**Beyond Quantity of Experience: Exploring the Role of Semantic Consistency in Chinese Character Knowledge**

Cheng-Yu Hsieh1, Marco Marelli2,3, and Kathleen Rastle1

1 Department of Psychology, Royal Holloway, University of London, Egham, UK

2 Department of Psychology, University of Milano-Bicocca, Milan, Italy

3 NeuroMI, Milan Center for Neuroscience, Milan, Italy

**Author Note**

Cheng-Yu Hsieh <https://orcid.org/0000-0002-2826-1435>

Marco Marelli <https://orcid.org/0000-0001-5831-5441>

Kathleen Rastle <https://orcid.org/0000-0002-3070-7555>

The data, analysis scripts and materials for this study are available at<https://osf.io/thfq6/>. We acknowledge funding to Cheng-Yu Hsieh from the Ministry of Education, Taiwan. The authors have no conflicts of interest to declare.

Correspondence concerning this article should be addressed to Cheng-Yu Hsieh, Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom. Email: Cheng-Yu.Hsieh.2021@live.rhul.ac.uk

**Abstract**

Most printed Chinese words are compounds built from the combination of meaningful characters. Yet, there is poor understanding of how individual characters contribute to the recognition of compounds. Using a mega-study of Chinese word recognition (Tse et al., 2017), we examined how the lexical decision of existing and novel Chinese compounds was influenced by two properties of individual characters: family size (the number of distinct words which embed a character) and family semantic consistency (the average semantic relatedness between a character and all words containing it). Results revealed that both variables influence word and nonword processing: words are recognised more quickly and accurately when they contain characters that occur frequently across different words and that make consistent meaningful contributions to those words, while nonwords containing those types of characters are rejected more slowly. These findings suggest that the learning of individual characters is based not only on the quantity of experience with them but also on the reliability of the semantic information they communicate. In addition, readers are able to generalise character knowledge acquired from previous word experiences to their daily encounters with familiar and unfamiliar words. We close by discussing how word experience shapes character knowledge when different ways of calculating family properties are considered.

*Keywords*: Chinese word recognition, family size, family semantic consistency, compound word processing, distributional semantic models

Beyond Quantity of Experience: Exploring the Role of Semantic Consistency in Chinese Character Knowledge

The Chinese language is written using characters, each of them representing a *morpheme*, or an orthographic unit associated with meaning(s). Most of these characters do not occur frequently in isolation; instead, characters are usually embedded in multiple-character compound words [(Myers, 2022; Yip, 2000)](https://www.zotero.org/google-docs/?xyxL9t). These multiple-character compounds often convey specific meanings by combining the meanings of constituent characters [(Myers, 2022)](https://www.zotero.org/google-docs/?XAloV6). For example, the compound word for ‘vase’ (花瓶 /huā píng/) contains characters meaning ‘flower’ (花 /huā/) and ‘bottle’ (瓶 /píng/), while the compound word for ‘anaesthesia’ (麻藥 /má yào/) contains characters meaning ‘paralyse’ (麻 /má/) and ‘medicine’ (藥 /yào/).

The frequent use of compounding in Chinese word formation might lead one to suspect that compound words are analysed in terms of their constituents during visual word recognition. However, the situation is complicated because the relationship between individual Chinese characters and the meanings of whole compounds is not always systematic. Chinese characters frequently convey more than one meaning [(Liu et al., 2007; Tsang & Chen, 2010)](https://www.zotero.org/google-docs/?X4tb1U); for example, the character 花 /huā/ is associated with the meanings for both ‘flower’ and ‘spend’ (as in 花錢 /huā qián/, or ‘spend money’), while the character 麻 /má/ is associated with the meanings for both ‘paralyse’ and ‘hemp’ (as in 麻繩 /má shéng/, or ‘hemp rope’). There are also some cases in which the meaning of the whole compound bears no relation to any meaning of the constituent characters; for example, the compound word for ‘nerves’ (神經 /shén jīng/) is built from characters frequently associated with ‘god’ (神 /shén/) and ‘pass’ (經 /jīng/). In other words, the contribution of a character to the meaning of a whole compound is highly dependent on the compound.

The lack of systematicity between the meanings of characters and whole words raises questions about the extent to which readers’ learned representations of characters influence the visual recognition of compound words. This situation also arises in English and in other languages – for example, English compounds can be fully transparent (e.g. blackbird), partially transparent (e.g. horseplay), or fully opaque (e.g. honeymoon) – and yet there is evidence that they are recognised in terms of their constituents [(Crepaldi et al., 2013)](https://www.zotero.org/google-docs/?UOtQEl). However, the constituents that make up compounds in English frequently stand on their own and are learned independently, perhaps leading to particularly robust representations. In contrast, most Chinese constituents tend not to occur in isolation [(Myers, 2022; Yip, 2000)](https://www.zotero.org/google-docs/?1tHfaT). This tendency is visualised in Figure 1 which plots the relative frequency of each character in a megastudy of Chinese compound word recognition reported by Tse et al. (2017). By relative frequency, we mean the token frequency of a character occurring in isolation divided by the token frequency of the character in the corpus. This measure has a high positive skew with over 80% of characters having a relative frequency of less than 0.50. The characters with a relative frequency above 0.50 are typically function words, including conjunctions, propositions, and pronouns.

**Figure 1**

*Density plot of the relative frequency of standing alone for 3,720 constituent characters appearing in the megastudy of Tse et al. (2017). The red dashed line represents the median of the relative character frequency of standing alone (Mdn = 0.29).*

**

This characterisation applies not only to exposure in ecological contexts, but also to the learning of character meanings in classroom settings. Children learn to read and write Chinese characters, including their pronunciation, the strokes needed to produce them, their internal structure, and some aspects of their meaning [(X. Wu et al., 1999)](https://www.zotero.org/google-docs/?Vm1gY1). However, instruction on the meanings of individual characters is typically tied to the compound examples chosen to illustrate those characters [(Hsiang et al., 2021; J. Wang & Leland, 2011)](https://www.zotero.org/google-docs/?6f5kOh). For example, the meaning of the character (麻 /má/) may be illustrated through compounds like 麻藥 (/má yào/, ‘anaesthesia’) and 麻木 (/má mù/, ‘numb’) embedded in sentences, prompting the children to infer the meaning ‘paralyse’ (as opposed to other meanings like ‘hemp’, ‘sesame’, or ‘trivial’). It is not always feasible to teach all possible meanings of a character during reading instruction. This instructional format is very different from English in which children typically learn single morpheme ‘stems’ (e.g. rain) before moving onto inflections (e.g. rained), derivations (e.g. rainy), and compounds (e.g. rainstorm). Indeed, [Rastle and Taylor (2018)](https://www.zotero.org/google-docs/?8APHmH) reported that around 80% of tokens encountered in the first year of reading instruction in English are stems in isolation (e.g. jump); the remaining 20% of tokens are mostly simple inflections (e.g. jumps, jumped).

Thus far, we have established that Chinese characters tend not to occur in isolation, and that instruction on the meanings of Chinese characters tends to arise through compound examples that contain those characters. It therefore seems likely that adults’ knowledge of individual Chinese characters is built over many years through encounters with those characters in compound words, following a brief period of initial instruction. These characteristics raise the possibility that readers develop more robust representations for characters that occur frequently and/or behave consistently across compound words than those that do not. It seems likely that characters with more robust representations may be more influential in the processing of Chinese compounds. This study therefore seeks to (a) quantify the consistency with which individual Chinese characters occur and contribute to the meanings of compound words; and (b) determine how variation in these family properties of characters influences the processing of both familiar and novel compounds.

This work informs a set of enduring theoretical issues pertaining to reading and reading acquisition typically studied with respect to English morphology (and described in the next section). However, investigating these issues in respect of Chinese reading enriches understanding not only because of the unique properties described previously, but also because the visual symbols of Chinese writing reflect *meaning* as opposed to reflecting *sound* in the alphabetic writing systems typically studied.[[1]](#footnote-1) Thus, meaning-based regularities should be particularly important in printed word representations in Chinese, a morphologically rich language [(Plaut & Gonnerman, 2000)](https://www.zotero.org/google-docs/?2tdaEE). In addition to investigating important general issues around acquisition and processing, this study also begins to redress the overreliance on English and alphabetic writing systems in reading research [(e.g. Blasi et al., 2022; Share, 2008)](https://www.zotero.org/google-docs/?OuOKQp).

**Statistical Knowledge of Form-Meaning Relationships in English**

Despite the fact that alphabetic writing systems transcribe the *sounds* of language (rather than the *meanings*), most research on morphological processing in reading has come from studies of English and other alphabetic writing systems (see e.g. Rastle, 2019a; Stevens & Plaut, 2022, for recent reviews). In alphabetic writing systems, the mapping between form and meaning is largely arbitrary for words with only one morpheme; similarity in form does not imply similarity in meaning (e.g. dog, dig, fog). However, morphemes have been described as “islands of regularity” in this mapping [(Rastle et al., 2000](https://www.zotero.org/google-docs/?ZDmB2x), p. 527; also [Plaut & Gonnerman, 2000)](https://www.zotero.org/google-docs/?LPmbm1), as stems occur and reoccur in words with similar meanings (cleanly, cleaner, cleanliness), and affixes transform the meanings of stems in a relatively systematic manner (teacher, builder, gardener; [Rastle et al., 2000](https://www.zotero.org/google-docs/?9jF0S5)). Most words in English (as in other languages) are built by combining morphemes, including stems with prefixes (e.g. rewash), suffixes (e.g. washable), and other stems (e.g. whitewash). However, the reliability with which English morphemes communicate meaning is imperfect; instead, the form-to-meaning relationship in English morphology is *graded* with some morphemes conferring a more consistent meaning than others [(Plaut & Gonnerman, 2000; Ulicheva et al., 2020, 2021)](https://www.zotero.org/google-docs/?mQPs4d).

