

# Computational Simulation and Experimental Verification of Chinese Lexical Semantic Processing

**Authors:** Li Xingshan, Zhang Qiwei, Huang Linjieqiong, Li Xingshan, Zhang Qiwei

**Date:** 2024-05-18T00:00:00+00:00

## Abstract

Chinese is a writing system with distinct characteristics that is widely used by Chinese populations worldwide. Due to its specificity, Western linguistic theories and models cannot be directly applied to Chinese. Existing research on Chinese lexical processing lacks systematic computational models to simulate the process of lexical semantic processing. This study aims to address the aforementioned issues through computational modeling and experimental research methods. The research will systematically review existing studies on Chinese lexical processing and conduct meta-analyses, constructing a model to simulate the processing of Chinese words in both isolated presentation and sentence contexts. The model is capable of processing both single-character and multi-character words, simulating the processing of and interactions among orthographic, phonological, and semantic representations, while also accounting for contextual influences. Finally, model hypotheses will be validated through experimental studies. The Chinese lexical semantic processing model established in this study will contribute to understanding the unique cognitive mechanisms of Chinese reading and the dynamic processes of lexical processing.

## Full Text

### Computational Modeling and Experimental Validation of Chinese Lexical and Semantic Processing

**LI Xingshan**<sup>1,2</sup>, **ZHANG Qiwei**<sup>1,2</sup>, **HUANG Linjieqiong**<sup>1</sup>

<sup>1</sup>CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, 100101, China

<sup>2</sup>Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049, China

**Abstract:** Chinese is a writing system widely used by Chinese people worldwide and has many distinct characteristics. Due to its uniqueness, theories and models of alphabetic languages cannot be directly applied to Chinese. Previous Chinese studies lack systematic computational models for lexical and semantic processing. To address this issue, this study first plans to conduct a systematic review and meta-analysis of previous literature. Next, computational models will be constructed to simulate the processing of Chinese words presented in isolation and during natural reading. The model has the following characteristics: a) it can process both single-character and multi-character words; b) it can simulate orthographic, phonological, and semantic processing of words, as well as their interactions; c) it can simulate the impact of contextual cues on word processing during sentence comprehension. Finally, the assumptions and predictions of the model are planned to be validated in experimental studies. The established model can guide experimental research and has significant theoretical significance. The research findings will help clarify the cognitive mechanisms of Chinese reading and the dynamic process of lexical processing.

**Keywords:** cognitive simulation, lexical processing, semantic processing, Chinese reading, computational modeling

## 1 Research Significance

In recent years, computational modeling has played an important role in understanding language cognition. On one hand, by continuously refining model parameters and structures, cognitive processing can be accurately characterized. On the other hand, models can simulate and explain numerous experimental findings within a unified framework, including data collected with different research designs and even inconsistent results. Western scholars have conducted extensive research on reading mechanisms in alphabetic languages and proposed corresponding theories and models (Dehaene, 2009; Engbert & Kliegl, 2011; Reichle, 2021; Reichle et al., 1998; Seidenberg, 2017). However, researchers have recognized that findings and computational models from alphabetic languages do not always generalize to different writing systems (Share, 2008).

Chinese is a logographic script with many unique features. First, Chinese uses characters rather than letters. Compared to letters or words, Chinese characters and their strings carry more semantic information and less phonological information. Second, Chinese words are typically short, mostly one or two characters, whereas words in languages like German and Finnish often exceed 10 letters. Finally, Chinese text lacks spaces to mark word boundaries, requiring readers to segment and identify words from continuous character strings. Therefore, models developed for alphabetic scripts are not always applicable to logographic systems (Li et al., 2022). Currently, no computational model has been specifically designed for Chinese lexical and semantic processing. Developing such a model would facilitate understanding of Chinese lexical processing mechanisms and guide future research. Moreover, comparing Chinese lexical processing models with those based on alphabetic languages would help illumi-

nate universal and script-specific mechanisms of reading cognition.

This research focuses on Chinese lexical and semantic processing, aiming to propose original theoretical hypotheses based on analyses and summaries of previous experimental findings, and to establish computational models that simulate semantic processing of isolated words. It will further explore how contextual factors influence lexical semantic processing and construct models capable of simulating lexical semantic processing during sentence reading. Finally, experiments will validate important assumptions and predictions of the model, which will be refined and optimized based on experimental results. This research is novel, interdisciplinary, and pioneering, contributing to understanding the script-specific cognitive mechanisms of Chinese reading and advancing cross-linguistic comparison and language cognitive science.

## 2.1 Cognitive Mechanisms of Chinese Compound Word Processing

A central question in lexical processing research is whether words are processed holistically or through decomposition. Research findings in this area are diverse, influenced by methodology, tasks, and corpora. In Chinese reading, some studies support holistic processing. First, eye-tracking experiments show that disrupting holistic word processing (e.g., inserting spaces within words) slows lexical processing (Bai et al., 2008; Li et al., 2013). Second, a word superiority effect exists in both alphabetic scripts and Chinese (申薇, 李兴珊, 2012; Reicher, 1969), where characters in real words (e.g., “审” in “审查”) are recognized faster and more accurately than characters in nonwords (e.g., “审” in “审卖”). Finally, foveal and parafoveal processing studies show larger preview effects for three-character idioms than phrases, suggesting idioms are represented and processed as independent units (Yu et al., 2016; Zang, 2019).

Other studies focus on character frequency effects, assuming they reflect decompositional processing. However, results are inconsistent: some find facilitative character frequency effects (王春茂, 彭聃龄, 1999; Tse & Yap, 2018; G. Yan et al., 2006), others find inhibitory effects (Cui et al., 2021; Xiong et al., 2023; Yu et al., 2021), and some find no significant effects (Li et al., 2014; Ma et al., 2015). A key feature of Chinese is that characters have rich and independent information representations; in most cases, one character corresponds to one morpheme, representing orthographic form, phonology, and rich semantics. Cao et al. (2023) found that Chinese readers can use morphemic features to aid word segmentation and identification. These script-specific features may make Chinese compound word processing mechanisms more complex.

Beyond orthographic processing, debates about holistic versus decompositional processing also exist at the semantic level. Masked priming studies show faster target recognition when primes and targets share morphemic semantics compared to when they share only orthographic form but differ in morphemic meaning (Tsang et al., 2014; Tsang & Chen, 2013; X. Zhou & Marslen-Wilson, 2000),

indicating morphemic semantics are processed during early lexical identification. Neuroimaging evidence also supports semantic decomposition. The left anterior temporal lobe has been implicated in semantic composition in phrase processing (Bemis & Pykkänen, 2011; Flick et al., 2018; Pykkänen, 2020; Westerlund & Pykkänen, 2014). A magnetoencephalography study comparing neural activity during processing of coordinate, modifier-head, and verb-object compound words versus monomorphemic words found enhanced activation in this region for compound words, suggesting semantic composition processes in Chinese compound word processing (Hsu et al., 2019). However, eye-tracking studies using sentence reading tasks found that constituent morpheme semantics do not participate in sentence semantic integration; instead, semantics are processed holistically at the word level (Shen et al., 2018; J. M. Yang et al., 2012; J. Zhou & Li, 2021).

