LEARNING TO READ CHINESE AS A SECOND LANGUAGE: BUILDING LEXICAL REPRESENTATIONS IN THE INITIAL STAGES OF CHARACTER LEARNING

by

WEIZE SUN

(*aka* KRISTIAN SUEN)

A thesis submitted to

the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

January 2018

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

How do learners of Chinese as a second language acquire characters? When and how do initial L2 Chinese learners build the phonological, semantic and orthographic representations of the graphics? My research investigated the very earliest stages of learning Chinese and the effect of grouping, focusing on how we initially build these representations especially for a writing system which differs from our original one. My research demonstrated, for the first time, that in a time span of only seven days, a considerable number of behavioural changes could be observed to confirm that the learning of Chinese characters takes places. Moreover, representations of the learnt Chinese characters were maintained in long-term memory over the course of several days without further inputs. In this thesis, I provided clear evidence for that learning Chinese characters in semantic group without semantic radical and in phonological group by homophones contributes to better learning results compared to ungrouped learning. Meantime, grouping in semantic category with shared radicals or in rhyming sets inhibits the specification of representations. Reasons for such effects could be attributed to the coactivation of relevant lexical representations as well as to the degree of specification of newly formed lexical representations. A timeline of the initial L2 Chinese learning was also reconstructed based on empirical evidence. The consolidation effect of sleep was also addressed and discussed in this thesis.

謹以此文獻給

我摯愛之父母

孫洪義先生與王怡善女士!

This thesis is dedicated to

my dearest father Mr. Hongyi Sun

and

my most beloved mother Mrs. Yishan Wang

for your everlasting and unconditional love and support.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude towards my supervisors, Prof. Linda Wheeldon and Dr. Katrien Segaert, who provided enormous guidance and support throughout my research. Linda and Katrien are respectable and resourceful scholars with great kindness and patience. They provided me with considerable encouragement when I felt confused and puzzled during the whole process of the current research. Their encouragement and kindness will always enlighten me throughout my career.

Moreover, I am grateful to my fellow colleagues and friends: Beinan Zhou, Alper Kumcu, Johnny Lau, Haobo Chen, and Paraskevi Argyriou for their help, support and encouragement.

I shall express my gratitude towards Prof. William S. -Y. Wang at the Chinese University of Hong Kong, who guided, encouraged and helped me selflessly when I took the first step into this field. I would also like to thank Prof. Amanda Wood for her guidance and assistance in the design of experiment.

I am indebted to Mr. Kevin Yang, who provided substantial support and encouragement to me during my studying career at Birmingham. I also would like to thank Mr. Alexander Wang for his encouragement during my writing-up period.

Heartfelt appreciation would definitely be expressed to my parents for their generous encouragement and support to me throughout my life.

However, I alone am responsible for any mistakes and errors that remain and appear in this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

OVERVIEW

As a human artefact, the goal of writing is to record spoken languages and further, the expression of thoughts in speech and minds. Today, many people read and write in more than one language and sometimes in more than one writing systems. All writing systems, regardless of the language they represent, require a lexical representation comprising phonology, orthography and semantics. However, the relationship between the three components of lexical representations differs across different writing systems. An *alphabetic* writing system, such as English, is characterised by a mapping from the orthographic form to the phonological form, although the nature and transparency of that mapping can vary from language to language. By contrast, the mapping from orthographic form to the phonology is far less obvious in the *logographic* writing system, such as the Chinese language, which instead has elements that map directly to semantics. For example, the radical 犭 in the character 狗 (dog) means *animals*.

Regardless of writing system, the mental lexicon is not something we are born with. A baby starts to build these lexical representations in his/her *first language* (L1) and transfers them into the mental lexicon through a very rapid process. That process, which is also known as *fast mapping*, could happen after only one exposure to new words at as early as the year of two (Carey & Bartlett, 1978; Markson & Bloom, 1997; Spiegel & Halberda, 2011; Wilkinson & Mazzitelli, 2003). Similarly, during the acquisition of a *second language* (L2), we adults need to build the phonological representation and orthographic representation in the L2 again. However, in this case we may already have a lexical semantic representation of the word existing in our L1. In an adult reading system, the process of visual-word recognition is affected by both form and meaning (Coltheart, Curtis, Atkins, & Haller, 1993; Seidenberg & McClelland, 1989). Overlap between words shows that skilled readers have built a complex network of relationships between words they know in their mental lexicon (Collins & Quillian, 1969). Therefore, the process of L2 acquisition must differ from L1 acquisition.

The aim of my thesis is to investigate the early stages of how we learn another language which uses a different writing system from our native one. Particularly, I focus on the learning of Chinese characters by native English speakers. Since the 1970s when China returned to the global stage, the mysterious country has attracted growingly more attention from the Western world, and not surprisingly, more people have started to learn the Chinese language. Nowadays, research about learning Chinese as a Foreign Language (CFL) has become a sub area of Second Language Acquisition (SLA). To date, a number of studies have looked at the learning of L2 Chinese by adult native English speakers (Liu, Dunlap, Fiez, & Perfetti, 2007; Liu, Perfetti, & Wang, 2006; Liu, Wang, & Perfetti, 2007; Yum, Midgley, Holcomb, & Grainger, 2014). The learning process of L2 Chinese may take quite a long time, and most of these studies are therefore longitudinal in nature: lasting for weeks, months or even years. There is evidence from brain imaging studies that L2 Chinese learners need to recruit extra brain areas when reading Chinese characters which may not be commonly used by alphabetic language users (Nelson, Liu, Fiez, & Perfetti, 2005). Their main focus on brain imaging changes means the behavioural task was either an unmeasured by-product or just a way to present the stimuli. Behavioural evidence from those adaptions during the initial stage of L2 Chinese character learning are still largely unknown. That is exactly the question that I am going to explore in this thesis.