One reason for the graded nature of English morphemes is that there are morphologically-structured words that are simply unrelated to their constituents; for example, a ‘witness’ is not the quality of being ‘wit’, and a ‘corner’ is not someone who ‘corns’ [(Rastle et al., 2004)](https://www.zotero.org/google-docs/?ErJN5O). In other cases, the meaning of the whole word is only marginally related to its constituents; for example, a ‘fruitful meeting’ is full of fruit only in a metaphorical sense. There are still other cases in which the meaning of the whole word is transparently related to the meanings of its constituents, but this relationship is slightly different in each exemplar [(Schmidtke et al., 2018)](https://www.zotero.org/google-docs/?BtFZOt). To illustrate, the words ‘milkman’, ‘snowman’ and ‘fireman’ are all men (of sorts), but while a ‘snowman’ is a man made of snow, a ‘milkman’ is not a man made of milk; and while a ‘milkman’ delivers the milk, a ‘fireman’ does the opposite of delivering fire. These types of cases all weaken the consistency with which morphemes communicate meaningful information in English writing; and in doing so, they may weaken readers’ ability to acquire understanding of the functions of those morphemes through experience with whole words.

The observation that morphemes communicate meaningful information in a graded manner has motivated research seeking to quantify different forms of morphemic regularity and to test the influence of these measures on printed word processing. It has long been known that stems belonging to large morphological families (i.e. stems that are part of many derivations and compounds) are recognised more quickly in isolation ([De Jong et al., 2000; Schreuder & Baayen, 1997](https://www.zotero.org/google-docs/?aDvPno)) and when they appear in derived words [(Bertram et al., 2000; De Jong et al., 2000)](https://www.zotero.org/google-docs/?7SffI9) and compounds [(De Jong et al., 2002; Kuperman et al., 2009)](https://www.zotero.org/google-docs/?kXD1gn) than stems belonging to small morphological families. This family size effect appears to be at least partly semantically-driven [(Amenta et al., 2020; Marelli et al., 2020; Stevens & Plaut, 2022)](https://www.zotero.org/google-docs/?yUVkZC); for example, the negative correlation between family size and lexical decision latency gets stronger if semantically opaque family members are removed from the family size counts for monomorphemic words [(Schreuder & Baayen, 1997)](https://www.zotero.org/google-docs/?d4WcN7) and derived words [(Bertram et al., 2000)](https://www.zotero.org/google-docs/?BZclvo). This result is interesting because it suggests that morpheme representations arise not because morphemes are orthographic units that occur frequently but because they are orthographic units that occur frequently with *consistent meaningful functions* (see e.g. Tamminen et al., 2015). However, this approach to quantifying the semantic consistency of morphemes relies on a binary classification of potential family members as ‘transparent’ or ‘opaque’; this classification is problematic given the graded nature of morphemes described previously.

One approach to addressing this problem has been to develop a more nuanced, continuous measure termed orthography-semantic consistency (OSC) that quantifies the semantic relationship between a stem and all words in which it is contained [(Marelli et al., 2015; Marelli & Amenta, 2018; see also Ulicheva et al., 2020, 2021 for a similar approach involving affixes)](https://www.zotero.org/google-docs/?POHrht). For instance, the word ‘widow’ has high OSC because it is semantically related to the words in which it is contained (e.g. widowed, widower); in contrast, the word ‘whisk’ has low OSC because it is semantically unrelated to the words in which it is contained (e.g. whisker, whiskey). Several studies now show that OSC impacts word recognition: words with higher OSC are recognised more quickly than words with lower OSC [(Hendrix & Sun, 2021; Marelli et al., 2015; Marelli & Amenta, 2018; Siegelman et al., 2022)](https://www.zotero.org/google-docs/?BFL3RL). These results suggest that morphemes with higher semantic consistency across all the compound words in which they occur yield stronger representations, and hence, are recognised more quickly. These conclusions resonate with our predictions regarding Chinese morphology; namely, that characters that behave more consistently in the compound words in which they occur may yield more salient representations that have greater influence on processing.

OSC is not strictly a measure of the morphological family, however. Marelli et al.’s (2015) definition includes words in the calculation that are not part of the morphological family (e.g. ‘lucrative’ included as part of the ‘rat’ family) and excludes words that are part of the morphological family (e.g. ‘ran’ not included in the ‘run’ family). Likewise, Siegelman et al.’s (2022) definition quantifies the semantic relationship between a stem and all words that can be changed to the stem with a single edit, including insertions, deletions, or substitutions (e.g. ‘ran’ included as part of the ‘run’ family). However, this approach excludes genuine family members with affixes longer than one letter (e.g. ‘irresponsibility’ not included in the ‘responsible’ family). These issues arise because morphemes are not physically demarcated in the English writing system. Thus, although OSC is typically interpreted as a measure of morpheme consistency (Stevens & Plaut, 2022), this interpretation requires caution.

The body of research on morpheme family size effects has begun to suggest that readers may be sensitive to the behaviour of a morpheme *across the whole lexicon*, both with respect to its frequency and its semantic consistency. Early parallel distributed processing (PDP) models sought to understand how morphological representations might emerge from exposure to stems in isolation and in inflections, derivations, and compounds [(Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999](https://www.zotero.org/google-docs/?UGr1px); see [Stevens & Plaut, 2022 for review)](https://www.zotero.org/google-docs/?trde3p). These simulations demonstrated that it is possible to capture graded morphological regularities through experience with whole words and that these emerging regularities are particularly salient in morphologically rich languages [(Plaut & Gonnerman, 2000)](https://www.zotero.org/google-docs/?2dECIl). However, the success of these PDP simulations derives in part from the use of input codes that make morpheme boundaries salient; for example, the use of input codes that are already morphemically segmented [(see Rastle & Davis, 2008; but see Milin et al., 2017)](https://www.zotero.org/google-docs/?nDMf0L). The use of morphemically-structured input codes allows these models to “see” meaningful orthographic similarity across whole words; however, it is unclear where a morphemically-structured input code would come from unless the reader had already acquired morphemic knowledge [(Rastle & Davis, 2008)](https://www.zotero.org/google-docs/?yb8ZHn). The difficulty for PDP models in identifying the meaningful, morphemic elements of a word without “help” from the modeller echoes the limitations of OSC described previously. These limitations arise because morphemes are not physically marked in the English writing system.

Recent distributional semantic approaches have also sought to determine how the functions of morphemes and their combinatorial properties emerge through exposure to whole words ([Amenta et al., 2020)](https://www.zotero.org/google-docs/?L0RIj0). In the trained FRACCS model ([Functional Representations of Affixes in Compositional Semantic Space; Marelli & Baroni, 2015)](https://www.zotero.org/google-docs/?gnv2Zs), a representation of the meaning of an affix is derived through analysis of vectors for all whole words containing that affix (e.g. -LESS; homeless, effortless). The quality of the assimilated meanings is then evaluated through studies of the meanings of novel derivations; for example, [Marelli and Baroni (2015)](https://www.zotero.org/google-docs/?w9xHiW) reported that the novel derived word ‘pastureless’ in the FRACCS model is closer in meaning to ‘barren’ than ‘pasture’, a result that matches human judgements. Similarly, the CAOSS model (Compounding as Abstract Operation in Semantic Space; [Marelli et al., 2017)](https://www.zotero.org/google-docs/?8zS1H7) is trained on vectors of modifier-head compounds in English (e.g. teabag, armchair) and generates modifier-specific and head-specific weight parameters. These parameters are then used to transform free words into modifier or head constituent vectors that can be combined in compound words (including novel compound words). The validity of these representations is evaluated using human lexical decision tasks. For example, [Günther and Marelli (2020)](https://www.zotero.org/google-docs/?oF765c) reported that a novel compound word elicits longer rejection times when, according to CAOSS, its constituents are easier to combine in a new compound meaning. That is, a novel compound such as ‘bridgemill’, whose CAOSS-computed meaning is very close to ‘bridge’ and ‘mill’, is more difficult to reject as a non-word in lexical decision as compared to a novel compound like ‘radiosauce’, whose CAOSS-computed meaning ends up relatively far from ‘radio’ and ‘sauce’. The successful implementation of these computational models suggests that the morpheme behaviour can be assimilated through the training on whole words.

**Meaningful Information in Chinese Characters**

The emerging picture from English and other alphabetic writing systems is that (a) morphemes contribute to the meanings of whole words in a graded manner; and (b) readers’ sensitivity to this form of consistency is reflected in language processing tasks. The present study asks similar questions regarding the behaviour of Chinese characters in compound words and readers’ sensitivity to this behaviour in recognition tasks. Chinese is a particularly interesting writing system to study in respect of these issues. We have already documented how the meaning of Chinese characters depends to a large extent on the compound words embedding that character. Thus, readers’ knowledge of individual characters must be strongly influenced by encounters with characters in compound words rather than being anchored to the knowledge of the meaning of that character in isolation. One consequence of this property is that morphemic information may be tightly bound to the meanings of whole compounds, and hence difficult to extract as general morphemic knowledge. On the other hand, the acquisition of morphemic information from compound words should be supported by the fact that the symbols of Chinese writing usually represent morphemes (i.e. each character is a morpheme) and because compounding is the major mechanism for word formation in Chinese [(Ceccagno & Basciano, 2007; Myers, 2006; Tsang & Chen, 2013)](https://www.zotero.org/google-docs/?WcyelI). In fact, the presence of physical boundaries between Chinese morphemes obviates the questions raised earlier about how readers (or models) know which orthographic chunks are likely to be meaningful [(see Rastle & Davis, 2008)](https://www.zotero.org/google-docs/?HPXNJ4). In the case of Chinese characters, the meaningful units of the writing system align with the physically-demarcated visual symbols of the writing system.