Previous research has yielded inconsistent results regarding compound word processing mechanisms. Whether processing is holistic or decompositional, and the relationship between morpheme and compound word processing, remain unanswered. Different studies on the same question have produced contradictory results. For example, character frequency effects are inconsistent across studies, making it difficult to determine whether they can serve as evidence for lexical decomposition. Therefore, it is necessary to explore the causes of these contradictory results and whether alternative explanations exist.

## 2.2 The Role of Phonological Pathway in Lexical Processing

Researchers have focused on two questions regarding phonological pathway involvement: whether phonological information is activated during visual word processing, and whether phonology mediates access to lexical semantics.

To probe phonological activation, researchers have used homophonic pseudowords. Studies in both alphabetic scripts and Chinese have found homophonic pseudoword effects: homophonic pseudowords are recognized more slowly than non-homophonic pseudowords due to interference from homophones, suggesting phonological information is activated during visual word processing (Van Orden & Kloos, 2005; X. Zhou et al., 2009). Additionally, sentence reading studies have found facilitative effects of phonological cues (W. Zhou et al., 2018) and tongue-twister effects (McCutchen & Perfetti, 1982; S. Zhang & Perfetti, 1993), further supporting phonological activation during visual word processing (see Rayner et al., 2011).

Two main theories exist regarding phonological involvement in Chinese lexical semantic comprehension. Phonological mediation theory posits that phonology is an essential component of lexical identification, providing early constraints rather than being an optional process or byproduct (Tan & Perfetti, 1997, 1998, 1999). Leck et al. (1995) used semantic classification tasks and found that performance on structurally complex single-character words was influenced by both orthographic and phonological information, while performance on simple-structure

single-character words was influenced only by orthography. Another study on Chinese two-character word identification found that homophonic primes with different meanings produced greater interference on target word semantic judgments than unrelated primes (Tan & Perfetti, 1999). Direct semantic access theory argues that orthographic form directly activates semantic information, with no phonological processing in Chinese visual word identification, and/or that semantic activation precedes phonological activation (Tan & Perfetti, 1997; Wong et al., 2014). Early behavioral studies used priming paradigms with manipulated intervals between primes and targets to examine the time course of orthographic, phonological, and semantic activation, assuming that the shortest interval showing priming effects reflects the time needed to activate representations. Different studies support different theories. Perfetti and Zhang (1995) found that phonological interference effects emerged earlier than semantic interference in single-character word identification tasks, consistent with phonological mediation theory. However, Zhou et al. (2000) found earlier semantic than phonological activation in two-character word judgment tasks, while in lexical naming tasks, phonological and semantic activation were nearly simultaneous—results not supporting phonological mediation. Such studies have two common problems: First, Perfetti et al. (2005) noted that relative time course cannot serve as evidence for phonological mediation because semantic activation might be slower than phonological activation even without phonological mediation. Second, these experimental manipulations differ substantially from natural reading, raising concerns about external validity.

Current research has not adequately addressed phonological processing in Chinese lexical identification, and the role of phonological pathway activation in semantic access requires further investigation. While some evidence suggests phonological information is activated during Chinese reading, whether semantics can be accessed via phonological mediation remains unresolved.

### 2.3 Context Effects on Chinese Word Processing

In natural reading, lexical processing does not occur in isolation but is embedded within sentences, influenced by neighboring words and sentence context. Word processing in sentences is more complex than processing isolated words in several ways.

First, Chinese sentences lack explicit word boundary markers, so Chinese reading first requires determining which characters constitute a word. Without low-level visual cues like spaces, Chinese readers must rely on higher-level information such as linguistic knowledge and context for word segmentation, unlike alphabetic scripts. Research on word segmentation mechanisms reveals two processing stages: in the first stage, readers generate initial segmentation based on local information such as word frequency; in the second stage, readers verify this initial segmentation against sentence context, with incorrect segmentation requiring additional correction time (Huang et al., 2021; Huang & Li, 2020; Inhoff & Wu, 2005; Li et al., 2009; Ma et al., 2014; J. Zhou & Li, 2021). Lexi-

cal processing models based on alphabetic languages struggle to simulate these Chinese-specific word and sentence processing characteristics.

Second, lexical processing during sentence reading is influenced by visual attention and reading direction. In studies of isolated Chinese words, all characters of a word typically appear simultaneously within the reader's perceptual span and can be processed together. However, during sentence reading, the eyes move from left to right with variable fixation positions. Consequently, multiple characters within a word may be processed at different times, with characters near the fixation point processed earlier than those farther away. Additionally, a single fixation can process multiple characters that may belong to the same or different words. These phenomena, absent in isolated word processing, may affect lexical processing mechanisms.

Third, sentence context influences lexical processing. Studies show that high-predictability words have shorter reading times than low-predictability words, indicating that readers continuously use prior context to make predictions that facilitate lexical semantic access (刘志方 et al., 2020; Cui et al., 2022; Rayner et al., 2005; Yao et al., 2022). Context also helps readers quickly select appropriate meanings for ambiguous words (Shen & Li, 2016). For example, “火星” can refer to the planet Mars or something that causes fire, with the specific meaning determined by context.

Thus, lexical processing in sentences differs from isolated word processing, and findings from isolated word studies may not apply to sentence reading. Word processing in sentences is often more complex than when words are presented alone, and the underlying mechanisms may differ. Current models of lexical processing rarely consider sentence context effects. Therefore, to fully understand the cognitive mechanisms of Chinese lexical semantic processing, it is necessary to further investigate the role of sentence features and context in lexical cognition.

## 2.4 Development of Cognitive Models of Lexical Processing

To understand language processing mechanisms, scholars have developed many cognitive models. The interactive activation model (McClelland & Rumelhart, 1981) has been highly influential, successfully explaining how bottom-up and top-down information interact. Its interactive activation framework applies to most cognitive processes. In reading, language content can only be understood by integrating visual text information with stored linguistic and world knowledge. Another well-known lexical processing model is the triangle model, which uses connectionist principles to assume distributed representations of orthographic, phonological, and semantic information, with lexical processing occurring through three bidirectionally connected subnetworks. This model simulates reading aloud from orthography (Plaut et al., 1996; Seidenberg & McClelland, 1989) and semantic access from orthography (Harm & Seidenberg, 2004). Plaut et al. (1996) focused on how readers pronounce irregular spelling-sound corre-

spondence words, arguing that both regular and irregular words are processed through a neural network-like system without separate pathways. Knowledge about mappings between orthography and phonology is encoded in weighted connections between units and adjusted through learning. This learning process, driven by statistical regularities of word frequencies in the training set, captures continuities between regular and irregular patterns, naturally reflecting these variations in internal representations and blurring distinctions between regular and irregular words, enabling flexible processing of irregular input. Harm and Seidenberg (2004) used the triangle model to successfully simulate cooperative division of labor between phonological mediation and direct semantic pathways in English lexical processing. While connectionist models can simulate some behavioral results, they have many parameters and limited interpretability.