In the area of language learning and teaching, there are many studies investigating the question of how best to present language materials for learning. In particular, whether to teach words in groups based on either semantic or phonological relationships (Henning, 1973). The results varied dramatically: some suggested that learners could benefit from those grouping methods (e.g., Finkebeiner & Nicol, 2003), while the other indicated the opposite extreme – such learning could bring confusion and affect the learning outcome (e.g., P. Nation, 2000; Wilcox & Medina, 2013). How to present vocabulary for learners in a way that yields good learning is still a topic of debate. On the one hand, in the Chinese language, there are a large number of characters which are homophones (e.g., 盒 (box) [xɤ˧˥] and 河 (river) [xɤ˧˥]) and many more characters are rhyming (e.g., 河 (river) [xɤ˧˥] and 格 (lattice) [kɤ˧˥]). Learning phonologically related characters together could considerably reduce the workload because in Chinese, there is more limited variation in phonological representation compared to English (in Chinese, all characters are monosyllabic; while in English, each word could contain multiple syllables). On the other hand, the large number of *pictophonetic* characters[1] and the productiveness of *semantic radicals*[2] also enable semantic grouping in Chinese character learning. However, such grouping methods may not necessarily bring better learning outcomes as homophones and rhyming sets may also result in confusion among individual characters within the set. The issue is one of the tensions between forming unique representations for the identification of individual characters and building representations of the relationship between characters demonstrated among skilled readers.

[1] *Pictophonetic characters* are those contain one semantic radical which is pictographic and another component for the indication of the sound. For example, the character 河 (*river*) [xɤ˧˥] consists a semantic radical 氵 which means *water* and the other part 可[khɤ˨˩˦] indicates the pronunciation.

[2] *Semantic radicals* are part of the characters which provide the meaning of a category. For example, the radical 氵 (water) in characters like 河 (*river*), and 湖 (*lake*) means *water*.

In contrast to previous studies, my research investigates the very earliest stages of learning Chinese and the effect of grouping. My interest lies in how we initially build the representations especially for a writing system which differs from our original one. I will also investigate effects of grouping on the initial stage of L2 Chinese character learning. In this thesis, I report a series of behavioural studies which aimed to address the following research questions:

1. How do alphabetic readers build phonological and orthographic representations of Chinese characters during the very beginning stage of learning to read the Chinese language?
2. What is the effect of learning Chinese characters in semantical and/or phonological groups for novice L2 Chinese learners?
3. Which aspects of the new representations are affected by different grouping criteria during learning?

My research demonstrates, for the first time, that in a time span of only seven days, a considerable number of behavioural changes can be observed to confirm that the learning of Chinese characters takes places. Moreover, representations of the learnt Chinese characters are maintained in long-term memory over the course of several days without further inputs.

This thesis is organised as follows. In Chapter 1, I provide some background about skilled reading in both alphabetic and logographic languages. Differences between the two writing systems are briefly discussed, followed by principles of Chinese character structures. I also review studies of the reading process of alphabetic and logographic languages. In Chapter 2, the literature about learning to read in the first and the second language is reviewed, followed by a detailed discussion of the early learning of lexical representations. Effects of grouping according to phonological, semantic and orthographic categorisation are also discussed.

In Chapter 3, two experiments which investigate the effect of semantic and phonological grouping on the initial stage of Chinese character learning are presented. These experiments consisted of a learning and testing session across two consecutive days. On Day 1, L1 English readers without prior knowledge of East Asian languages learnt 20 Chinese characters in a self-paced E-Prime programme. They either learnt them in related sets (semantic related sets with shared semantic radicals in Experiment 1 and homophone sets in Experiment 2) or in unrelated sets. They were tested with a lexical judgement task (Talamas, Kroll, & Dufour, 1999). During the task, participants judged pairings of characters with Chinese sounds or English words. On Day 2, they took another testing session as I was interested in the effect of grouping following sleep consolidation (Gaskell & Dumay, 2003). I found that learning characters according to semantic radicals had inhibitory effect on results of learning in Experiment 1. In contrast, grouping by homophones had a facilitatory effect of learning in Experiment 2. The consolidation effect of sleep was observed: error rates dropped after sleep overall and sometimes reaction time dropped as well.

In Chapter 4 and 5, I expanded the experiment to run across seven days. The reason is that although I found evidence of learning after only one training session in the studies reported in Chapter 3, the level of learning was limited and did not reach a high level of accuracy. Moreover, I amended the learning materials to separate the effect of orthography from the effect of semantics. Specifically, in Chapter 4, I used pseudo-characters which the semantic radical was replaced with another semantic radical. In Chapters 4 and 5, I again used lexical judgement tasks to test the degree of learning of individual characters. Here I designed a categorisation task which aimed to test their performance over learnt items into their semantic and phonological categorisations. Based on the lexical judgement task I used in all experiments, I also included a primed version of that task to test their learning results of individual characters further with a set of primed conditions. In Chapter 4, I found a facilitatory effect of semantic grouping. I also, for the first time, gathered solid evidence that the pathway from L2 to the concept has been built only after three learning sessions. In Chapter 5, by contrast, I observed an inhibitory effect of phonological grouping on learning. I also found that phonological representations of Chinese characters have been well built among learners, especially for the identification of rhymes.

