chinese through many channels, though they pronounce out-of-use characters with modern pronunciations. Familiarity could well have an impact on the interpretation of a character. For example, 目 and 眼 are archaic and contemporary characters for words meaning "eye," respectively. Note that the former is a component in the latter, but no longer stands alone to encode a modern word. Although exposure to archaic words is almost unavoidable in this highly traditional environment, it is more extensive for some readers than others. The general level of literacy and academic concentration may affect the analytic potential of the Chinese reader.

### **Final remarks**

The foregoing discussion was intended to clarify the distinction between the character and the word, the attributes of the character's components, and the complex relationship of the speaker to the phonology encoded in the character. We hope that we have made the literature more accessible for those who do not read Chinese, and provided a useful summary of the system for all working in the field of Chinese psycholinguistics.

VOLUME I CHINESE

THE HANDBOOK

# of East Asian Psycholinguistics

EDITED BY Ping Li, Li Hai Tan, Elizabeth Bates  
and Ovid J. L. Tzeng

GENERAL EDITOR Ping Li

CAMBRIDGE