Short Communication

Knowing Chinese character grammar

James Myers

Graduate Institute of Linguistics, National Chung Cheng University, 168 University Road, Min-Hsiung, Chia-Yi 62102, Taiwan

ARTICLE INFO

ABSTRACT

Chinese character structure has often been described as representing a kind of grammar, but the notion of character grammar has hardly been explored. Patterns in character element reduplication are particularly grammar-like, displaying discrete combinatoriality, binarity, phonology-like final prominence, and potentially the need for symbolic rules (X → XX). To test knowledge of these patterns, Chinese readers were asked to judge the acceptability of fake characters varying both in grammaticality (obeying or violating reduplication constraints) and in lexicality (of the reduplicative configurations). While lexical knowledge was important (lexicality improved acceptability and grammatical configurations were accepted more quickly when also lexical), grammatical knowledge was important as well, with grammaticality improving acceptability equally for lexical and nonlexical configurations. Acceptability was also higher for more frequent reduplicative elements, suggesting that the reduplicative configurations were decomposed. Chinese characters present an as-yet untapped resource for exploring fundamental questions about the nature of the human capacity for grammar.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

There is a vast experimental literature exploring grammatical knowledge (Cohn, Fougeron, & Huffman, 2011; Kawahara, 2011; Myers, 2009; Phillips & Wagers, 2007). Chinese characters also show systematic patterns, and researchers have often applied the term grammar to them as well (Kordek, 2013; Ladd, 2014; Sproat, 2000; Wang, 1983). Yet much more is known about the processing challenges that characters pose to learners (Chan & Nunes, 1998), readers (Honorof & Feldman, 2006), and writers (Chen & Cherng, 2013) than about the knowledge of character grammar per se.

The core of Chinese character grammar is discrete combinatoriality: characters are usually decomposable into smaller elements, which are often decomposable in turn. Most characters (85%, as estimated by Perfetti & Tan, 1999) are decomposed, at the first step, into a semantic component (associated with the meaning of the whole character) and a phonetic component (associated with the character's pronunciation). A typical example (from Ladd, 2014, p. 129) is 鐳 *léi* 'radium', decomposed into the semantic component 金 *jīn* 'gold, metal' and phonetic component 雷 *léi* 'thunder', where the latter, in turn, is composed of 雨 *yǔ* 'rain' and 田 *tián*

'field', which themselves share stroke complexes with 兩 *liǎng* 'two' and 十 *shí* 'ten', respectively. Ladd (2014) compares this hierarchical structure to the duality of patterning of Hockett (1960), with the deeper levels reminiscent of the phonology of spoken or signed languages.

Chinese readers are sensitive to the combinability of character components (Hsu, Tsai, Lee, & Tzeng, 2009), their typical positions (Taft, Zhu, & Peng, 1999), consistency in the pronunciation of the phonetic components (Lee, Huang, Kuo, Tsai, & Tzeng, 2010), and the overall arrangement of character elements (Yeh & Li, 2002). Writers, who need more detailed character representations than readers, are also influenced by the elements within phonetic components (Chen & Cherng, 2013). The processing challenges posed by orthography are partly universal; English reading is also influenced by stochastic orthographic patterns (Bailey & Hahn, 2001; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), and the brain's visual word form area is activated both by Chinese characters (Liu et al., 2008) and by alphabetic orthographies (Dehaene & Cohen, 2011).

In other words, the structure of information shapes how that information is processed. But orthographic structure is itself a product of the mind, and the phonology-like aspects of character structure go beyond mere combinatoriality. For example, the semantic component is typically reduced (shrinking or even losing strokes) when it is on the left or the top, but rarely when it is on the right or bottom (Myers, 1996). This pattern may relate to the order

in which strokes are normally written, from left to right and top to bottom, putting prominence on the stroke-final position, just as phrase-final syllables tend to be longer in speech (Beckman & Edwards, 1990) and in signed languages (Sandler, 1993). However, semantic components are closed-class (belying the common term *radical*, they are more analogous to affixes than roots), so it is difficult to test whether Chinese readers have active knowledge of these patterns.

This paper tests another phonology-like pattern, involving the configuration of reduplicated constituents (Kordek, 2013). Many characters contain horizontal configurations of two copies of the same element, like 朋 *péng* 'friend' (cf. 月 *yuè* 'moon, month'). Other reduplicative configurations are vertical, like 炎 *yán* 'inflammation' (cf. 火 *huǒ* 'fire'), while still others consist of three identical elements arranged in a upward-pointing triangle, like 森 *sēn* 'forest' (cf. 木 *mù* 'wood'). (A rare fourth type of reduplicative configuration, not studied here, forms a two-by-two square, as in 叕 *zhuó* 'join together'; cf. 又 *yòu* 'also'). Character element reduplication is partly iconic: a forest (森) contains more trees (木) than do woods (林 *lín*). Yet iconicity is not a reliable cue to whole-character meaning: 朋 and 月 are totally unrelated, 哥 *gē* 'older brother' relates to 可 *kě* 'may' only in pronunciation, and iconicity is irrelevant when it appears in phonetic components. Iconicity is also found in spoken and signed reduplication (Aronoff, Meir, & Sandler, 2005; Hurch, 2005), but these patterns nevertheless also conform to phonological constraints (Brentari, 1998; McCarthy & Prince, 1994).