In the general discussion (Chapter 6), findings of the semantic and phonological grouping studies are compared and contrasted. Based on my findings, I briefly reconstruct the timeline during the initial stage of L2 Chinese character learning. I then discuss the consolidation effect of sleep. Real-world applications of my findings and future directions of similar research are also addressed.

CHAPTER 1

READING NETWORKS: A COMPARISON OF ALPHABETIC LANGUAGES AND LOGOGRAPHIC LANGUAGES

This chapter focuses on skilled reading processes across different writing systems. I will start with a brief introduction to the Chinese writing system and structures of Chinese characters. Followed by that, I will review existing literature on skilled reading process on written word identification in alphabetic and logographic scripts, ending with a brief summary of key points reviewed in this chapter.

Logographic Writing System and the Structure of Chinese Characters

The very first aspect of written languages we encounter is the script of the language. However, scripts do not necessarily indicate which writing system the language belongs to. Writing systems reflect *design principles*, not the visual forms of the writing (Perfetti et al., 2007). For example, Korean and Chinese languages may look similar because of the shape of their written unit. However, Korean uses an alphabetic writing system while the Chinese language is a logographic one. Here is an example to illustrate the logographic nature of Chinese (see Table 1.1 and Figure 1.1, p. 8). The Chinese character “门” means *door* or *gate*. The words in both languages are monosyllabic. The Korean counterpart in hangul “문” can be further divided into letters, which stand for individual phonemes while the Chinese character “门” is a morpheme and pictographic through the evolution of the character in different fonts chronologically.

Table 1.1.

Examples of Korean hangul and Chinese character.

	Chinese character					Korean hangul				
Form	门					문				
	丶	+	丨	+	𠃌	ㅁ	+	ㅜ	+	ㄴ
Pronunciation	[mən˥]					[mun]				
						[m]	+	[u]	+	[n]
Meaning	*door* or *gate*									

The Chinese language employs a written system which can be described as logographic, morpho-syllabic or morphophonological according to different scholars' claims (DeFrancis, 1989; Perfetti et al., 2007). Usually, a single character, which is represented by a spoken syllable, is a morpheme and also a word (DeFrancis, 1989). Unlike alphabetic languages where a graphic unit, i.e. a letter maps to a phoneme, the mapping from graphic forms in Chinese is from an individual graphic unit, which is a character in this case, onto a syllable and morpheme.

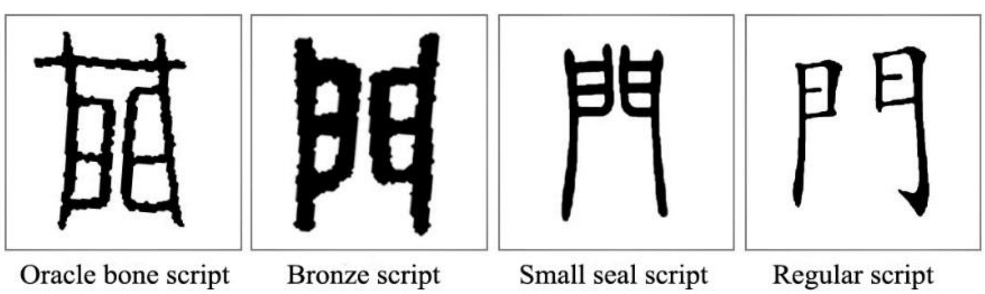


Figure 1.1. Evolution of Chinese character through different periods.

Table 1.2

The Six principles of composing Chinese characters

Category	Definition	Example	
xiangxing (Pictographic characters)	Characters formed by drawing the object while strokes imitate features of the object	日 *sun*	月 *moon*
zhishi (Simple ideographic characters)	Characters created by 1) an iconic form which could indicate an abstract idea or 2) putting an indicative sign to an existing pictographic character	上 *up* 木 + 一 = 本 *wood* + (sign) = *root*	下 *down* 木 + 一 = 末 *wood* + (sign) = *apex*
huiyi (Compound ideographic characters)	Two or more pictographic characters to form a new character which represents the meaning associated with their components	止 + 戈 = 武 *toe* + *dagger* = *military*	人 + 言 = 信 *person* + *speech* = *trust*
xingsheng (Pictophonetic characters)	Characters created by combining two components: a semantic component that suggests a meaning category and a phonetic component to (not necessarily accurately) indicate the sound of the whole character	氵 + 工 = 江 *water* + (*sound*) = *river*	氵 + 木 = 沐 *water* + (*sound*) = *wash*
zhuanzhu (Derivative cognate characters)	Characters which share the same etymological root or meaning, which could be orthographic, semantic or phonetic derived	考 *old*	老 *old*
jiajie (Phonetic loan characters)	Characters initially formed to represent one concept but borrowed to write a homophone which indicates a different meaning, while a new character was created for the original concept, usually containing the original character as a component	自 *self* (originally *nose*, new word for the original concept 鼻)	北 *north* (originally *back*, new word for the original concept 背)

(Xu, 121/1963, p. 314; Yin, 2006, p. 1)

Despite a long debate over these ways in which Chinese characters are constructed, the most influential theory *Liushu* (六书, six scripts), the *Six Principles of Composing Chinese Characters*, was proposed by antient linguist Xu Shen during the first century AD and revised by various scholars. The Six Principles include *xiangxing* (象形, pictographic), *zhishi* (指事, simple ideographic), *huiyi* (会意, compound ideographic), *xingsheng* (形声, pictophonetic), *zhuanzhu* (转注, derivative cognate), and *jiajie* (假借, phonetic loan) (Xu, 121/1963, p. 314; Yin, 2006, p. 1). Table 1.2 (p. 9) illustrates the Six Principles.