These models are important for understanding alphabetic script mechanisms, particularly in mapping orthography to phonology. However, they are generally not applicable to Chinese due to fundamental differences. In alphabetic scripts, letters correspond to phonemes, and phoneme blending quickly yields word pronunciation for lexical identification and semantic access (Tan & Perfetti, 1997). In Chinese, each character corresponds to a syllable; readers cannot blend sub-character components to derive pronunciation and can only obtain it after recognizing the whole character. Therefore, some scholars argue that Chinese semantic access relies less on phonological mediation (Perfetti et al., 2005).

Previous computational models of Chinese reading have been important for guiding research (Li et al., 2009; Li & Pollatsek, 2020; Perfetti et al., 2005; J. F. Yang et al., 2009). However, existing Chinese reading models have several problems. First, most models focus on character-level or single-character word processing (Perfetti et al., 2005; J. F. Yang et al., 2009). These models introduce radical and character representation nodes to simulate effects of radical features and character frequency on character recognition (e.g., Chang et al., 2016; Hsiao & Shillcock, 2004, 2005; Xing et al., 2002, 2004). Although Perfetti et al. (2005) further introduced semantic and phonological nodes to enable output from orthography to phonology and semantics, their model remains limited to single-character word identification. However, in modern Chinese, multi-character words comprising two or more characters constitute the majority. Processing mechanisms differ between single- and multi-character words; multi-character word identification involves questions of holistic versus decompositional processing, and if decompositional, how individual morpheme meanings combine to form multi-character word semantics and how ambiguous characters' specific meanings are determined within words.

Second, based on the interactive activation framework, researchers have proposed model frameworks for Chinese multi-character word processing to explain important findings in Chinese character and word recognition (Peng et al., 1999; Taft & Nguyen-Hoan, 2010; Tan & Perfetti, 1999). Tan and Perfetti's (1999) visual word recognition model includes orthographic, phonological, and semantic modules with interactive activation between them, and contains bidirectionally

connected character and multi-character word representations in orthographic and phonological modules. This model assumes character identification facilitates multi-character word recognition but cannot explain inhibitory character frequency effects found in some studies. X. Zhou and Marslen-Wilson's (2000) multi-level lexical processing framework includes orthographic, phonological, and semantic modules but lacks independent word-form and word-phonology representations, making it difficult to explain widely observed word-based processing patterns in Chinese reading such as the word superiority effect (Chen et al., 2018; Shen et al., 2018; J. M. Yang et al., 2012). Peng et al.'s (1999) model does not involve phonological representation but includes orthographic and lexical representation systems, storing characters in the orthographic system and morphemes and words in the representation system. Character nodes in the orthographic layer activate corresponding morpheme and lexical nodes in the representation layer through hierarchical connections, with connection strength increasing with word frequency. Within the lexical representation layer, semantically transparent words have excitatory connections between morpheme and lexical nodes, while opaque words have inhibitory connections. This model predicts that semantic transparency and whole-word frequency jointly modulate character frequency effects, but it is unclear when semantic processing begins and how it relates to word-form representation. Taft and Zhu's (1997) initial framework for Chinese multi-character word processing included character and multi-character word representations in orthographic, phonological, and semantic modules. In their updated framework, Taft et al. (1999) added an abstract lemma layer connecting orthography, phonology, and semantics, without whole-word orthographic or phonological representations. However, what information the lemma layer represents and the role of phonological pathways in Chinese multi-character word processing remain unclear. Notably, these Chinese lexical processing models based on interactive activation frameworks have not been implemented as computer programs or fitted to experimental data. Strictly speaking, these frameworks are not computational models and cannot simulate the complex dynamic processes of word processing. Thus, multi-character word processing mechanisms are central to understanding Chinese lexical processing, but no formal computational model currently addresses this issue.

Third, most implemented lexical models have focused on orthographic processing and pathways from orthography to phonology, with less involvement in semantic processing of multi-character words (Xing et al., 2002, 2004; J. F. Yang et al., 2006, 2009, 2013). Given the complex character-word relationships in Chinese, lexical semantic processing remains challenging. For lexical processing in sentence reading, models must also explain how context influences lexical semantic access. Li and Pollatsek's (2020) Chinese Reading Model (CRM) simulates word segmentation, word identification, and eye-movement control during sentence processing, including character and word recognition. The model assumes that when the eyes fixate a location, all perceivable characters are recognized in parallel. All words formed by these recognized characters are activated and compete; when a word wins the competition, it is identified and segmented. Thus,

the model posits that Chinese word segmentation and identification occur simultaneously as a unified process. The CRM successfully explains cognitive mechanisms of word segmentation and eye-movement control in Chinese reading. However, it primarily focuses on processing from orthography to lexical entries, without addressing the roles of phonology and semantics. Therefore, the model needs further development to explain the functions of phonology and semantics in reading.

Current research on cognitive models of Chinese lexical processing faces several unresolved issues and lacks systematic computational models of lexical semantic processing. On one hand, models developed for alphabetic scripts cannot be directly applied to Chinese. On the other hand, previous Chinese cognitive models have focused either on orthography-to-phonology processing or orthography-to-lexical-entry identification; no model has comprehensively described the dynamic interactions among orthography, phonology, and semantics during Chinese lexical processing. Moreover, most Chinese reading models have focused only on single-character word processing, lacking simulation of multi-character word processing. These limitations restrict their applicability.

### 3 Problem Statement

This research focuses on Chinese lexical semantic processing, using computational modeling to explore its cognitive mechanisms and experimental validation to test theoretical hypotheses. It will address three core questions:

First, from a whole-part competition perspective, it will examine whether single-character words are activated and compete with whole words during Chinese compound word processing, thereby explaining the relationship between morpheme and compound word processing. Most Chinese multi-character words are compound words composed of two or more morphemes. Previous theories attempting to distinguish holistic versus decompositional processing have many unexplained findings. This study aims to propose and validate a new theoretical hypothesis: during compound word processing, local single-character word representations and whole-word representations are activated at visual, phonological, and semantic levels and compete with each other, with the winning competitor being selected.

Second, it will examine the division of labor between phonological mediation and direct semantic pathways during reading. Chinese is a logographic script with different spelling-to-sound mechanisms than alphabetic scripts, and its orthography-phonology-semantics processing mechanisms also differ. Previous Chinese character and word recognition models have only considered single-character word processing and have not addressed phonological pathway issues in multi-character word processing. However, multi-character and single-character word processing differ. Chinese has many homophonic characters (averaging 4 homophones per character), making semantic access via single-character pronunciation inefficient due to multiple activated semantics. How-

ever, multi-character words have fewer homophones, allowing easier activation of corresponding semantic units, making phonological mediation more likely for semantic access. This research will investigate how phonological mediation and direct semantic pathways interactively cooperate to access semantics during multi-character word processing. Through modeling, it will examine how orthography, phonology, and semantics are dynamically activated during Chinese lexical processing, focusing on the cognitive framework and processes, particularly how phonological information is activated, whether activated phonology can access semantics, and the relationship between phonological mediation and direct semantic pathways.