Research has already shown that Chinese words are recognised more quickly if they contain an initial character from a large family [(Li et al., 2015, 2017; Liu et al., 2007; Wu et al., 2013; Xiong et al., 2021)](https://www.zotero.org/google-docs/?F5gW03), with the *family* of a Chinese character defined as the number of words that contain that particular character. This result provides evidence that the recognition of Chinese compounds involves the analysis of individual characters. It makes sense that Chinese characters from larger families should be recognised more easily given that these characters should be learned more robustly (as a result of learners’ experiences with them in many contexts; [Tamminen et al., 2015](https://www.zotero.org/google-docs/?3kTv2i)). However, the relevant Chinese studies to date consist of small-scale factorial manipulations with small numbers of items, and thus it is unclear whether the family size effect extends to non-initial characters. More importantly, we are unaware of any research that has probed how the semantic consistency of Chinese characters modulates the family size effect.

The metric that we establish for quantifying the behaviour of Chinese characters across compound words is *family semantic consistency*. This metric is defined in a similar manner to OSC [(Hendrix & Sun, 2021; Marelli et al., 2015; Marelli & Amenta, 2018; Siegelman et al., 2022)](https://www.zotero.org/google-docs/?Y7AKRZ): the average semantic relatedness between a character and its family members. However, family semantic consistency in the context of Chinese writing departs from the OSC metric in two important ways. The first is that the obtained consistency values in Chinese necessarily reflect morphological families. In contrast to the English metric, the pool of orthographic relatives that contain a character are morphological relatives. The second is that given the marked morpheme boundaries in Chinese, we are able to examine how the consistency of each component in a compound word impacts the recognition of that compound word without having to use researcher knowledge to determine which are the relevant orthographic segments. By contrast, computing a morpheme-based version of OSC in English would require researcher knowledge to determine how to parse a morphologically-complex form into relevant orthographic segments.

In this study, we tested the extent to which family size and family semantic consistency metrics influence the visual recognition of existing and novel Chinese compounds, using response time and accuracy measures from the megastudy by [Tse et al. (2017)](https://www.zotero.org/google-docs/?mhzZCy). Our prediction was that characters with a higher family size and a more semantically-consistent family would yield stronger learned representations, and hence should facilitate the recognition of existing compounds containing those characters, and should make it difficult to reject novel compounds containing those characters. In addition, we hypothesised an interaction between family size and family semantic consistency, such that the effect of family size becomes weaker when the family lacks semantic consistency.

**Method**

We calculated family size and family semantic consistency metrics using a corpus of traditional Chinese script that we created along with a publicly-available distributional semantic model for traditional Chinese script. We then assessed the extent to which these metrics influenced lexical decision responses in a megastudy of Chinese word recognition [(Tse et al., 2017)](https://www.zotero.org/google-docs/?yAqcpJ).

**Corpus of traditional Chinese script**

We established a new corpus by concatenating three corpora of traditional Chinese script based on Taiwanese Mandarin: Ministry of Education, Taiwan (2000), Academia Sinica, Taiwan (2001), and [Wu and Liu (1988)](https://www.zotero.org/google-docs/?4rw0Ra). This merged corpus has approximately 11.3 million word tokens in total. To verify the validity of this merged corpus, we examined the predictive power of the (log-transformed) word frequency measure obtained from this corpus using the megastudy of Tse et al. (2017) (see *Dataset* below for details) using linear regression. Words with zero frequency (i.e. not observed in the corpus) were dealt with using Laplace smoothing in which 1 is added to each word frequency value so that all words have a frequency value of at least one [(Brysbaert & Diependaele, 2013)](https://www.zotero.org/google-docs/?j2Yklx). R2 of the regression model was .236, which was slightly larger than the R2 (.202) of the whole word frequency effect based only on Academia Sinica (2001).

**Word vectors from word2vec models**

Vectors representing the meanings of Chinese characters and compound words were obtained using the pre-trained word2vec model [(Mikolov et al., 2013)](https://www.zotero.org/google-docs/?tkHqBi) developed by Academia Sinica, Taiwan (available at<https://ckip.iis.sinica.edu.tw/project/embedding>). This model was trained using the skip-grams with negative sampling (SGNS) method, which predicts the context words given a target word. 300-dimensional vectors for 500,000 word types can be obtained in this model.

We selected the Academic Sinica model following an evaluation of four publicly available word2vec models for traditional Chinese script, including Facebook, University of Oslo, Academia Sinica, and ToastyNews (see Table 1 for details). Our aim was to determine which word2vec model provides semantic representations that best approximate human semantic relatedness judgements. The analysis employed two publicly-available datasets, WordSim-240 and WordSim-296 (the number embedded in the names denotes the number of data points (i.e. word pairs), available at<https://github.com/ray1007/GWE>). Each dataset consists of a list of word pairs along with corresponding human ratings of semantic relatedness. WordSim-240 and WordSim-296 were translated into traditional Chinese script from simplified Chinese script and proved to align with traditional Chinese neural network models by [Su and Lee (2017)](https://www.zotero.org/google-docs/?1tEnHH). Because not all word pairs in these datasets are represented in the models being evaluated, we selected the subset of pairs from each of these datasets that were represented in all the four models. This procedure resulted in a drop rate of 7.5% for WordSim-240 (18 word pairs) and 12.8% for WordSim-296 (38 word pairs).

**Table 1**

*Information of all the publicly available word2vec models for traditional Chinese*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Developer | Algorithm | Number of dimensions | Number of word types (Million) | Sources of training data |
| Facebook | fastText CBOW | 300 | 17 | Common Crawl (2017) & Wikipedia (2017) |
| University of Oslo | word2vec skip-gram | 100 | 2 | Chinese CoNLL17 corpus (2017) |
| Academia Sinica | word2vec skip-gram | 300 | 0.5 | Academia Sinica Balanced Corpus & Chinese Gigaword Corpus |
| ToastyNews | fastText CBOW | 300 | 0.9 | ToastyNews crawled Hong Kong news websites |

*Note.* Unlike the operation in the English language, word2vec skip-gram and fastText algorithms train Chinese models in an identical fashion both using character-level, or subword-level information. CBOW, continuous bag of words.

Table 2 presents Spearman’s rank correlations that were computed between the human ratings of semantic relatedness and the cosine similarities extracted from the four word2vec models. The higher the value, the better the performance. Analyses on WordSim-240 (*N* = 222) and WordSim-296 (*N* = 258) both suggested that the word2vec model developed by Academia Sinica had the best overall performance.

**Table 2**

*Spearman’s rank correlations between human-labelled semantic relatedness and cosine similarities of word pairs on two datasets: WordSim-240 and WordSim-296*

|  |  |  |
| --- | --- | --- |
|  | WordSim-240(*N* = 222) | WordSim-296(*N* = 258) |
| Facebook | .560 | .697 |
| University of Oslo | .557 | .575 |
| Academia Sinica | .627 | .707 |
| ToastyNews | .564 | .668 |

*Note*. *N*, number of word pairs used to compute each Spearman’s rank correlation.

**Basic definition of family size and family semantic consistency**

 We define a morphological family as a group of two-character compound words sharing a constituent character. Family members are hence two-character compound words that embed a given constituent. For example, words containing the character 花 /huā/ in Table 3 belong with the family of 花 /huā/ while words containing the character 瓶 /píng/ belong with the family of 瓶 /píng/. The family size metric was obtained by counting the number of family members in the corpus that we developed.

**Table 3**

*Examples of Chinese compound words containing 花 /huā/ and 瓶 /píng/*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Literal meaning by constituents | Actual meaning | Whole word frequency | Cosine similarity with 花 /huā/ | Cosine similarity with 瓶 /píng/ |
| Words containing 花 /huā/ |   |   |   |
| 花瓶 /huā píng/ | *flower-bottle* | *vase* | 38 | .276 | .341 |
| 花束 /huā shù/ | *flower-bunch* | *bouquet* | 4 | .378 | - |
| 花費 /huā fèi/ | *spend-spend* | *to spend* | 432 | .640 | - |
| 花生 /huā sheng/ | *flower-grow* | *peanut* | 103 | .256 | - |
| 開花 /kāi huā/ | *open-flower* | *to blossom* | 99 | .424 | - |
| 棉花 /mián huā/ | *cotton-flower* | *cotton* | 44 | .192 | - |
| Words containing 瓶 /píng/ |   |   |   |
| 瓶蓋 /píng gài/ | *bottle-lid* | *bottle lid* | 34 | - | .647 |
| 瓶頸 /píng jǐng/ | *bottle-neck* | *bottleneck* | 254 | - | .134 |
| 花瓶 /huā píng/ | *flower-bottle* | *vase* | 38 | .276 | .341 |

*Note*. Whole word frequency was obtained from the corpus we created. Cosine similarity was calculated with word vectors from the word2vec model by Academia Sinica, Taiwan.