The Six Principles theory is by no means perfect. For example, the least productive principle, the derivative cognate, is neither fully clarified in Xu's original work nor given a well-accepted definition in the field of character studies. Many modifications are proposed since then. For example, the *Sanshu* (三书, three scripts) theory, which is proposed by Tang in 1934. According to his model, Three Principles of character formation are *xiangxing* (象形, pictographic), *xiangyi* (象意, ideographic), and *xingsheng* (形声, pictophonetic) (1934/1981, p. 401; 1949/2005, p. 61).

Although a single Chinese character is a morpheme, most Chinese characters can be divided into radicals. Radicals are components of character construction. A radical can consist of one or more strokes, which are the set of line patterns to form Chinese characters.

Characters which are formed of two or more radicals are called *compound characters*, and by contrast, characters only containing one radical are *simple characters*. At the radical level, one radical could serve as a *semantic indicator*, for example, the radical 氵 (*water*) in characters like

河 (river), and 湖 (lake), or a *phonetic component*, such as the radical 木 [mu˥˩] in the character 沐 [mu˥˩].

Nearly 70% of compound characters are pictophonetic (Han, 2012, p. 69). The large number of pictophonetic characters implies that the sharing of semantic and phonological radicals is quite common in written Chinese. Most Chinese dictionaries use both phonological clues and semantic clues as indices when retrieving a character, for instance, a character 沐 could be retrieved by its pronunciation [mu˥˩] (as in *pinyin*[3] mù) as well as by the semantic radical 氵 (water).

Reading Processes in Alphabetic and Logographic Scripts: Effects of Form and Meaning

Regardless of writing system, in order to understand a written language, one must obtain the skill of reading. Reading is the process of transforming information from print to meaning or speech. Successful reading relies on a "mental information processing system", which is capable of doing those transformations (Coltheart, 2005). However, to what extent the "processing system" differs among writing systems or whether there is a universal reading mechanism applicable to us all remains a matter of debate.

Numerous studies about reading processes, especially about reading aloud, contributed to the development of reading models of alphabetic scripts. There are two sets of models which are prominent within the field, the symbolic models and the connectionist models. The two sets of models share a basic assumption. They all assume that the pronunciation and meaning of a word

[3] Pinyin (拼音) is the official romanisation system of standard Chinese in China.

is generated through the interaction between the orthographic form and lexical knowledge. However, they differ upon the postulation of the lexicon: the former assumes its existence while the latter, does not (Coltheart et al., 1993; Rayner & Reichle, 2010; Seidenberg & McClelland, 1989).

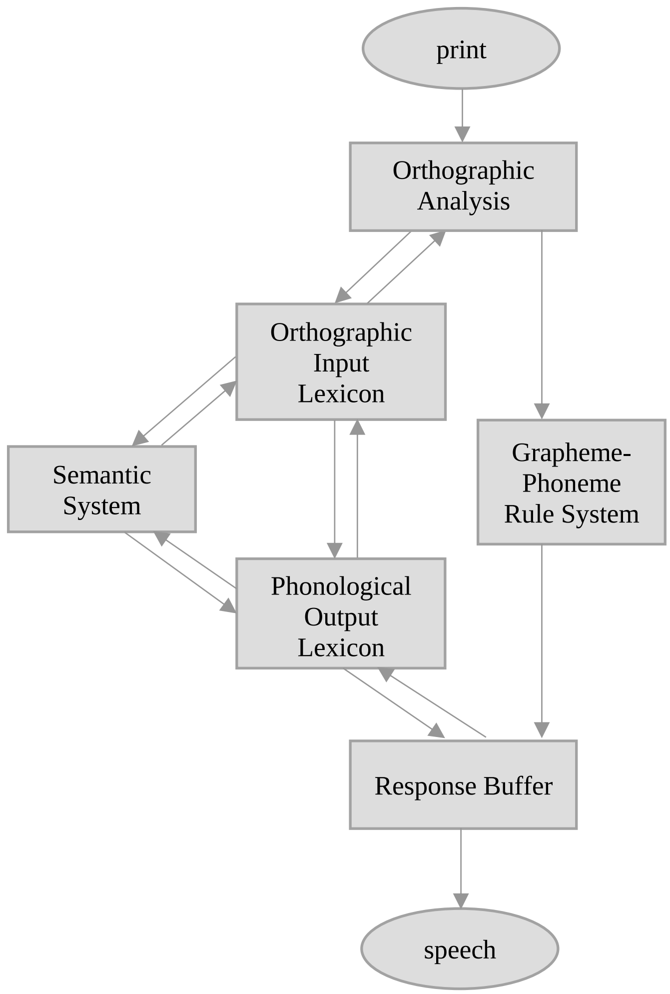


Figure 1.2. The dual-route model of visual word recognition and reading aloud. Adapted from “DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud,” by M. Coltheart, K. Rastle, C. Perry, R. Langdon, and J. Ziegler, 2001, *Psychological Review,108(1)*, p. 213. Copyright 2001 by the American Psychological Association.