Third, from a sentence reading perspective, it will explore cognitive mechanisms of how context influences lexical semantic processing. Current lexical processing research and models rarely consider sentence context effects. During sentence reading, lexical processing is often influenced by context. This study will examine what assumptions are needed to model context effects on lexical semantic processing during sentence reading, including how context influences ambiguous word comprehension, word segmentation, and word prediction. It will test two hypothesized processing modes: (1) before seeing the target word, context pre-activates some lexical semantic representations, affecting target word processing; (2) after seeing the target word, prior context influences lexical processing by affecting competitive advantages among different semantic representations. Notably, these hypotheses are not mutually exclusive. During modeling, the model may use one or both mechanisms to simulate different tasks.

In summary, Chinese lexical processing is a complex cognitive process. This research addresses this issue from different perspectives of lexical processing using various methods, comprising four parts. Part 1 will comprehensively review Chinese lexical processing mechanisms through meta-analysis to understand overall findings, identify potential factors influencing different results, and propose theoretical hypotheses. Based on these hypotheses, Part 2 will construct computational models simulating semantic processing of Chinese isolated words. These models will simulate dynamic activation of orthography, phonology, and semantics during lexical processing and fit classic findings. By temporarily ignoring context effects, model complexity remains manageable. Building on this, Part 3 will construct computational models simulating lexical semantic processing during sentence reading, focusing on problems arising during sentence processing and explaining inconsistencies between isolated word and sentence reading findings. Part 4 will design experiments to validate important model assumptions, particularly the competition mechanism between single-character and whole words proposed for compound word processing. Experimental results will help refine the model and iteratively improve it to better understand Chinese lexical semantic processing.

## 4 Research Plan

This research focuses on Chinese lexical semantic processing, using meta-analysis, computational modeling, and experimental methods to investigate compound word processing, interactions among orthographic-phonological-semantic modules, and lexical processing in sentence reading. It aims to innovate theoretical hypotheses, construct Chinese-specific computational models of lexical semantic processing, and validate them experimentally. This work will reveal the dynamic process of Chinese lexical semantic processing in both isolated and sentence contexts, elucidating script-specific cognitive mechanisms of Chinese reading. The research will proceed in four parts, with the overall framework shown in Figure 1 [Figure 1: see original paper].

### 4.1 Meta-Analysis of Chinese Lexical Processing Studies

Study 1 will conduct three meta-analyses addressing core scientific questions to estimate effect sizes of morpheme, phonology, and context in compound word processing and test the reliability of previous findings. The three meta-analyses follow similar procedures but differ in inclusion criteria and coding rules.

For the meta-analysis examining morpheme processing effects on compound word identification, the sample will include all experimental studies manipulating compound word morpheme features, with conditions producing stronger morpheme activation as the experimental group and weaker activation as the control group. All included studies will be coded to estimate overall effect size. If morpheme activation affects compound word identification, a significant overall effect should be observed, with effect size reflecting the actual impact. Subsequent analyses will examine heterogeneity and potential sources, using weighted regression to test potential moderators including presentation mode, manipulated morpheme features, and word frequency, and test for publication bias. Moderating effects are expected: (1) larger morpheme effects when target words are presented in isolation versus embedded in sentences; (2) larger morpheme effects when manipulating semantic features than phonological or orthographic features; (3) larger morpheme effects for low-frequency than high-frequency target words. Based on the overall effect size, studies with effect sizes closest to the mean will be selected as representative for subsequent model development.

For the meta-analysis examining phonological effects and whether phonology mediates semantic access, all experimental studies using phonological features as independent variables with visual lexical processing tasks will be included. Phonologically related conditions will serve as the experimental group and unrelated conditions as the control group. Potential moderators causing heterogeneity include experimental paradigm, task, processing of whole versus partial phonology, word frequency, and participant age. If the overall phonological effect is significant, it supports phonological activation during lexical processing. If semantic processing tasks show significant facilitatory phonological effects,

this indicates use of phonological mediation for semantic access. Conversely, absent phonological effects at certain moderator levels suggest weak phonological activation and influence.

Finally, a meta-analysis will examine context effects in lexical processing, including studies manipulating contextual features or relationships between target words and context. Conditions with relatively favorable context will be coded as the experimental group and unfavorable or neutral context as the control group to estimate overall effect size. Context is expected to significantly influence lexical processing, showing a large overall effect. Potential moderators include manipulated variables, task, presentation mode, duration, and participant age. Meta-regression results and effect size differences across variable levels will identify primary internal factors influencing context effects, such as conditions producing larger contextual influences. Meta-analysis results will provide theoretical hypotheses and data support for models of lexical semantic processing during sentence reading.

## 4.2 Construction of Chinese Isolated Word Processing Model

Study 2 will use model simulation to examine the dynamic processing from orthography to semantics in Chinese reading from whole-part competition and modular processing perspectives. It aims to construct computational models of Chinese isolated word processing to explore the roles of morpheme semantics and phonological pathways in lexical semantic processing.

**4.2.1 Model Assumptions** The Chinese isolated word semantic processing model incorporates two key assumptions. First, the model assumes activation of phonological representations during visual lexical semantic processing, thus including orthographic, phonological, and semantic modules with feedforward and feedback connections between them. This means both direct orthography-to-semantics pathways and orthography-phonology-semantics mediation pathways are activated during lexical processing. Model experiments will test this assumption.

Second, when processing compound words, the model assumes that nested single-character words and whole-word representations are activated at orthographic, phonological, and semantic levels, and that they compete. Due to competition, single-character word activation inhibits whole-word activation. Unlike most previous models assuming either holistic or decompositional processing, this model places single-character and whole-word representations at the same level with simultaneous activation, with competition outcomes across orthographic, phonological, and semantic levels determining whether whole words or single-character words win. In model experiments, activation levels of different nodes change dynamically over time, testing this assumption.

**4.2.2 Model Structure and Implementation** The model has a complete triangle structure with orthographic, phonological, and semantic modules, using

hierarchical network structures with nodes representing characters, pronunciations, lexical entries, word pronunciations, and semantics of single- and multi-character words (Figure 2 [Figure 2: see original paper]). Modules interact through interactive activation frameworks enabling bottom-up and top-down information exchange. Basic assumptions are embodied in connection patterns and strengths between and within modules. Model input is visual word information, output is identified lexical entries, meanings, and pronunciations, with dynamic processing simulated through recurrent updating of node activation levels from input to output.