Family semantic consistency refers to the average semantic relatedness of family members with the shared constituent character. Semantic relatedness was calculated as pairwise cosine similarity between word vectors derived from a word2vec model for traditional Chinese script [(Marelli & Baroni, 2015; H.-C. Wang et al., 2014)](https://www.zotero.org/google-docs/?Ey2Tm0). In this study, cosine similarity values range from 0 to 1, with higher cosine similarity values indicating greater semantic relatedness. Thus, the higher the cosine similarity, the greater the semantic consistency of the family.

**Comparing different metrics for family size and family semantic consistency**

With the corpus and the word2vec model, we sought to obtain four family-related metrics for each compound word (*N* = 25,281) and nonword (*N* = 25,085) in the megastudy of Chinese two-character word recognition (Tse et al., 2017): family size of character 1 (hereafter, C1 family size), family size of character 2 (hereafter, C2 family size), family semantic consistency of character 1 (hereafter, C1 family semantic consistency) and family semantic consistency of character 2 (hereafter, C2 family semantic consistency). However, in determining the pool of family members upon which the consistency measure is computed [(Marelli & Amenta, 2018; Siegelman et al., 2022)](https://www.zotero.org/google-docs/?Mcfmz8), two theoretically important questions arise regarding the nature of the family.

The first question is whether family should be defined in terms of type or token frequency (see [Siegelman et al., 2022](https://www.zotero.org/google-docs/?ISLHs5)). Type-based family size is the number of words in which a particular character occurs, while token-based family size is the summed frequency of the family members. This distinction is important because it relates to how readers acquire knowledge of characters: is it more important that the character occurs in a number of different words or that the character occurs frequently (even if only in a single word; see e.g. [Tamminen et al., 2015](https://www.zotero.org/google-docs/?Bp9aTK))?

The second question is whether the family members of a particular character should be constrained by the position of the character. That is, should family members of a particular character contain that character in the same position across all words, or should the family include words that contain the character in a different position? The question of whether position information influences the calculation of family size is theoretically important because it addresses how readers’ character knowledge is formed. If the family of a particular character is constrained by position, then that means that an individual’s knowledge of a character that occurs in the word-initial position is not influenced by encounters with that character in a different position. Conversely, if the family of a particular character is not constrained by position, then that means that any encounter with a particular character contributes to knowledge of that character, irrespective of where it occurs in a word.

Taking into account these two parameters (type-based versus token-based; position-specific versus position-general), we obtain four different measures of all family-related variables (see Table 4). Suppose that Table 3 exhaustively lists all words containing 花 /huā/ and 瓶 /píng/. Table 4 shows that for the target word 花瓶 /huā píng/, C1 family size without positional constraint is six by type and 720 by token, since there are six words sharing the character 花 /huā/. However, C1 family size with positional constraint is four by type and 577 by token, since there are only four words that start with the character 花 /huā/. Likewise, C2 family size without positional constraint is three by type and 326 by token, since three words contain the character 瓶 /píng/, while C2 family size with positional constraint is 1 by type and 38 by token, since only one word ends with 瓶 /píng/.

**Table 4**

*Family size and family semantic consistency metrics under four definitions for the target word 花瓶 (/huā píng/, flower-bottle, meaning vase)*

|  |  |  |  |
| --- | --- | --- | --- |
| Family-related metrics | Position-specific |   | Position-general |
| Type-based | Token-based |   | Type-based | Token-based |
| C1 FS | 4 | 577 |   | 6 | 720 |
| C2 FS | 1 | 38 |   | 3 | 326 |
| C1 consistency | 0.388 | 0.546 |   | 0.361 | 0.507 |
| C2 consistency | 0.341 | 0.341 |   | 0.374 | 0.217 |

*Note.* The values calculated here are all based on Table 3, with the assumption that Table 3 exhaustively lists all words containing 花 /huā/ and 瓶 /píng/. FS, family size; consistency, family semantic consistency.

Family semantic consistency by type is the average cosine similarities of family members, while family semantic consistency by token is the frequency-weighted average cosine similarities of family members (i.e. cosine similarities weighted by frequency divided by the summed frequency of family members). In Table 4, for 花瓶 /huā píng/, C1 family semantic consistency without positional constraint is .361 by type and .507 by token, while C1 family semantic consistency with positional constraint is .388 by type and .546 by token. Likewise, C2 family semantic consistency without positional constraint is .374 by type and .217 by token, while C2 family semantic consistency with positional constraint is .341 by type and .341 by token. Family semantic consistency measures under the four definitions can be found in the Material folder at<https://osf.io/thfq6/>.

**Lexical Decision Dataset**

Our analyses were conducted on lexical decision data from a megastudy of Chinese word recognition (Tse et al., 2017). This dataset comprises response time and accuracy measures for over 25,000 two-character compound words and nonwords (i.e. words composed of two real characters but not included as dictionary entries) printed using traditional Chinese characters. The dataset includes responses to each of these items from about 33 Cantonese speakers in Hong Kong.

Prior to the statistical analyses, items whose accuracy across participants was less than 70% were removed as outliers (Tse et al., 2017; Tse & Yap, 2018). Implementation of this criterion left 22,808 words (drop rate = 9.8%) and 23,807 nonwords (drop rate = 5.1%). Further, because family semantic consistency values of some characters could not be obtained (they were too low in frequency to be represented in the word2vec model used), latencies for 22,701 words (drop rate = 0.4%) and 23,047 nonwords (drop rate = 3.2%) were eventually entered in the analysis (for raw data, please refer to the Dataset folder at<https://osf.io/thfq6/>). This study was not preregistered.

**Results**

 To examine the effects of family size and family semantic consistency, we first investigated which of the four different approaches for defining the morphological family explained the most variance in word recognition response times. Our subsequent analyses then use the highest-performing metric.

**Metric comparison**

To determine which metrics for family size and family semantic consistency best predict the lexical decision data, we built eight linear mixed-effect models for four definitions (type vs token; position-specific vs position-general) based on inverse transformed lexical decision latencies [(Box & Cox, 1964)](https://www.zotero.org/google-docs/?6DlKV6) of two types of items (word, nonword) using the lme4 package in R (Bates et al., 2015) with maximum likelihood estimation applied (Faraway, 2016). In each model, log-transformed C1 family size, C1 family semantic consistency, log-transformed C2 family size and C2 semantic consistency under one of the four definitions were entered as predictors of interest in the model and centred to the mean to protect the estimates of these main effect terms from the interference of multicollinearity (Iacobucci et al., 2016, 2017). Interactions between family size and family semantic consistency (for both C1 and C2) were also entered into each model. Log-transformed whole word frequency was entered additionally as a controlling predictor for models associated with word data, and by-C1 and by-C2 random intercepts were included. We then compared the performance of all the models using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Smaller values indicate a better fit of the model to the data. We also computed Akaike weights, representing the relative probability of a model to provide the best fit (in terms of Kullback-Leibler discrepancy) given the data and the set of hypothesised models, in order to examine if the difference in AIC values was non-negligible [(Wagenmakers & Farrell, 2004)](https://www.zotero.org/google-docs/?LWBbJL). The R script (model\_comparison.R) is available at

<https://osf.io/thfq6/>.

Table 5 displays AIC, BIC and Akaike weights for all the considered models. Based on the word data, both AIC and BIC indicated that the position-specific, type-based metric produced the best performance. The ratio of Akaike weights suggests that this model was 4.7 x 1011 times more likely to be the best model than the next-best model (position-general, type-based) [(Wagenmakers & Farrell, 2004)](https://www.zotero.org/google-docs/?YPb9rd). The AIC and BIC values associated with the nonword data also favour the position-specific, type-based model, with the ratio of Akaike weights suggesting that this model was 1.1 x 1017 more likely to be the best model than the next-best model (position-specific, token-based) [(Wagenmakers & Farrell, 2004)](https://www.zotero.org/google-docs/?Yuj3Ng). Subsequent analyses thus focus only on the position-specific, type-based model.

**Table 5**

*AIC, BIC and Akaike weights for model comparison on inverse transformed lexical decision latencies of words and nonwords from Chinese Lexicon Project (Tse et al., 2017)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Baseline | Position-specific |  | Position-general |
| Type-based | Token-based | Type-based | Token-based |
| *Word data* |   |   |  |   |   |
| AIC | 28875 | -29154 | -29038 |   | -29099 | -28987 |
| BIC | -28835 | -29065 | -28950 |   | -29011 | -28899 |
| Akaike weights | - | 1 | 9.2x10-26 |   | 1.5x10-12 | 7.1x10-37 |
| *Nonword data* |   |   |  |   |   |
| AIC | -41060 | -41548 | -41479 |   | -41313 | -41304 |
| BIC | -41028 | -41468 | -41398 |   | -41233 | -41224 |
| Akaike weights | - | 1 | 8.1x10-16 |   | 7.7x10-52 | 8.9x10-54 |

*Note*. For word data, the baseline model included whole word frequency and random intercepts, while for nonword data, the baseline model only included random intercepts. AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.

**Statistical analysis on lexical decision latencies of words and nonwords**

We refitted models of the position-specific, type-based metric for word (Model 1) and nonword data (Model 2), with restricted maximum likelihood estimation applied for accurate coefficient estimation (Baayen et al., 2008). Interaction terms were removed where they did not contribute significantly to the final model. While *p* values were obtained using the Satterthwaite approximation in the lmerTest package in R [(Kuznetsova et al., 2017)](https://www.zotero.org/google-docs/?9flfTv), semi-partial *R2* for the fixed effects was derived using the r2glmm package in R [(Jaeger, 2017)](https://www.zotero.org/google-docs/?TMI6y9), where the approach of [Nakagawa and Schielzeth (2013)](https://www.zotero.org/google-docs/?3EGt4a) was applied. The R script (lmer\_RT.R) is available at<https://osf.io/thfq6/>. Table 6 displays the results of the final linear mixed-effects models for word (Model 1) and nonword (Model 2) response times.