The symbolic dual-route model (see Figure 1.2, p. 12) assumes phonology and orthography as independent processing units in the lexicon. It offers two parallelly operated routes for reading English letter strings: a *lexical* and a *nonlexical* route. The lexical route works by retrieving the phonological representation of a real word in a mental lexicon at the lexical level. That is realised through the mapping from graphemes to orthographic inputs, followed by the activation of phonological outputs from orthographic units. The nonlexical route uses linguistic rules about orthographic segments to phonological segments at the sublexical level, without referring to lexicon. When focusing on the process of visual words identification only, the model provides two pathways for the access into the lexicon: one directly from orthographic input to semantics, and the other from orthographic inputs to phonological outputs, and then into semantic system. One must use the former pathway to read phonological irregular words (*yacht*), while the latter must be used to read pseudo-words (*nufe*). In this view, phonology *mediates* the identification of printed words. (Coltheart, 2005; Coltheart et al., 1993; Rayner, Pollatsek, & Schotter, 2013)

The connectionist models, by contrast, only provide one single pathway to generate phonological representations from the orthographic input. Instead of assuming a lexicon, this model postulates that lexical information is stored and distributed in the "connections" and "correlations" between processing units containing orthographic, phonological, and semantic properties of words. When processing visual words, the stimulus interacts with the lexical information. Figure 1.3 (p. 14) clearly presents the role of phonology as a *product* or *output* of computation from orthography during word identification (Rayner & Reichle, 2010; Seidenberg & McClelland, 1989).

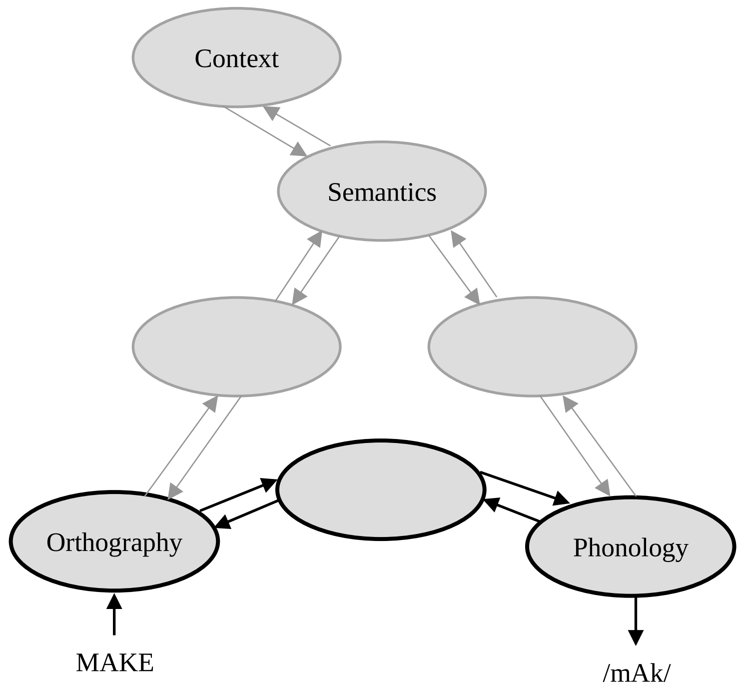


Figure 1.3. The connectionist model proposed by Seidenberg and McClelland (1989). Adapted from “A Distributed, Developmental Model of Word Recognition and Naming,” by M. Seidenberg, and J. McClelland, 1989, *Psychological Review*, *96(4)*, p.259. Copyright 1989 by the American Psychological Association.

Two models introduced above either consider the access to phonology as a mediator or simply deny the involvement of phonology during the process of word identification. It seems whether the process of word identification requires phonology remains a matter of debate. However, there are empirical evidence challenging the two models, especially in the following two aspects:

1. Are phonological codes involved during the identification of printed words? If so, when?

2. What is the role of phonology during the process of word identification? Is it merely a mediator, a product or something else?

Models mentioned above place phonology either before or after lexical access. However, this is not necessarily the case. As indicated by the *identification-with-phonology* hypothesis of *lexical constituency model* rather well (Perfetti et al., 2007; Perfetti, Liu, & Tan, 2005; Perfetti & Liu, 2005; Perfetti & Tan, 1998; Perfetti & Zhang, 1995), phonology is a *constituent* of word identification, which is neither a *pre-lexical mediator* nor a *post-lexical by-product*. That means words and morphemes could be phonologically specified. In this model, three interlinked constituents of word representations are required during the identification of words, which are phonology, orthography, and semantics. Each constituent has many processing units at its own level. Orthographic and semantic unit only represents the character and the meaning of the character in the mental lexicon respectively. However, a phonological unit could be an onset, a rime, or a tone. The model provides an explanation for the asynchrony between orthographic and phonological processing in Chinese word reading. The identification of a visual word in Chinese starts from the processing of sub-lexical radicals. In the computational lexical constituency model, the activation of phonology and semantics could only be triggered only after the threshold of orthographic processing is reached.