Reduplicative configurations in Chinese characters obey formal constraints as well: triangular configurations cannot point downward (two elements over one), and horizontal and vertical configurations cannot contain three elements (the sole exception being 𤼵 *líng* 'raindrops', an archaic character that now only appears in 靈 *líng* 'spirit'). No such restrictions apply to mere strokes (e.g., 州 *zhōu* 'prefecture', 三 *sān* 'three') or to combinations of distinct components (e.g., the three 火 *huǒ* 'fire' in 熒 *yíng* 'glimmering' do not form an illicit downward pointing triangle because the two 火 at the top form a constituent with 冖 *mì* 'cover', also seen in 營 *yíng* 'camp'). These patterns are not merely formal, but reminiscent of familiar phonological constraints. In particular, reduplicative configurations obey binarity, just as stress feet are much more likely to be disyllabic than trisyllabic (Gordon, 2002). The upward-pointing triangles obey binarity both horizontally and vertically, while also making the configuration "bottom-heavy", consistent with the stroke-final prominence noted earlier.

Before speculating on how to explain such patterns, we must first establish that they are part of the active knowledge of Chinese readers. We thus ran an acceptability judgment experiment (Bailey & Hahn, 2001; Topolinski & Strack, 2009) using speeded binary (yes/no) responses (Weskott & Fanselow, 2011), testing fake characters that crossed grammaticality (e.g., upward-pointing vs. downward-pointing triangular configurations) with lexicality of the reduplicative configuration (i.e., whether the grammatical version of this configuration appears within real characters), while also taking into account character element frequency and visual complexity.

2. Method

2.1. Participants

The participants were 20 university students in southern Taiwan. All were native speakers of Mandarin Chinese with normal or corrected-to-normal vision. Participants were paid for their participation and provided written consent (among other things, to share their response data online).

2.2. Materials and design

Forty-eight sets of four fake characters each were created by editing traditional characters in Microsoft MingLiU font. All characters contained reduplicated elements in a horizontal, vertical, or triangular configuration (16 sets each). The four characters in each set contained a semantic component in its standard position, with the reduplicative configuration forming the remainder and crossing grammaticality and lexicality. In grammatical configurations, reduplication obeyed the constraints discussed above. Each grammatical configuration was paired with an ungrammatical one formed of the same element, but where horizontal and vertical configurations contained three repetitions and triangular configurations formed a downward pointing triangle. Grammatical configurations were called lexical if they also appear in real characters; matching ungrammatical characters were created as just described. In nonlexical configurations, the reduplicated element is never reduplicated in real characters. Sample materials are shown in Table 1 ('NA' indicates that reduplication of the given element, e.g., 支 *zhī* 'branch', is not lexically attested).

A three-way ANOVA on the log number of strokes showed effects of configuration shape ($F(2, 180) = 7.70$, $p < .001$), lexicality ($F(1, 180) = 4.15$, $p < .05$), and grammaticality ($F(1, 180) = 15.63$, $p < .001$), with an interaction between grammaticality and shape ($F(2, 180) = 3.85$, $p < .05$). These potential confounds with visual complexity are unavoidable because lexical reduplication, particularly vertical reduplication, favors simpler elements, and ungrammatical horizontal and vertical (but not triangular) configurations necessarily contain more strokes. Thus we included the log number of strokes as a covariate in the analyses described below.

We also calculated the type frequencies of the reduplicated base elements (e.g., 木). Character components were extracted with the help of a Wikimedia resource (Chinese Characters Decomposition, 2015) that recursively decomposes 21,170 traditional and simplified characters while also providing information about reduplicative configuration shape. Type frequencies were based on just the 6962 traditional characters in the Academia Sinica Balanced Corpus of Modern Chinese (Huang, Chen, Chen, & Chang, 1997).

An additional 120 fillers were created by editing other real characters (see Table 2). The fillers, all composed of real semantic and phonetic components in their standard positions but in novel combinations, were designed to vary gradiently in acceptability: 40 had no further modification (best), 40 added or removed strokes (worse), and 40 reflected an asymmetrical element vertically or horizontally (worst).

Test items were divided into four lists of 48 items each in a Latin square design, so that all four participant groups saw all test item types (defined by configuration shape, grammaticality and lexicality) but never from the same matched set. All participants saw all 120 filler items.

2.3. Procedure

The experiment was run with PsychoPy v. 1.82 (Peirce, 2007, 2009). Participants were told they would see a series of characters that were not real Chinese characters. They were asked to decide if they were like or not like Chinese characters by pressing, respectively, a key on the right or left side of a computer keyboard. Trials consisted of a 500 ms display of a fixation cross at the center of the screen, followed by 500 ms of a blank screen, and finally a fake character that remained at the center of the screen for 3000 ms or until the participant pressed one of the response keys, after which the next trial began. Characters were displayed in black on a white background, subtending approximately 4.5° vertically and horizontally from a viewing distance of 50 cm. Prior to the