**Table 6**

*Effects of family size and family semantic consistency on lexical decision latencies of words (Model 1) and nonwords (Model 2) from Chinese Lexicon Project (Tse et al., 2017)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | DV | Predictor | Estimate | *SE* | *t* | *p* | Semi-partial *R2* |
| 1 | Inv-RT ofreal words | Intercept | 1.408 | 0.002 | 573.942 | <.0001 | - |
|   | Log-word freq | 0.040 | 0.000 | 80.755 | <.0001 | 23.0% |
|   | Log-C1 FS | 0.016 | 0.001 | 12.279 | <.0001 | 1.1% |
|   | C1 consistency | 0.086 | 0.013 | 6.694 | <.0001 | 0.3% |
|   | Log-C2 FS | 0.014 | 0.001 | 10.496 | <.0001 | 1.0% |
|   | C2 consistency | 0.093 | 0.019 | 4.930 | <.0001 | 0.3% |
|   |   | Log-C2 FS x C2 consistency | 0.027 | 0.009 | 3.009 | .0026 | 0.1% |
| 2 | Inv-RT ofnonwords | Intercept | 1.380 | 0.002 | 780.525 | <.0001 | - |
|   | Log-C1 FS | -0.018 | 0.001 | -15.814 | <.0001 | 2.6% |
|   | C1 consistency | -0.039 | 0.011 | -3.528 | .0004 | 0.1% |
|   | Log-C2 FS | -0.015 | 0.001 | -16.280 | <.0001 | 2.6% |
|   | C2 consistency | -0.027 | 0.010 | -2.698 | .0070 | 0.1% |

*Note.* Inv-RT, inverse transformed reaction times; DV, dependent variable; C1, first character, C2, second character; log, log-transformed; word freq, whole word frequency; FS, family size; consistency, family semantic consistency.

 The analysis of response times for words revealed that all the main effects were significant (*p*s < .001): C1 family size, C2 family size, C1 family semantic consistency and C2 family semantic consistency. The direction of these effects is such that higher family size and higher family semantic consistency speed response time. The inclusion of the C2 family size by C2 family semantic consistency interaction improved the model, *χ*2(1) = 9.05, *p* = .003, but the inclusion of the C1 family size by C1 family semantic consistency interaction did not, *χ*2(1) = 2.75, *p* = .097.

To better understand the significant interaction term in the model, we visualised it using the sjPlot package in R [(Lüdecke, 2023)](https://www.zotero.org/google-docs/?S2Q2gB). Figure 2 presents three values of family semantic consistency, namely, the mean value, the value one *SD* above the mean, and the value one *SD* below the mean [(Aiken et al., 1991)](https://www.zotero.org/google-docs/?TxZtxZ). Family semantic consistency was treated as a continuous variable in the regression analysis; it is represented as separate categories in Figure 2 to help visualise the interaction. This visualisation reveals that the effect of family size is particularly large when there is high family semantic consistency.

**Figure 2**

*An interaction plot for log-transformed C2 family size and C2 family semantic consistency in Model 1 (on latency data of words)*



*Note.* Inv-RT, inverse transformed reaction times; C2, second character; consistency, family semantic consistency.

The analysis of nonword response times revealed that the significant main effects include (*p*s < .01): C1 family size, C2 family size, C1 family semantic consistency and C2 family semantic consistency. Higher family size and higher family semantic consistency slows reaction times of rejecting nonwords. Likelihood ratio tests revealed that neither of the interactions improved the model fit: C1 family size by C1 family semantic consistency (*χ*2(1) = 0.44, *p* = .509); C2 family size by C2 family semantic consistency (*χ*2(1) = 0.03, *p* = .874).

**Statistical analysis on lexical decision accuracy of words and nonwords**

Generalised linear mixed-effect models were fitted on the number of correct and incorrect responses for each word (Model 3) and each nonword (Model 4) using the glmmTMB package in R [(Brooks et al., 2017)](https://www.zotero.org/google-docs/?9RuqG1), with restricted maximum likelihood applied for coefficient estimation. Given overdispersion present in the data, a logit link function with a beta-binomial family was employed for the model [(Agresti, 2012; Harrison, 2015)](https://www.zotero.org/google-docs/?27TiUD). The main effects of log-transformed C1 family size, C1 family semantic consistency, log-transformed C2 family size and C2 semantic consistency were entered as predictors of interest in the model and centred to the mean to protect the estimates of these main effect terms from the interference of multicollinearity [(Iacobucci et al., 2016, 2017)](https://www.zotero.org/google-docs/?q75g0w). We also tested for an interaction between family size and family semantic consistency (for both C1 and C2) using likelihood ratio tests; interaction terms were removed where they did not contribute significantly to the model. Family size and family semantic consistency values were based on the position-specific, type-based metric, as indicated by the results of the metric comparison analysis. Log-transformed whole word frequency was controlled for the analysis on word accuracy, and by-C1 and by-C2 random intercepts were included. Odds ratios, referring to the difference in odds of being correct (vs incorrect) in a single unit change on the predictor, were reported as the effect size. Odds ratios greater than one indicate increase in odds of being correct while smaller than one indicates decrease in odds of being correct given a single unit increase on the predictor. The R script (glmmTMB\_Acc.R) is available at<https://osf.io/thfq6/>. Table 7 displays the results of the final linear mixed-effects models for word (Model 3) and nonword (Model 4) accuracy.

 **Table 7**

*Effects of family size and family semantic consistency on lexical decision accuracy of words (Model 3) and nonwords (Model 4) from Chinese Lexicon Project (Tse et al., 2017)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | DV | Predictor | Estimate | *SE* | Wald *Z* | *p* | Odds ratio |
| 3 | Accuracy ofreal words | Intercept | 1.800 | 0.015 | 116.697 | <.0001 | 6.049 |
|   | Log-word freq | 0.225 | 0.004 | 57.424 | <.0001 | 1.253 |
|   | Log-C1 FS | 0.059 | 0.008 | 6.908 | <.0001 | 1.060 |
|   | C1 consistency | 0.722 | 0.109 | 6.605 | <.0001 | 2.059 |
|   | Log-C2 FS | 0.019 | 0.008 | 2.382 | .0172 | 1.019 |
|   | C2 consistency | 0.641 | 0.114 | 5.632 | <.0001 | 1.898 |
|   |   | Log-C1 FS x C1 consistency | 0.198 | 0.058 | 3.436 | .0005 | 1.219 |
|   |   | Log-C2 FS x C2 consistency | 0.183 | 0.058 | 3.157 | .0015 | 1.201 |
| 4 | Accuracy ofnonwords | Intercept | 2.415 | 0.012 | 208.187 | <.0001 | 11.193 |
|   | Log-C1 FS | -0.164 | 0.009 | -19.090 | <.0001 | 0.890 |
|   | C1 consistency | 0.150 | 0.114 | 1.308 | .1909 | 1.161 |
|   | Log-C2 FS | -0.178 | 0.007 | -24.215 | <.0001 | 0.837 |
|   | C2 consistency | 0.134 | 0.106 | 1.264 | .2061 | 1.143 |
|   | Log-C1 FS x C1 consistency | 0.153 | 0.062 | 2.457 | .0140 | 1.165 |
|   | Log-C2 FS x C2 consistency | 0.120 | 0.058 | 2.087 | .0369 | 1.128 |

*Note.* DV, dependent variable; C1, first character, C2, second character; log, log-transformed; word freq, whole word frequency; FS, family size; consistency, family semantic consistency.

The analysis of accuracy for words revealed that all the main effects were significant (*p*s < .05): C1 family size, C2 family size, C1 family semantic consistency and C2 family semantic consistency. The direction of these effects is such that higher family size and higher family semantic consistency improve accuracy. The model was improved with the inclusion of the C1 family size by C1 family semantic consistency interaction, *χ*2(1) = 13.50, *p* < .001, and the inclusion of the C2 family size by C2 family semantic consistency interaction, *χ*2(1) = 11.74, *p* < .001. Plots in Figure 3 further reveal that the effect of family size is particularly large when there is high family semantic consistency.

**Figure 3**

*Interaction plots for (a) log-transformed C1 family size and C1 family semantic consistency, and (b) log-transformed C2 family size and C2 family semantic consistency in Model 3 (on accuracy data of words)*

**

*Note.* C1, first character, C2, second character; consistency, family semantic consistency.

The analysis of nonword accuracy revealed that only the main effects of C1 family size and C2 family size were significant (*p*s < .001). Higher family size makes it more difficult to reject nonwords accurately. The model was improved with the inclusion of the C1 family size by C1 family semantic consistency interaction, *χ*2(1) = 5.98, *p* = .015, and the inclusion of the C2 family size by C2 family semantic consistency interaction, *χ*2(1) = 4.38, *p* = .036. This interaction reflects the fact that the effect of family size is smaller when family semantic consistency is larger (see Figure 4). The direction of this interaction is unexpected and it is inconsistent with the non-significant interaction effects for nonword latencies. We also note that the significance values for this interaction are at threshold despite substantial power in our design. These two issues lead us to question the robustness of this interaction effect.

**Figure 4**

*Interaction plots for (a) log-transformed C1 family size and C1 family semantic consistency, and (b) log-transformed C2 family size and C2 family semantic consistency in Model 4 (on accuracy data of nonwords)*

**

*Note.* C1, first character, C2, second character; consistency, family semantic consistency.