The answer to the first question lies in the evidence of the time-course of printed word processing, especially among alphabetic languages. During the reading of alphabetic languages, before triggering word-level phonology, the orthographic specification of all letter units does not necessarily need to be completed, which is called the *cascade processing style* (Coltheart et al., 1993; Seidenberg & McClelland, 1989). During the processing of phonological information, graphemes activate phonemes, and phonemes further combine into syllables and word. That process is known as *phonological assembly*. Therefore, there is a synchrony in time between the processing of orthographic form and the access to phonological information. Time window of such synchrony could be observed from the results of *masked priming paradigm*. In this paradigm, a prime word is presented swiftly (shorter than awareness), followed by the target word. Typically, the task is a lexical decision (Forster & Davis, 1984). The brief presentation of the prime will either facilitate or inhibit the processing of the target word. Using masked priming technique, Perfetti and Bell (1991) found evidence of the synchrony between orthographic and phonological processing in English among 135 native American English undergraduates. They found that orthographic and phonological facilitation both emerged at masked prime durations of 35 ms and continued to develop through 65 ms. Similar effects were discovered in other alphabetic languages, such as in French (Ferrand & Grainger, 1994). With a forward masked primed lexical decision task with 84 native French speakers, Ferrand and Grainger also found both the orthographic and the phonological facilitation effect became significant when the prime duration was around 29 ms and continued in the first 40 ms time window. Both studies showed that the orthographic effect slightly preceded the phonological effect. This phenomenon could also be seen in other studies, where orthographical code activation was found 20-30 ms before phonological code activation (Ferrand & Grainger, 1993; Ziegler, Ferrand, Jacobs, Rey, & Grainger, 2000). This swift lapse in time between orthographic and phonological processing was predicted by the *bi-modal interactive-activation model* (BIAM). In this model, when processing visual words, there was an extra but swift mapping from graphemes onto phonemes before the rapid activation of phonological representations. As shown in Figure 1.4 (p. 17), the slight delay in time between orthographic processing and phonological processing was the process from Stage

3 to Stage 4. This model was backed up with both behavioural evidence from masked priming paradigm as well as from *event-related potential* (ERP) experiments, which reflects the cascade during the processing of visual word identification (Diependaele, Ziegler, & Grainger, 2010; Grainger & Ferrand, 1994; Grainger & Holcomb, 2009).

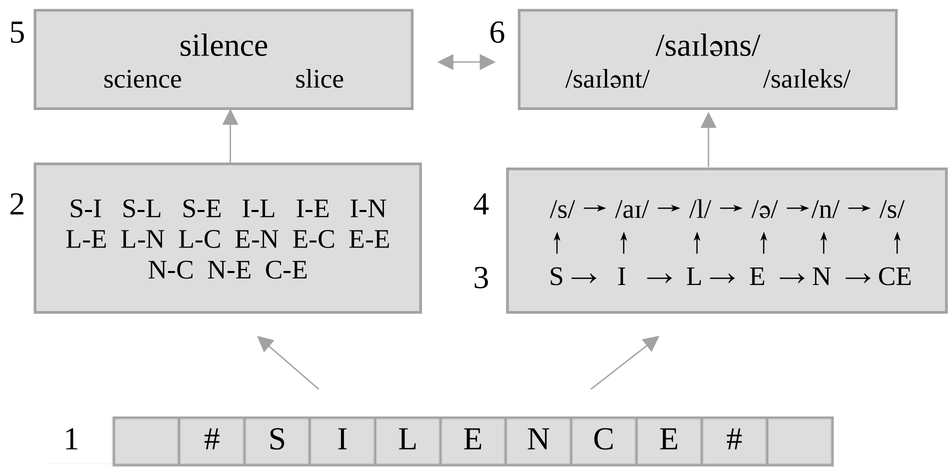


Figure 1.4. The orthographic and phonological pathways of the BIAM. Adapted from "Watching the Word Go By: On the Time-course of Component Processes in Visual Word Recognition," by J. Grainger and P. Holcomb, 2009, *Language and Linguistics Compass*, *3(1)*, p.131. Copyright 2009 by Blackwell Publishing Ltd. Adapted with permission.

It is worth pointing out that the lexical access via phonological codes may not be observed in some tasks such as *picture-word interference tasks* (PWI). Using this paradigm, Damian and Martin (1998) instructed 96 participants to name the picture in English and deliberately ignore the word distractor. The *stimulus-onset asynchrony* (SOA) was either 0 or 100 ms. The printed word was either a semantic related word to the picture name or a homophone of the picture name. At the SOA of 0 ms, the authors only found semantic interference but no phonological effects.

No effect was observed for the 100 ms SOA. According to their results, the role of phonology in visual word identification might be task dependent. Nevertheless, absence of phonological interference in a certain task does not rule out the possibility of phonological coding in reading.

However, for logographic writing system like Chinese, the segmental structure which is found in alphabetic writing systems does not exist. Therefore, phonological assembly cannot occur in Chinese character reading. The lack of a phonological assembly route does not indicate the lack of phonological activation during the reading of Chinese characters. However, for a written system like Chinese, only after a full orthographic specification of the character can the word-level phonology be activated, which is known as the *threshold processing style* (Perfetti et al., 2005; L. H. Tan, Hoosain, & Siok, 1996). Perfetti and Tan (1998) reported the asynchrony between the orthographic and phonological processing in a primed naming paradigm among native Mandarin speakers. There were four kinds of primes: graphically similar, homophonic, semantically related and unrelated. They found a graphic facilitation at 43 ms SOA. However, the facilitatory graphic effect switched to inhibition at 57 ms SOA, when the homophonic primes showed facilitation. Then at 85 ms SOA, the facilitation of semantic primes emerged with the homophonic facilitation continued. It is clear that in the time window between 43 ms and 57 ms SOA, there was a time point when the onset of phonological facilitation and the start of graphic inhabitation happened at the same time. Therefore, for a single Chinese character, the activation of phonological information happened after the full processing of the orthographic information.