For implementation, visual, character, and lexical entry layers follow CRM implementation. Phonological layer activation occurs when orthographic unit activation exceeds thresholds, activating corresponding character phonology nodes and then all related lexical phonology units. In the semantic layer, semantic information is assumed to be represented in semantic space, with semantically related words closer together. A word's semantics is represented as a set of activated semantic units (Landauer & Dumais, 1997). When word activation exceeds threshold, it activates a small region in semantic space and all units within it. Units at the center have highest activation, decreasing with distance from the center. Semantic space information will be obtained using word embeddings, with 200-dimensional vectors for each word from wordVec pre-training (Mikolov et al., 2013) and cosine similarity calculating semantic distances.

The model assumes interactive activation connections between adjacent layers but no direct connections between nonadjacent layers. This interactive connectivity gives the model novel characteristics. For example, Chinese has many polyphonic characters and homophones, making it difficult for single-character models to determine pronunciation or access semantics directly from phonology. This model processes multi-character word phonology interactively, enabling correct polyphonic character pronunciation determination through feedback connections. Multi-character words have fewer homophones, making it easier to activate accurate semantics and improving phonological mediation efficiency. These features are absent in single-character word models.

**4.2.3 Model Experiments** After model construction, structural and parameter adjustments will examine its characteristics to reveal lexical processing mechanisms. Simulation experiments will test whether model assumptions can explain important existing findings. To examine phonological pathway involvement, model parameters will be adjusted to alter division of labor between pathways. Depending on parameters, the model can activate semantics via only direct semantic pathways, only phonological mediation pathways, or both. Modeling will explore which approach better fits findings to determine final model structure. Additionally, removing model components will test impacts on model behavior, examining whether some findings cannot be fit without phonological mediation, which would provide strong support for its necessity. Simulation experiments will also observe activation timing of phonological and semantic

units, comparing dynamics across processing units to provide new insights into time courses of different lexical semantic processing pathways.

The model will simulate Chinese isolated word processing, evaluating how well competition-based models fit Chinese reading findings. Simulations will include: (1) effects of character properties (e.g., semantic transparency) on processing speed in compound word processing; (2) phonological processing studies examining phonological activation and its role in lexical processing, particularly polyphonic character pronunciation determination; (3) semantic priming effects, including whole-word and morpheme semantic priming. These simulations aim to evaluate model performance, reveal how model structure produces relevant effects, and deepen understanding of cognitive mechanisms.

### 4.3 Sentence Reading Model

Study 3 will integrate the isolated word processing model from Study 2 with CRM's eye-movement control module, incorporating phonological processing, semantic processing, and sentence context effects with new assumptions to construct a lexical semantic processing model during sentence reading that accurately simulates lexical processing and its relationship with eye-movement control. As reading aims to comprehend text semantics, unlike CRM's assumption that visual lexical processing degree is the primary determinant of when the eyes move, the new assumption posits that semantic unit activation level is the primary factor determining saccade timing. This will help more accurately simulate lexical processing during sentence reading and better explain important findings related to semantic processing. Another key assumption is that identified words influence subsequent word processing. After a word's semantics are extracted, semantic units constituting its meaning maintain maximal activation, unaffected by other words being processed, while influencing subsequent word processing. Adjusting connection strengths between semantic units can simulate context effects on ambiguous word meaning selection.

**4.3.1 Model Experiments** The newly constructed model will simulate important findings: (1) Preview effects. Studies using the boundary paradigm examine how semantic relatedness between preview and target words influences fixation times, demonstrating that readers process lexical semantic information before fixation. This is important for understanding information processing at a fixation and whether words are processed in parallel or serially. (2) Ambiguous word meaning selection. Simulating Shen and Li (2016) to explore how context influences ambiguous word meaning selection, showing how readers use contextual information in real-time to select appropriate meanings. (3) Word predictability effects. When readers can predict upcoming words from context, reading times on those words decrease (Rayner et al., 2005). This predictive function of context is an important indicator for testing model validity. (4) Holistic processing of nested words. Simulating how Chinese readers holistically process nested words and integrate them with context, such as processing “老板

娘” being unaffected by contextual plausibility of “老板” (J. M. Yang et al., 2012; J. Zhou & Li, 2021). These findings are important for understanding lexical processing and integration units during sentence reading. If model assumptions are correct, the model should fit these key data patterns; otherwise, further revision is needed.

#### 4.4 Experimental Validation of Core Model Assumptions and Important Predictions

This research plans to use experimental studies to validate important model assumptions and predictions, focusing on validating the competition mechanism between single-character and whole words proposed for compound word processing. Two experiments will examine whether competition exists between partial single-character words and whole multi-character words during processing of isolated words and words embedded in sentences. If model assumptions and predictions are validated, they will provide empirical support for the model and deepen understanding of Chinese lexical processing. If experimental results contradict model assumptions and predictions, model revision will be needed. This process will iteratively improve understanding of Chinese lexical processing cognitive mechanisms.

**Experiment 1** will validate an important assumption of the isolated word processing model: during compound word processing, both whole words and nested single-character words are activated and compete. A 2 (whole-word frequency: high vs. low)  $\times$  2 (single-character word frequency: high vs. low) within-subjects design will be used. The manipulated single-character word frequency refers to the frequency of the first character of a compound word when it appears as a single-character word in the corpus. Chinese single-character words have both character frequency and word frequency features that are highly correlated but represent different concepts. For example, for “树”, character frequency indicates how often the character appears in all text (including “树”, “树立”, “大树”), while word frequency only includes its frequency as a single-character word among all words. Previous research on character frequency effects in compound word processing has not controlled for potential word frequency confounds. This experiment aims to test the model’s assumption of lexical-level competition between single-character and whole words, thus manipulating single-character word frequency while matching character frequency. Chinese two-character compound words will be presented in isolation for lexical decision tasks, with response keys and reaction times recorded. According to the competition hypothesis, higher single-character word frequency should lead to longer lexical decision times and slower processing for two-character words. An interaction between whole-word frequency and single-character word frequency is expected, with stronger inhibitory effects of single-character word frequency when whole-word frequency is low. Without lexical-level competition, single-character word frequency should have no effect on processing speed.

**Experiment 2** will examine lexical processing mechanisms in sentences using

a 2 (semantic similarity between single-character word and whole word: high vs. low)  $\times$  2 (single-character word frequency: high vs. low) within-subjects design. It will test whether nested single-character words are activated and compete with whole words during multi-character word processing in sentences, and whether competition occurs at orthographic or semantic levels. Target words are Chinese two-character compound words embedded in identical sentence frames across four conditions, with eye movements recorded during sentence comprehension. By comparing fixation times on target words, the experiment examines whether compound word processing is influenced by semantic similarity between constituents and the whole word, and by single-character word frequency. If the competition hypothesis holds, inhibitory effects of single-character word frequency are expected, with longer fixation times under high-frequency conditions. If sentence context provides top-down activation reducing reliance on individual character representations, no single-character word frequency effect is expected. If semantic-level competition occurs in sentence reading, single-character word frequency inhibition is expected only when semantic similarity between the single-character word and whole word is low, with longer processing times for compound words containing high-frequency constituents.