**Discussion**

The past 20 years of research on reading and reading acquisition has begun to reveal the importance of understanding properties of the writing system being learned and processed [(e.g. Frost, 2012; Liversedge et al., 2016; Ziegler & Goswami, 2005)](https://www.zotero.org/google-docs/?OWNER3). This body of research has begun to suggest that properties of the skilled processing system reflect important characteristics of the writing system discovered through the long process of reading acquisition [(e.g. Rastle, 2019b)](https://www.zotero.org/google-docs/?WToEFx). In this vein, this study sought to determine how properties of individual characters influence the recognition of Chinese compound words, and to understand why those effects arise.

Using a megastudy of lexical decision data [(Tse et al., 2017)](https://www.zotero.org/google-docs/?Zw8BZb), we examined the extent to which the visual recognition of two-character Chinese words is influenced by the family size and family semantic consistency of their constituent characters. Family size is simply a measure of how frequently a particular character occurs in different words, while family semantic consistency measures the reliability with which that character conveys semantic information across those same words. These measures reflect aspects of an individual’s experiences with characters across the words in which they occur. Both measures can be calculated by types or by tokens, and both can be calculated in a position-specific or position-general manner. These different ways of calculating the family properties of characters reflect important theoretical considerations about how experience shapes knowledge, as we discuss below.

**Family Semantic Consistency Effect**

Our headline result was that the recognition of Chinese compounds is influenced not only by the family size of constituent characters but also by the consistency with which those characters communicate semantic information. Chinese compounds comprising characters that have a high family size and that are more semantically consistent are recognised more quickly than those comprising characters with a lower family size and lower semantic consistency. Moreover, there was evidence of an interaction between these effects, such that the effect of family size decreases when the character does not behave consistently across compound words. This result indicates that frequent experience with a character is not as helpful during processing if that character does not communicate reliable semantic information.

These properties of individual characters influenced not only the recognition of existing Chinese compounds but also the rejection of novel Chinese compounds. Novel Chinese compounds comprising characters with a higher family size and with higher semantic consistency were harder to reject than those with a lower family size and lower semantic consistency. This result is interesting because it suggests that a reader’s decision to reject an unfamiliar compound is influenced by information about the semantic properties of individual characters. Our results do not speak to whether there is any analysis of the extent to which those individual characters are meaningful *in combination* [(see e.g. Günther & Marelli, 2020; Marelli et al., 2017)](https://www.zotero.org/google-docs/?H8FJad).

Our results suggest that readers exhibit sensitivity to the characters embedded in compounds. When learning Chinese characters, readers do not always regard them as standalone units; instead, they tend to learn them in relation to compound words, and further generalise their understanding to other compound words. Character representations may emerge via experiences with compound words, with the robustness of these representations dependent on the behaviour of characters across distinct morphological family members. This hypothesis is in line with the behaviour predicted by the PDP model (Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999) and the CAOSS model (Marelli et al., 2017).

Our findings are also consistent with insights from other writing systems suggesting that morphemes become important units of skilled processing to the extent that they reflect a consistent relationship between orthographic form and meaning [(e.g. Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999; Ulicheva et al., 2020)](https://www.zotero.org/google-docs/?6CT4qb). We believe that the family size effect is semantically driven in Chinese (and also in other languages; [De Jong et al., 2000; Schreuder & Baayen, 1997)](https://www.zotero.org/google-docs/?qtH024) because morphemes are learned more robustly when they communicate reliable semantic information (see also Tamminen et al., 2015). These conclusions are supported by the effects observed in the present study and are inconsistent with recent proposals suggesting that morpheme representations may reflect relatively primitive visual statistical learning mechanisms [(Lelonkiewicz et al., 2020; Vidal et al., 2021)](https://www.zotero.org/google-docs/?o3M7oN). It may be that adults can learn visual statistical regularities devoid of meaning in artificial laboratory settings (as in these recent studies), but our data suggest that visual regularities alone do not drive the acquisition of morpheme knowledge in reading. Morpheme representations are most salient when orthographic forms occur frequently and communicate consistent semantic information.

Despite the importance of morphemic, character-level processing highlighted in this study, it is important to highlight the substantial difference in effect sizes between the whole-word frequency effect (accounting for around 23% of the variance in response times for words) and morpheme-level effects (accounting for around 1% of the variance in response times for words). This result may suggest a preference for holistic processing, or at least that holistic processing is more efficient, especially for high-frequency two-character words (Cui et al., 2021; Shen et al., 2018). However, this result does not imply that morpheme analysis lacks importance in Chinese printed word recognition (see G[ü](https://www.zotero.org/google-docs/?H8FJad)nther & Marelli, 2018, 2020 for related arguments). Unfamiliar words or nonwords lack a holistic lexical representation and must be analysed in terms of characters in order to compute a possible meaning. It is likely that adult readers encounter unfamiliar words on a regular basis (Brysbaert et al., 2016), so it is perhaps not surprising that these readers appear to analyse familiar and unfamiliar compound words by default in terms of their characters. The use of this analytic process also suggests that skilled Chinese readers have acquired an understanding that characters are important meaningful elements within words (Yan et al., 2010). Our results speak in favour of the argument that morpheme-level information is readily available when processing compound words.

**Insights from Model Comparison Analysis**

Our model comparison results provide powerful evidence that family-related metrics explain the most variance in response latency for both existing and novel compounds when they are calculated in a type-based, position-specific manner. This result provides important insight into how experiences with Chinese characters translate to the long-term knowledge used in skilled processing. The fact that type-based family counts were more predictive than token-based family counts suggests that experiencing a Chinese character across different compound words is more important than the raw experience of a character (for example, in a single high-frequency compound). This result is consistent with findings from artificial morpheme learning paradigms (Tamminen et al., 2015), and from previous studies of family size effects in lexical decision [(De Jong et al., 2000; Schreuder & Baayen, 1997)](https://www.zotero.org/google-docs/?ldlTrT). This result also aligns more broadly with a body of research suggesting that words that occur in a wide range of semantic contexts are easier to recognise [(Hoffman & Woollams, 2015; Hsiao et al., 2020)](https://www.zotero.org/google-docs/?JqcCqv) and easier to learn [(Joseph & Nation, 2018; Mak et al., 2021; Norman et al., 2022)](https://www.zotero.org/google-docs/?BlZ1r8) than words that occur in a smaller range of semantic contexts. Taken together, these studies may suggest that it is the diversity of experience (of a morpheme, or of a word) in the context that drives learning as opposed to experience in and of itself.

The other remarkable insight emerging from the model comparison analysis is that readers are sensitive to the position in which character constituents occur. Put another way, it seems that readers learn “positionally bound” constituents based on family members that share that constituent in the same position. This result is surprising because it suggests that readers’ experiences of a Chinese character as the second constituent in a compound may not inform their knowledge of that character when it occurs as the first constituent in a compound, for example. On the face of it, this type of positionally-bound learning would seem inefficient. Although there is evidence that learned representations of English affixes are position-specific [(Crepaldi et al., 2010)](https://www.zotero.org/google-docs/?H6P4qe), English prefixes and suffixes are actually bound to specific positions. In contrast, the vast majority of Chinese characters can occur in all positions [(Liang et al., 2017)](https://www.zotero.org/google-docs/?bErucX).

One possibility is that the meaning of Chinese characters is systematically influenced by the position in which they occur, and it is for this reason that readers learn to prioritise experiences in a position-specific manner. Libben (2014) illustrates this point with the example constituent ‘key’. When ‘key’ acts as the head of a compound (e.g. latchkey, housekey), it refers primarily to a concrete item used for locking and unlocking. By contrast, when ‘key’ acts as the modifier of a compound (e.g. keyword, keynote), it usually bears an abstract, metaphorical meaning. The ‘reversible word’ phenomenon in Chinese offers an even more striking example. These are words with distinct meanings that consist of the same two constituent characters in reversed order. For instance, both comprising 牛 (/niú/, *cattle*) and 乳 (/rǔ/, *milk*), 牛乳 means ‘cow’s milk’ while 乳牛 means ‘dairy cattle’. In these compound words, 牛 acts as the modifier in 牛乳 but the head in 乳牛. These reversible words are not particularly rare in Chinese and illustrate how the functions and meanings of constituents vary with position. The position-specificity of compound constituents is particularly important when we consider the coinage of novel compounds. A new family member is typically created by substituting one constituent while keeping the other constituent fixed in the existing compound, rather than randomly combining two constituents together (Libben, 2010; Singh, 2006). For example, the existence of ‘rock music’ and ‘pop music’ makes the advent of ‘ambient music’ more acceptable.

To examine how position might influence the meanings of Chinese characters, we developed an analysis using a metric called content similarity (see the Discussion folder at<https://osf.io/thfq6/>). Content similarity is defined as the average cosine similarity of every family member with every other family member [(Ulicheva et al., 2021; H.-C. Wang et al., 2014)](https://www.zotero.org/google-docs/?lRHR6C). If the position of Chinese characters adds systematicity to their meanings, then we might anticipate that position-specific content similarity should be larger than position-general content similarity. Using the word dataset (*N* = 21,668) in the megastudy [(Tse et al., 2017)](https://www.zotero.org/google-docs/?H0lWIQ), we found a statistically significant effect in the predicted direction, although with a relatively small effect size. For C1, the position-specific content similarity (*M* = 0.246, *SD* = 0.082) was larger than position-general content similarity (*M* = 0.242, *SD* = 0.073), *t*(20,217) = 19.26, *p* < .0001, Cohen’s *dz* = 0.13, 95% CI [0.12, 0.14]; for C2, the position-specific content similarity (*M* = 0.254, *SD* = 0.080) was larger than the position-general one (*M* = 0.242, *SD* = 0.071), *t*(20,217) = 46.86, *p* < .0001, Cohen’s *dz* = 0.32, 95% CI [0.30, 0.33]. These results suggest that defining family size in a position-specific manner improves the semantic relatedness among positional family members, especially for C2 family members. This analysis provides only a starting point to understanding why position-specific metrics appear superior to position-general ones when it comes to family size, but it does provide an initial indication that the position in which a Chinese character occurs may convey important semantic information captured during reading acquisition.