Results of studies above lead to an important conclusion: phonology does take part in the process of printed word identification and it happens rapidly and automatically. For alphabetic languages, phonological processing and orthographic processing start almost simultaneously, while phonological effects lags behind orthographic effects slightly. However, for the Chinese language, character level phonology could only be activated once orthographic processing fully completes.

Concerning the second issue, phonological mediation during the access to meaning is predicted in the dual route model. The process only comes into play when the direct access is not possible. The *phonologically mediated priming paradigm* (PMP) is used to investigate the facilitatory effect of phonology. In a typical PMP experiment, the naming time of the target (e.g., *step*) is shorter for those primed by a homophone of the semantic related (e.g., *stare*) than by a control word (e.g., *stars*) (Lesch & Pollatsek, 1993). Due to the vast number of homophones in Chinese, process of such mediation becomes problematic. That is to say the constraint brought by pronunciation of a character is not strong enough for morpheme selection. This is reflected by results of Tan and Perfetti's study (1997). By controlling the homophone density of the prime, they found that phonological mediation is affected by the number of homophones that a certain character has. Namely, the identification of target was only facilitated by primes with few homophones. When there were many homophones of the prime, such facilitation disappeared.

As discussed above, the lexical constituency model provides a unified key to the two questions found in L1 reading process regardless of writing system. Moreover, there is evidence that the lexical constituency model could be extended to L2 learning contexts, for example, among L2 Chinese learners (Liu, Wang, et al., 2007). Details will be addressed in the following chapter (see p. 29).

Here it is clear that orthographic as well as phonological features of a printed word are involved during lexical access. When a word is presented to a skilled reader, the written word triggers several neighbours in the mental lexicon. Those representations compete for the best candidates for the identification of the word. One of these representations matches best and therefore inhibits the activation of other representations (Grainger & Jacobs, 1996; Jacobs & Grainger, 1994; H.-C. Wang et al., 2017). The process does not only limit to visual word recognition (Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989), but applies to auditory word recognition of alphabetic languages (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Swingley & Aslin, 2007) as well as in Chinese (Qu & Damian, 2017). Many of the evidence of such competition process come from inhibitory effect using masked priming technique for isolated single word reading as well as another paradigm for reading in context: *fast priming* (Sereno & Rayner, 1992).

```
   *---------*-----*----*-
a) Remember to bring your own| dvpl next time.  | Non-word preview
                             -*                 |
b) Remember to bring your own| look next time.  | Prime
                             -*                 |
c) Remember to bring your own| book next time.  | Target
                              *-----*----*      |
d) Remember to bring your own| book next time.  |
```

Figure 1.5. Schematic diagram of the sequence of events of the fast-priming paradigm, where "|" represents the invisible boundary and "*" represents fixations.

Figure 1.5 illustrates a typical fast-priming experiment trial. At the beginning, the target word in the sentence is previewed as a non-word (*dvpl*, see sentence a). Then, when the participant glances across the invisible boundary, the non-word is replaced by a prime word (b).

Swiftly after the presentation of the prime word, it is replaced by the target word (c). Participant continues reading the sentence until s/he finishes (d).

It was found by several fast-priming studies that when prime and target were orthographically overlapped (*beach-bench*), the processing time of the target became significantly shorter than those are orthographically unrelated (*beach-thing*) between the SOA of 29 and 32 ms. When target was primed by a homophone (*beach-beech*), facilitatory priming effect was found in the time window from 29 to 35 ms. When the target was primed by a semantically related word (*hate-love*), the semantic priming could only be observed at 32 ms SOA (H. W. Lee, Rayner, & Pollatsek, 1999). In another study (Y. A. Lee, Binder, Kim, Pollatsek, & Rayner, 1999), the SOA range of orthographic priming effect was further enlarged to 24 and 42 ms. Results of both studies showed that the activation of phonology started nearly at the same time of the semantic codes. However, there are two questions that shall be pointed out here: Firstly, the effect of phonological overlaps was sometimes confounded with orthographic overlaps, based on the stimuli provided in Y. A. Lee et al's study (1999); Secondly, it was not clear that whether the extent of facilitation would be affected by the position and/or extent of overlap.

Frisson, Bélanger, and Rayner (2014) carried out a study with 78 participants in order to address these two questions using fast priming and masked priming techniques. In their study, they designed four types of priming-target types. Four types were: high orthographic and high phonological (P+O+, *track-crack*), high orthographic and low phonological (P–O+, *bear-gear*), low orthographic and high phonological (P+O–, *fruit-chute*). At the same time, they also contrasted two types of orthographic overlaps: the beginning overlap (*swoop-swoon*) and the rhyme overlap (*track-crack*). Prime durations in the fast priming were 32 ms and 50 ms while for masked priming only at 50 ms. For fast priming tasks, they found facilitatory priming effects for both P+O+ and P–O+ conditions. There was no significant difference between the two P+O+ conditions (i.e., beginning overlap and rhyme overlap). They also found a facilitation priming effect only at the P+O– condition at 50 ms SOA. However, results of masked priming tasks showed inhibitory effects for end overlapping pairs in P–O+ and P+O+ conditions. This study provided further evidence for that orthographic overlap effect could be facilitatory in reading reflected through the fast-priming paradigm and inhibitory in lexical decision during masked priming experiments. Fast priming paradigm accounted for the immediate involvement of orthography and later the involvement of phonology during reading. This suggested a more crucial role of orthography than phonology in word identification process.