## 5 Theoretical Construction

Lexical processing is fundamental to reading comprehension. However, due to the uniqueness of the Chinese writing system where individual characters carry rich semantic information, lexical processing models based on alphabetic languages cannot be directly applied to Chinese (Li et al., 2022). Some Chinese lexical processing models focus only on single-character word processing, introducing radical and character representation nodes to simulate effects of radical features and character frequency on character recognition (Chang et al., 2016; Hsiao & Shillcock, 2004, 2005; Xing et al., 2002, 2004; J. F. Yang et al., 2006, 2009, 2013). Other models for Chinese multi-character word processing emphasize interactive activation between character and word identification but do not involve phonological and semantic processing (Li & Pollatsek, 2020; Yu et al., 2021), or have not been computationally implemented, making it difficult to simulate dynamic lexical processing (Peng et al., 1999; Taft et al., 1999; Taft & Zhu, 1997). Although existing models provide deep understanding of Chinese character and word recognition and reading processes, many challenges remain. For example, Chinese has many compound words, and accurate processing cannot rely solely on concatenating single-character word phonology and semantics. This research focuses on three key scientific questions in Chinese lexical semantic processing, proposes original theoretical hypotheses, extends the established Chinese reading model (Li & Pollatsek, 2020) at phonological and semantic levels, constructs two cognitive computational models for isolated word processing and sentence reading, and designs experiments to validate their assumptions and predictions. This research has the following theoretical significance.

First, regarding Chinese compound word processing mechanisms, previous re-

search has debated whether words are processed holistically or through decompositional pathways (Cao et al., 2016; Taft et al., 1994; Tsang & Chen, 2014; G. Yan et al., 2006). This research proposes a novel possibility: at orthographic and semantic levels, all words that can be formed by characters within the perceptual span are activated and compete, including single-character and multi-character words. For example, when seeing “树立”, the word forms of “树”, “立”, and “树立” are all activated at the lexical level and compete, with the winner being segmented and identified. Similarly, multi-character words and constituent single-character words are activated and compete at the semantic level; increased activation of one reduces activation of others, with the word reaching higher activation and crossing threshold earliest gaining priority for semantic access. This theory has not been reported in previous research. This study proposes a new theoretical hypothesis for compound word processing that may significantly advance Chinese lexical cognitive processing research.

Second, the constructed model can simulate interactions among orthography, phonology, and semantics during lexical processing, revealing the division of labor and functions between direct semantic and phonological mediation pathways. Previous single-character word models suggest Chinese semantic access relies less on phonological mediation partly because numerous homophones reduce semantic access efficiency (Wong et al., 2014; H. Zhang et al., 2020; X. L. Zhou et al., 1999). However, multi-character words, which constitute most Chinese vocabulary, rarely have phonological ambiguity, making it easier to activate corresponding semantic units. Multi-character word processing mechanisms differ from single-character words and are more likely to use phonological mediation for semantic access. The model includes a phonological module: when orthographic unit activation exceeds threshold, corresponding phonological units and all related lexical phonology units are activated. Multi-character word phonological processing is interactive; unlike single-character models, this model can determine correct polyphonic character pronunciation through top-down connections. For example, when “长” appears alone, its pronunciation cannot be determined as “cháng” or “zhǎng”, but in “成长”, “长” can be determined as “zhǎng”. These features are absent in previous single-character word models. The model fully considers dynamic interactions among orthographic, phonological, and semantic modules in multi-character word processing, explaining Chinese lexical processing mechanisms from the perspective of coordination and division of labor among different pathways. This will help refine existing lexical processing theories and compare universality and specificity across language processing mechanisms.

Third, this research constructs computational models of lexical semantic processing during sentence reading that fully consider effects of sentence context on lexical access and eye-movement control. Currently, few computational models have been developed for Chinese lexical processing, and existing models focus only on word segmentation and identification without involving semantic processing (Li & Pollatsek, 2020; Yu et al., 2021). Unlike previous lexical processing models, this model will focus on problems and eye-movement patterns that may

arise during lexical semantic access in sentence processing. A key theoretical assumption is that prior context importantly influences word segmentation and meaning selection. The model will fit important findings from Chinese sentence reading, including but not limited to semantic preview effects (Shen et al., 2018; M. Yan et al., 2009) and semantic plausibility effects (J. M. Yang et al., 2012; Yao et al., 2022; J. Zhou & Li, 2021). By inputting contextual semantic information, adjusting model structure and parameters, and observing dynamic changes in simulations, the model will facilitate deeper understanding of how lexical semantics are accessed during sentence reading, how ambiguous words select meanings, and how semantic information integrates with prior context, and will explain inconsistencies between isolated word and sentence reading findings. Additionally, while experimental research on lexical semantic processing during sentence reading is limited, computational model development will guide experimental studies.

Finally, this research contributes methodologically and technologically. Compared with previous models, the proposed model has three characteristics: first, it can process both single-character and multi-character words; second, it can simulate orthographic, phonological, and semantic processing and their interactions; third, it can simulate context effects on lexical processing during sentence processing. These features enable the model to characterize dynamic Chinese lexical processing and simulate Chinese lexical semantic processing mechanisms more broadly.

In summary, this research leverages computational models' accuracy, systematicity, and ability to describe dynamic processes to assist understanding of complex Chinese character and word recognition, while better guiding experimental research. It organically combines experimental and modeling research, using experiments to validate model assumptions and predictions, with results driving model revision to iteratively improve understanding of Chinese lexical semantic processing mechanisms.

## References

- 曹海波, 兰泽波, 高峰, 于海涛, 李鹏, 王敬欣. (2023). 词素位置概率在中文阅读中的作用: 词汇判断和眼动研究. *心理学报*, 55(2), 159-176. <https://doi.org/10.3724/SP.J.1041.2023.00159>
- 刘志方, 全文, 张智君, 赵亚军. (2020). 语境预测性对阅读中字词加工过程的影响. *心理学报*, 52(9), 1031-1047. <https://doi.org/10.3724/SP.J.1041.2020.01031>
- 申薇, 李兴珊. (2012). 中文阅读中词优效应的特异性. *科学通报*, 57(35), 3414-3420. <https://doi.org/10.1360/972012-666>
- 王春茂, 彭聃龄. (1999). 合成词加工中的词频、词素频率及语义透明度. *心理学报*, 31(3), 266-273.
- Bai, X., Yan, G., Liversedge, S., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology. Human Perception and Performance*, 34, 1277-1287. <https://doi.org/10.1037/0096-1523.34.5.1277>

- Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, 31(8), 2801-2814. <https://doi.org/10.1523/JNEUROSCI.5003-10.2011>
- Cao, H., Gao, M., & Yan, H. (2016). Character decomposition and transposition processes in Chinese compound words modulate attentional blink. *Frontiers in Psychology*, 7, 923. <https://doi.org/10.3389/fpsyg.2016.00923>
- Chang, Y.-N., Welbourne, S., & Lee, C.-Y. (2016). Exploring orthographic neighborhood size effects in a computational model of Chinese character naming. *Cognitive Psychology*, 91, 1-23. <https://doi.org/10.1016/j.cogpsych.2016.09.001>
- Chen, L., Perfetti, C., Leng, Y., & Li, Y. (2018). Word superiority effect for native Chinese readers and low-proficiency Chinese learners. *Applied Psycholinguistics*, 39(6), 1097-1115. <https://doi.org/10.1017/S0142716418000255>
- Cui, L., Wang, J., Zhang, Y., Cong, F., Zhang, W., & Hyona, 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>
- Cui, L., Zang, C., Xu, X., Zhang, W., Su, Y., & Liversedge, S. P. (2022). Predictability effects and parafoveal processing of compound words in natural Chinese reading. *Quarterly Journal of Experimental Psychology*, 75(1), 18-29. <https://doi.org/10.1177/17470218211048193>
- Dehaene, S. (2009). *Reading in the brain: The science and evolution of a human invention*. New York, NY: Penguin Press. <https://api.semanticscholar.org/CorpusID:141548807>
- Engbert, R., & Kliegl, R. (2011). Parallel graded attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements*. (pp. 787-800). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199539789.001.0001>
- Flick, G., Oseki, Y., Kaczmarek, A. R., Al Kaabi, M., Marantz, A., & Pylkkänen, L. (2018). Building phrases in language production: An MEG study of the lexical interface of the temporal lobe. *Cortex*, 101, 113-124. <https://doi.org/10.1016/j.cortex.2018.06.004>
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111(3), 662-720. <https://doi.org/10.1037/0033-295X.111.3.662>
- Hsiao, J. H., & Shillcock, R. (2004). Connectionist modelling of Chinese character pronunciation based on foveal splitting. In *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 601-606). Mahwah, NJ: Lawrence Erlbaum Associates. <http://csjarchive.cogsci.rpi.edu/Proceedings/2004/CogSci04.pdf>

Hsiao, J. H., & Shillcock, R. (2005). Foveal splitting causes differential processing of Chinese orthography in the male and female brain. *Brain Research. Cognitive Brain Research*, 25(2), 531-536. <https://doi.org/10.1016/j.cogbrainres.2005.08.005>

Hsu, C.-H., Pykkänen, L., & Lee, C.-Y. (2019). Effects of morphological complexity in left temporal cortex: An MEG study of reading Chinese disyllabic words. *Journal of Neurolinguistics*, 49, 168-177. <https://doi.org/10.1016/j.jneuroling.2018.06.004>

Huang, L., & Li, X. (2020). Early, but not overwhelming: The effect of prior context on segmenting overlapping ambiguous strings when reading Chinese. *Quarterly Journal of Experimental Psychology*, 73(9), 1382-1395. <https://doi.org/10.1177/1747021820926012>

Huang, L., Staub, A., & Li, X. (2021). Prior context influences lexical competition when segmenting Chinese overlapping ambiguous strings. *Journal of Memory and Language*, 118, 104218. <https://doi.org/10.1016/j.jml.2021.104218>

Inhoff, A. W., & Wu, C. (2005). Eye movements and the identification of spatially ambiguous words during Chinese sentence reading. *Memory & Cognition*, 33(8), 1345-1356. <https://doi.org/10.3758/BF03193367>

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240.

Leck, K. J., Weekes, B. S., & Chen, M. J. (1995). Visual and phonological pathways to the lexicon: Evidence from Chinese readers. *Memory & Cognition*, 23(4), 468-476. <https://doi.org/10.3758/bf03197248>

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, 143(2), 895-913. <https://doi.org/10.1037/a0033580>

Li, X., Gu, J., Liu, P., & Rayner, K. (2013). The advantage of word-based processing in Chinese reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 879-889. <https://doi.org/10.1037/a0030337>

Li, X., Huang, L., Yao, P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology*, 1(3), 133-144. <https://doi.org/10.1038/s44159-022-00022-6>

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, 127(6), 934-965. <https://doi.org/10.1037/rev0000248>

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology*, 58(4), 525-552.

<https://doi.org/10.1016/j.cogpsych.2009.02.003>

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1046-1059. <https://doi.org/10.1037/a0035389>

Ma, G., Li, X., & Rayner, K. (2015). Readers extract character frequency information from nonfixated-target word at long pretarget fixations during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, 41(5), 1409-1419. <https://doi.org/10.1037/xhp0000072>

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375-407. <https://doi.org/10.1037/0033-295X.88.5.375>

McCutchen, D., & Perfetti, C. A. (1982). The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior*, 21(6), 672-687. [https://doi.org/10.1016/S0022-5371\(82\)90870-2](https://doi.org/10.1016/S0022-5371(82)90870-2)

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, 3111-3119.

Peng, D. L., Liu, Y., & Wang, C. (1999). How is access representation organized? The relation of polymorphemic words and their morphemes in Chinese. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 65-89). Erlbaum.

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>

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 24-33. <https://doi.org/10.1037/0278-7393.21.1.24>

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56-115. <https://doi.org/10.1037/0033-295X.103.1.56>

Pylkkänen, L. (2020). Neural basis of basic composition: What we have learned from the red-boat studies and their extensions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791), 20190299. <https://doi.org/10.1098/rstb.2019.0299>

Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review*, 12(6), 1089-1093. <https://doi.org/10.3758/BF03206448>

- Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2011). *Psychology of Reading: 2nd Edition (1st ed.)*. London, UK: Psychology Press. <https://doi.org/10.4324/9780203155158>
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81(2), 275-280. <https://doi.org/10.1037/h0027768>
- Reichle, E. D. (2021). *Computational models of reading: A handbook*. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/oso/9780195370669.001.0001>
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125-157. <https://doi.org/10.1037/0033-295x.105.1.125>
- Seidenberg, M. S. (2017). *Language at the Speed of Sight: How We Read, Why so Many Can't, and What Can Be Done about It*. New York: Basic Books.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568. <https://doi.org/10.1037/0033-295x.96.4.523>
- 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.
- Shen, W., & Li, X. (2016). Processing and representation of ambiguous words in Chinese reading: Evidence from eye movements. *Frontiers in Psychology*, 7, 1713. <https://doi.org/10.3389/fpsyg.2016.01713>
- 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>
- Taft, M., Huang, J., & Zhu, X. (1994). The influence of character frequency on word recognition responses in Chinese. In H. W. Chang, J. T. Hung, C. W. Hue, & O. Tzeng (Eds.), *Advances in the study of Chinese language processing* (pp. 59-73). Taipei: National Taiwan University.
- Taft, M., & Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes*, 25(2), 277-296. <https://doi.org/10.1080/01690960903043261>
- Taft, M., & Zhu, X. P. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology-Learning Memory and Cognition*, 23(3), 761-775. <https://doi.org/10.1037/0278-7393.23.3.761>
- Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language*, 40(4), 498-519. <https://doi.org/10.1006/jmla.1998.2625>