**Future Directions**

It remains unclear whether our results based on lexical decision data can be generalised to more natural and ecological settings (e.g. sentence reading using eye-tracking) in which Chinese readers have to process text with no explicit word boundaries. Xiong et al. (2023) reported no difference in character-level effects across lexical decision and sentence reading, and on this basis, we predict that facilitatory effects of family metrics may also be observed in sentence reading. However, this may be the case only when target words are unpredictable from the prior context (Yao et al., 2022). If target words are highly predictable, then this may lead to pre-activation of the whole word, which may make recognition too rapid for a morpheme-level analysis.

In addition, although a recent megastudy suggested that the role of phonetic radicals in lexical decision is minimal (Tse et al., 2022), the role of semantic radicals still remains unclear. Our discussion of readers’ sensitivity to orthography-semantics mapping in the current study is restricted to the relationship between *characters* and *compound words*. This relationship should not be confused with another metric, semantic radical consistency: the average semantic relatedness between a *semantic radical* (e.g. 艹 meaning ‘grass’) and all *characters* containing it (e.g. 花 meaning ‘flower’ or ‘to spend’ and 莓 meaning ‘berry’) (Chen et al., 2004). This metric and family semantic consistency capture different semantic information at different levels; high semantic radical consistency does not guarantee high family semantic consistency across the corpus. Though research thus far indicates no effect of semantic radical consistency on the lexical decision of characters (Chen et al., 2004), it would be interesting to explore this variable and its effects on recognition more thoroughly: for example, studying the extent to which radicals are consistent or inconsistent across the characters in which they occur, exploring how semantic radical consistency relates to the character-level consistency studied in this paper, and investigating under what circumstances (if any) radical consistency influences whole word processing.

**Concluding Remarks**

In conclusion, the present study revealed that the visual recognition of Chinese compounds is an analytic process, influenced by the frequency with which constituent characters occur and convey reliable semantic information in distinct compound words. Chinese characters are often ambiguous in meaning(s), and the fact that they tend not to occur in isolation suggests that readers’ knowledge of these characters is accumulated through experiences with them in different compound words. Once such knowledge is acquired, readers have the capability to generalise it in their daily encounters with familiar or unfamiliar compound words. Our work suggests that characters that occur in many different compound words may yield the most robust representations perhaps as a consequence of general learning mechanisms. Likewise, we have provided evidence that readers’ knowledge of Chinese characters accumulates in a position-specific manner, and that this may signify that position contributes systematically to character meaning. It will be important in future work to understand the extent to which readers take account of the meanings of individual characters *in combination* during the recognition process.

**References**

Agresti, A. (2012). *Categorical data analysis* (3rd ed.). John Wiley & Sons.

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Amenta, S., Günther, F., & Marelli, M. (2020). A (distributional) semantic perspective on the processing of morphologically complex words. *The Mental Lexicon*, *15*(1), 62–78. https://doi.org/10.1075/ml.00014.ame

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*(3), 390–405. https://doi.org/10.1006/jmla.1999.2681

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.3929/ethz-b-000240890

Brysbaert, M., & Diependaele, K. (2013). Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice. *Behavior Research Methods*, *45*(2), 422–430. https://doi.org/10.3758/s13428-012-0270-5

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant’s age. *Frontiers in Psychology*, *7*, 1116. https://doi.org/10.3389/fpsyg.2016.01116

Ceccagno, A., & Basciano, B. (2007). Compound headedness in Chinese: An analysis of neologisms. *Morphology*, *17*(2), 207–231. https://doi.org/10.1007/s11525-008-9119-0

Chen, M. J., & Weekes, B. S. (2004). Effects of semantic radicals on Chinese character categorization and character decision. *Chinese Journal of Psychology*. *46*(2-3), 181–196. http://doi.org/10.6129/CJP.2004.460203.07

Chinese Knowledge Information Processing Group. (2001). *Word list with accumulated word frequency in Sinica Corpus 3.1*. Taipei: Institute of Information Science, Academic Sinica

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). ‘Fell’ primes ‘fall’, but does ‘bell’ prime ‘ball’? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, *63*(1), 83–99. https://doi.org/10.1016/j.jml.2010.03.002

Crepaldi, D., Rastle, K., Davis, C. J., & Lupker, S. J. (2013). Seeing stems everywhere: Position-independent identification of stem morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 510–525. https://doi.org/10.1037/a0029713

Cui, L., Wang, J., Zhang, Y., Cong, F., Zhang, W., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology*, *74*(4), 610–633. https://doi.org/10.1177/1747021820973661

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., & Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language*, *81*(1–3), 555–567. https://doi.org/10.1006/brln.2001.2547

De Jong, N. H., Schreuder, R., & Harald Baayen, R. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, *15*(4–5), 329–365. https://doi.org/10.1080/01690960050119625

Faraway, J. J. (2016). *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*. CRC press.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*(5), 263–279. https://doi.org/10.1017/S0140525X11001841

Günther, F., & Marelli, M. (2019). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1872–1882. https://doi.org/10.1037/xlm0000677

Günther, F., & Marelli, M. (2020). Trying to make it work: Compositional effects in the processing of compound “nonwords”. *Quarterly Journal of Experimental Psychology*, *73*(7), 1082–1091. https://doi.org/10.1177/1747021820902019

Harrison, X. A. (2015). A comparison of observation-level random effect and beta-binomial models for modelling overdispersion in binomial data in ecology & evolution. *PeerJ*, *3*, e1114. https://doi.org/10.7717/peerj.1114

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 157. https://doi.org/10.1037/xlm0000819

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 385–402. https://doi.org/10.1037/a0038995

Hsiang, T. P., Graham, S., Wang, Z., & Gong, Y. (2021). How Chinese characters are taught: An analysis of three popular textbooks used in Macao. *Technology, Knowledge and Learning*, *26*(2), 355–381. https://doi.org/10.1007/s10758-021-09492-8

Hsiao, Y., Bird, M., Norris, H., Pagán, A., & Nation, K. (2020). The influence of item-level contextual history on lexical and semantic judgments by children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(12), 2367–2383. https://doi.org/10.1037/xlm0000795

Hsieh, C.-Y., Lin, W.-C., Li, M.-F., & Wu, J.-T. (2021). The concept of neighborhood underlying the phonetic consistency effect in Chinese character recognition: A four-decade review and beyond. *Reading and Writing*, *34*(5), 1225–1246. https://doi.org/10.1007/s11145-020-10119-0

Hsieh, C.-Y., Marelli, M., & Rastle, K. (2023, August 1). Beyond Quantity of Experience: Exploring the Role of Semantic Consistency in Chinese Character Knowledge. Retrieved from osf.io/thfq6

Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2016). Mean centering helps alleviate “micro” but not “macro” multicollinearity. *Behavior Research Methods*, *48*(4), 1308–1317. https://doi.org/10.3758/s13428-015-0624-x

Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2017). Mean centering, multicollinearity, and moderators in multiple regression: The reconciliation redux. *Behavior Research Methods*, *49*(1), 403–404. https://doi.org/10.3758/s13428-016-0827-9

Jaeger, B. (2017). r2glmm: Computes R squared for mixed (multilevel) models, *R package version 0.1.2*. Available at https://github.com/bcjaeger/r2glmm.

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876–895. https://doi.org/10.1037/a0013484

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language*, *115*, 104152. https://doi.org/10.1016/j.jml.2020.104152

Li, M.-F., Gao, X.-Y., Chou, T.-L., & Wu, J.-T. (2017). Neighborhood frequency effect in Chinese word recognition: Evidence from naming and lexical decision. *Journal of Psycholinguistic Research*, *46*(1), 227–245. https://doi.org/10.1007/s10936-016-9431-5

Li, M.-F., Lin, W.-C., Chou, T.-L., Yang, F.-L., & Wu, J.-T. (2015). The role of orthographic neighborhood size effects in Chinese word recognition. *Journal of Psycholinguistic Research*, *44*(3), 219–236. https://doi.org/10.1007/s10936-014-9340-4

Liang, F., Blythe, H. I., Bai, X., Yan, G., Li, X., Zang, C., & Liversedge, S. P. (2017). The role of character positional frequency on Chinese word learning during natural reading. *PloS one*, *12*(11), e0187656. https://doi.org/10.1371/journal.pone.0187656

Libben, G. (2010). Compound words, semantic transparency, and morphological transcendence. In S. Oslen (Ed.) *New Impulses in Word-Formation* (pp. 317–330). Helmut Buske Verlag.

Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, 31(1–2), 8–25. https://doi.org/10.1080/02643294.2013.874994

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, *39*(2), 192–198. https://doi.org/10.3758/BF03193147

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, *147*, 1–20. https://doi.org/10.1016/j.cognition.2015.10.013

Lüdecke, M. D. (2023). sjPlot: Data visualization for statistics in social science. *R package version 2.8.14*. https://CRAN.R-project.org/package=sjPlot.

Mak, M. H. C., Hsiao, Y., & Nation, K. (2021). Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language*, *118*, 104203. https://doi.org/10.1016/j.jml.2020.104203

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (OSC) estimates for 15,017 English words. *Behavior Research Methods*, *50*(4), 1482–1495. https://doi.org/10.3758/s13428-018-1017-8

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography-Semantics Consistency on word recognition. *Quarterly Journal of Experimental Psychology*, *68*(8), 1571–1583. https://doi.org/10.1080/17470218.2014.95970

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515. https://doi.org/10.1037/a0039267

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as Abstract Operation in Semantic Space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition*, *166*, 207–224. https://doi.org/10.1016/j.cognition.2017.05.026

Marelli, M., Traficante, D., & Burani, C. (2020). Reading morphologically complex words: Experimental evidence and learning models. In V. Pirrelli, I. Plag & W. U. Dressler, (Eds.), *Word Knowledge and Word Usage* (pp. 553–592). de Gruyter Mouton. https://doi.org/10.1515/9783110440577-014

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*. https://doi.org/10.48550/arXiv.1310.4546

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PloS one*, *12*(2), e0171935. https://doi.org/10.1371/journal.pone.0171935

Myers, J. (2006). Processing Chinese Compounds: A survey of the literature. In G. Libben & G. Jarema (Eds.), *The Representation and Processing of Compound Words* (pp. 169–196). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199228911.003.0008

Myers, J. (2022). Wordhood and disyllabicity in Chinese. In C.-R. Huang, Y.-H. Lin, & I.-H. Chen (Eds.), *The Cambridge Handbook of Chinese Linguistics* (pp. 47–73). Cambridge University Press. https://doi.org/10.1017/9781108329019.005

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed‐effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

National Language Committee (2000). *Report of frequently used characters and words*. Taipei: Ministry of Education. Retrieved from https://language.moe.gov.tw/001/Upload/files/SITE\_CONTENT/M0001/PIN/c11.htm?open

Norman, R., Hulme, R. C., Sarantopoulos, C., Chandran, V., Shen, H., Rodd, J. M., … & Taylor, J. H. (2022). Contextual diversity during word learning through reading benefits generalisation of learned meanings to new contexts. *Quarterly Journal of Experimental Psychology*, *76*(7), 1658–1671. https://doi.org/10.1177/17470218221126976

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: some implications of research on Chinese for general theories of reading. *Psychological Review*, *112*(1), 43–59. https://doi.org/10.1037/0033-295X.112.1.43

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*(4–5), 445–485. https://doi.org/10.1080/01690960050119661

Rastle, K. (2019a). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54. https://doi.org/10.1016/j.cortex.2018.02.008

Rastle, K. (2019b). EPS mid-career prize lecture 2017: Writing systems, reading, and language. *Quarterly Journal of Experimental Psychology*, *72*(4), 677–692 https://doi.org/10.1177/1747021819829696

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7–8), 942–971. https://doi.org/10.1080/01690960802069730

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(4–5), 507–537. https://doi.org/10.1080/01690960050119689

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother’s brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10.3758/BF03196742

Rastle, K., & Taylor, J. S. H. (2018). Print-sound regularities are more important than print-meaning regularities in the initial stages of learning to read: Response to Bowers & Bowers (2018). *Quarterly Journal of Experimental Psychology*, *71*(7), 1501–1505. https://doi.org/10.1177/1747021818775053

Rueckl, J. G., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language*, *68*(1–2), 110–117. https://doi.org/10.1006/brln.1999.2106

Schmidtke, D., Gagné, C. L., Kuperman, V., & Spalding, T. L. (2018). Language experience shapes relational knowledge of compound words. *Psychonomic Bulletin & Review*, *25*(4), 1468–1487. https://doi.org/10.3758/s13423-018-1478-x

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, *37*(1), 118–139. https://doi.org/10.1006/jmla.1997.2510

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an ‘outlier’ orthography. *Psychological Bulletin*, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Shen, W., Li, X., & Pollatsek, A. (2018). The processing of Chinese compound words with ambiguous morphemes in sentence context. *Quarterly Journal of Experimental Psychology*, *71*(1), 131–139. https://doi.org/10.1080/17470218.2016.1270975

Siegelman, N., Rueckl, J. G., Lo, J. C. M., Kearns, D. M., Morris, R. D., & Compton, D. L. (2022). Quantifying the regularities between orthography and semantics and their impact on group- and individual-level behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(6), 839–855. https://doi.org/10.1037/xlm0001109

Singh, R. (206). Whole word morphology. In K. Brown, (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed.) (pp. 578–579). Oxford: Elsevier. https://doi.org/10.1016/B0-08-044854-2/04248-6

Stevens, P., & Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychonomic Bulletin & Review*, *29*(5), 1673–1702. https://doi.org/10.3758/s13423-022-02086-0

Su, T., & Lee, H. (2017). Learning Chinese word representations from glyphs of characters. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 264–273. https://doi.org/10.18653/v1/D17-1025

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39. https://doi.org/10.1016/j.cogpsych.2015.03.003

Tsang, Y.-K., & Chen, H.-C. (2010). Morphemic ambiguity resolution in Chinese: Activation of the subordinate meaning with a prior dominant-biased context. *Psychonomic Bulletin & Review*, *17*(6), 875–881. https://doi.org/10.3758/PBR.17.6.875

Tsang, Y.-K., & Chen, H.-C. (2013). Morpho-semantic processing in word recognition: Evidence from balanced and biased ambiguous morphemes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1990–2001. https://doi.org/10.1037/a0033701

Tse, C.-S., Chan, Y. L., Yap, M. J., & Tsang, H. C. (2022). The Chinese Lexicon Project II: A megastudy of speeded naming performance for 25,000+ traditional Chinese two-character words. *Behavior Research Methods*, 1–21. https://doi.org/10.3758/s13428-022-02022-z

Tse, C.-S., & Yap, M. J. (2018). The role of lexical variables in the visual recognition of two-character Chinese compound words: A megastudy analysis. *Quarterly Journal of Experimental Psychology*, *71*(9), 2022–2038. https://doi.org/10.1177/1747021817738965

Tse, C.-S., Yap, M. J., Chan, Y.-L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*, *49*(4), 1503–1519. https://doi.org/10.3758/s13428-016-0810-5

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers’ sensitivity to meaningful regularities in English writing. *Cognition*, *195*, 103810. https://doi.org/10.1016/j.cognition.2018.09.013

Ulicheva, A., Marelli, M., & Rastle, K. (2021). Sensitivity to meaningful regularities acquired through experience. *Morphology*, *31*(3), 275–296. https://doi.org/10.1007/s11525-020-09363-5

Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology*, *31*(6), 1261–1267.e3. https://doi.org/10.1016/j.cub.2020.12.017

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wang, H.-C., Hsu, L.-C., Tien, Y.-M., & Pomplun, M. (2014). Predicting raters’ transparency judgments of English and Chinese morphological constituents using latent semantic analysis. *Behavior Research Methods*, *46*(1), 284–306. https://doi.org/10.3758/s13428-013-0360-z

Wang, J., & Leland, C. H. (2011). Beginning students’ perceptions of effective activities for Chinese character recognition. *Reading in a Foreign Language*, *23*(2), 208–224. http://hdl.handle.net/10125/66851

Wu, J. T, Yang, F. L. & Lin, W. C. (2013) Beyond phonology matters in character recognition. Chinese Journal of Psychology, *55*(3), 289–318. https://doi.org/10.6129/CJP.20130608

Wu, J. T., & Liu, I. M. (1988). A data base system about the psychological features of Chinese characters and words. In I. M. Liu, H.-C. Chen, & M. J. Chen. (Eds.) *Cognitive Aspects of the Chinese Language* (pp. 171–186). Asian Research Service.

Wu, X., Li, W., & Anderson, R. C. (1999). Reading instruction in China. *Journal of Curriculum Studies*, *31*(5), 571–586. https://doi.org/10.1080/002202799183016

Xiong, J., Yu, L., Veldre, A., Reichle, E. D., & Andrews, S. (2023). A multitask comparison of word-and character-frequency effects in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. *49*(5), 793–811 https://doi.org/10.1037/xlm0001192

Xiong, J., Zhang, Y., & Ju, P. (2021). The effects of orthographic neighborhood size and the influence of individual differences in linguistic skills during the recognition of Chinese words. *Frontiers in Psychology*, *12*, 727894. https://doi.org/10.3389/fpsyg.2021.727894

Yan, M., Kliegl, R., Richter, E. M., Nuthmann, A., & Shu, H. (2010). Flexible saccade-target selection in Chinese reading. *Quarterly Journal of Experimental Psychology*, *63*(4), 705–725. https://doi.org/10.1080/17470210903114858

Yao, P., Staub, A., & Li, X. (2022). Predictability eliminates neighborhood effects during Chinese sentence reading. *Psychonomic Bulletin & Review*, *29*(1), 243-252. https://doi.org/10.3758/s13423-021-01966-1

Yip, P.-C. (2000). *The Chinese lexicon: A comprehensive survey*. Routledge. https://doi.org/10.4324/9780203442265

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

1. Although about 70% of Chinese characters contain phonetic radicals, and while these radicals may influence the recognition process (Perfetti et al., 2005), they do not systematically reflect sound like letters do in alphabetic writing systems. Unlike in alphabetic systems, it is nearly impossible to determine the sound of a pseudo-character, as the pronunciation of each Chinese character can only be acquired only through rote memorisation (see Hsieh et al., 2021 for review). [↑](#footnote-ref-1)