- Tan, L. H., & Perfetti, C. A. (1997). Visual Chinese character recognition: Does phonological information mediate access to meaning? *Journal of Memory and Language*, 37(1), 41-57. <https://doi.org/10.1006/jmla.1997.2508>
- Tan, L. H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. In C. K. Leong & K. Tamaoka (Eds.), *Cognitive Processing of the Chinese and the Japanese Languages* (Vol. 14, pp. 11-46). Springer Netherlands. [https://doi.org/10.1007/978-94-015-9161-4\\_2](https://doi.org/10.1007/978-94-015-9161-4_2)
- Tan, L. H., & Perfetti, C. A. (1999). Phonological activation in visual identification of Chinese two-character words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 382-393.
- Tsang, Y.-K., & Chen, H.-C. (2013). Early morphological processing is sensitive to morphemic meanings: Evidence from processing ambiguous morphemes. *Journal of Memory and Language*, 68(3), 223-239. <https://doi.org/10.1016/j.jml.2012.11.003>
- Tsang, Y.-K., & Chen, H.-C. (2014). Activation of morphemic meanings in processing opaque words. *Psychonomic Bulletin & Review*, 21(5), 1281-1286. <https://doi.org/10.3758/s13423-014-0589-8>
- Tsang, Y.-K., Wong, A. W.-K., Huang, J., & Chen, H.-C. (2014). Morpho-orthographic and morpho-semantic processing in word recognition and production: Evidence from ambiguous morphemes. *Language, Cognition and Neuroscience*, 29(5), 513-528. <https://doi.org/10.1080/01690965.2013.790554>
- 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>
- Van Orden, G. C., & Kloos, H. (2005). The Question of Phonology and Reading. In *The science of reading: A handbook* (pp. 61-78). Blackwell Publishing. <https://doi.org/10.1002/9780470757642.ch4>
- Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition: A critical review of the evidence. *Neuropsychologia*, 53, 199-219. <https://doi.org/10.1016/j.neuropsychologia.2013.11.018>
- Wong, A. W.-K., Wu, Y., & Chew, H.-C. (2014). Limited role of phonology in reading Chinese two-character compounds: Evidence from an ERP study. *Neuroscience*, 256, 342-351. <https://doi.org/10.1016/j.neuroscience.2013.10.035>
- Xing, H., Shu, H., & Li, P. (2002). A self-organizing connectionist model of character acquisition in Chinese. In W.D. Gray & C.D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 950-955). Mahwah, NJ: Lawrence Erlbaum. <https://doi.org/10.4324/9781315782379-198>

- Xing, H., Shu, H., & Li, P. (2004). The acquisition of Chinese characters: Corpus analyses and connectionist simulations. *Journal of Cognitive Science*, 5(1), 1-49.
- 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>
- Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, 97(2), 259-268. <https://doi.org/10.1348/000712605X70066>
- Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, 16(3), 561-566. <https://doi.org/10.3758/pbr.16.3.561>
- Yang, J. F., Zevin, J. D., Shu, H., McCandliss, B. D., & Li, P. (2006). A "triangle model" of Chinese reading. In R. Sun (Eds.), *Proceedings of the Twenty Eighth Annual Conference of the Cognitive Science Society* (pp. 912-917). Erlbaum.
- Yang, J. F., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language*, 61(2), 238-257. <https://doi.org/10.1016/j.jml.2009.05.001>
- Yang, J. F., Shu, H., McCandliss, B. D., & Zevin, J. D. (2013). Orthographic influences on division of labor in learning to read Chinese and English: Insights from computational modeling. *Bilingualism: Language and Cognition*, 16(2), 354-366. <https://doi.org/10.1017/S1366728912000296>
- Yang, J. M., Staub, A., Li, N., Wang, S., & Rayner, K. (2012). Plausibility effects when reading one- and two-character words in Chinese: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(6), 1804-1817. <https://doi.org/10.1037/a0028478>
- Yao, P., Alkhamash, R., & Li, X. (2022). Plausibility and syntactic reanalysis in processing novel noun-noun combinations during Chinese reading: Evidence from native and non-native speakers. *Scientific Studies of Reading*, 26(5), 379-394. <https://doi.org/10.1080/10888438.2021.2020796>
- Yao, P., Staub, A., & Li, X. (2022). Predictability eliminates neighborhood effects during Chinese sentence reading. *Psychonomic Bulletin & Review*, 29(1), 291-300. <https://doi.org/10.3758/s13423-021-01966-1>
- Yu, L., Cutter, M. G., Yan, G., Bai, X., Fu, Y., Drieghe, D., & Livesedge, S. P. (2016). Word  $n + 2$  preview effects in three-character Chinese idioms and phrases. *Language, Cognition and Neuroscience*, 31(9), 1130-1149. <https://doi.org/10.1080/23273798.2016.1197954>
- Yu, L., Liu, Y., & Reichle, E. D. (2021). A corpus-based versus experimental

examination of word- and character-frequency effects in Chinese reading: Theoretical implications for models of reading. *Journal of Experimental Psychology: General*, 150(8), 1577–1596. <https://doi.org/10.1037/xge0001014>

Zhang, C. (2019). New perspectives on serialism and parallelism in oculomotor control during reading: The multi-constituent hypothesis. *Vision*, 3(4), 50. <https://doi.org/10.3390/vision3040050>

Zhang, H., Su, I.-F., Chen, F., Ng, M. L., Wang, L., & Yan, N. (2020). The time course of orthographic and semantic activation in Chinese character recognition: Evidence from an ERP study. *Language, Cognition and Neuroscience*, 35(3), 304–316. <https://doi.org/10.1080/23273798.2019.1652762>

Zhang, S., & Perfetti, C. A. (1993). The tongue-twister effect in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1082–1093. <https://doi.org/10.1037/0278-7393.19.5.1082>

Zhou, J., & Li, X. (2021). On the segmentation of Chinese incremental words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(8), 1353–1368. <https://doi.org/10.1037/xlm0000984>

Zhou, W., Shu, H., Miller, K., & Yan, M. (2018). Reliance on orthography and phonology in reading of Chinese: A developmental study. *Journal of Research in Reading*, 41(2), 370–391. <https://doi.org/10.1111/1467-9817.12111>

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning Memory and Cognition*, 26(5), 1245–1265. <https://doi.org/10.1037//0278-7393.26.5.1245>

Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology in reading Chinese compound words. *Language and Cognitive Processes*, 14(5–6), 525–565. <https://doi.org/10.1080/016909699386185>

Zhou, X., Ye, Z., Cheung, H., & Chen, H.-C. (2009). Processing the Chinese language: An introduction. *Language and Cognitive Processes*, 24(7–8), 965–979. <https://doi.org/10.1080/01690960903201281>

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv – Machine translation. Verify with original.*