It is worth pointing out that phonological partial overlapping (e.g., rhymes) and phonological identical overlapping (homophones) are processed differently during reading. Evidence for that difference could be seen from fast-priming paradigm. The facilitatory priming effects could be observed in the time window from 29 to 35 ms SOA for homophones (H. W. Lee et al., 1999). However, such priming effect was only found at 50 ms SOA in Frisson et al.'s study (2014). Such discrepancy was also found in other paradigms such as four-field masking procedure. Using the four-field masking paradigm (mask-prime-mask-target), Lukatela and Turvey (1994, 1996) showed dramatic difference between rhyme priming and homophone priming. They found a facilitatory priming effect of homophones at 250 ms SOA while a consistent inhibitory rhyme priming effect at 36 and 70 ms.

Nevertheless, results of all the studies discussed above show that skilled reading in the first language involves the multiple activation of lexical representations that share orthographic, phonological and semantic features with the target word.

In this chapter, I firstly introduced the logographic writing system and addressed the structure of Chinese characters. I reviewed major models for skilled L1 reading of alphabetic and logographic languages and discussed the role of phonology in reading those languages. Through the review of the overlap effects in L1 reading, it is clear that successful reading in L1 necessarily needs multiple activation at phonological, orthographic as well as semantic levels. In the next chapter, I would like to review and discuss this process in a L2 learning context and address my own research questions through that.

CHAPTER 2

LEARNING CHINESE AS A SECOND LANGUAGE

In Chapter 1, I have introduced alphabetic and logographic writing systems and the process of skilled reading in those systems. It is clear that the reading process of Chinese scripts involves processing in threshold style compared to reading alphabetic scripts, which is a cascade processing style (Perfetti et al., 2005). Character level phonology could only be activated once the orthographic processing is fully completed. In this chapter I turn to the issue of learning to read Chinese as a second language. I will briefly introduce a developmental learning theory of L2 vocabulary. Then, I will review the literature on the longitudinal studies of L2 Chinese character learning with an emphasis on the very beginning stage of the learning process. I will also introduce a novel perspective to investigate the initial stage of L2 Chinese character learning: grouping. Grouping words according to different criteria could make it possible to observe the relationship among the newly learnt L2 vocabulary and their representations. Skilled reading in the first language involves the multiple activations of lexical representations that share orthographic, phonological and semantic features with the target word. My interest is in how such overlap affects the learning of L2 vocabulary.

Vocabulary plays an essential role in second language acquisition since it enables comprehension and communication within the new language. The learning of vocabulary of in second language as an adult differs from the learning of L1 vocabularies. Second language learning in this situation occurs with a fully developed conceptual system and, of course, with a first language already in place. An important theory of second language acquisition is the *Revised*

Hierarchical Model (RHM), which was proposed by Kroll and Stewart (1994) (as shown in the figure below).

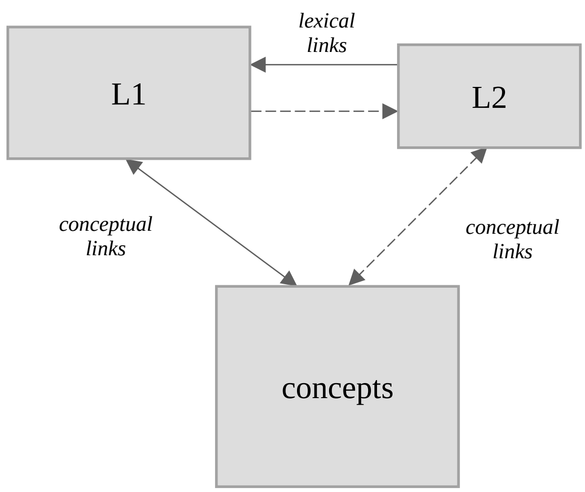


Figure 2.1. The Revised Hierarchical model. Adapted from “Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations,” by J. Kroll, and E. Stewart, 1994, *Journal of Memory and Language*, *33*, p. 158. Copyright 1994 by Academic Press, Inc. Adapted with permission.

According to RHM, during the initial stage of learning a second language, learners may need to rely on their first language to mediate the access to meaning stored in concepts of the newly learnt L2 lexicon. However, when learners reach a certain level of proficiency in their L2, the route via L1 lexicon will be less necessary and they will gain direct access to semantics via the L2 lexicon (Kroll & Hermans, 2011; Kroll & Stewart, 1994; Kroll, Van Heel, Tokowicz, & Green, 2010). The model addresses developmental aspects bilingualism as it predicts that increasing levels of L2 proficiency result in changes to its organisation in long-term memory (Duyck & Brysbaert, 2010).

The RHM was initially proposed to explain the phenomenon that late bilinguals take longer time for forward transition (L1 to L2) than backward translating (L2 to L1). Kroll and Curley (1988) found that fluent L2 speakers spent significantly longer time to translate semantically related lists of words into L2 than they did for those in random groups. They also found for picture naming tasks in L1, all participants were slower for semantically related lists than semantically unrelated lists (Kroll & Curley, 1988). In Kroll and Stewart’s study (1994), they demonstrated that the category interference existed not only among native speakers during picture naming tasks, but could also be found in translation tasks among advanced bilinguals. They used picture naming tasks as well as forward (L1-L2) and backward (L2-L1) translation tasks. They found that both English monolinguals and Dutch-English bilinguals were all faster for the random lists than the semantically related lists in picture naming and bilingual translation respectively. However, there was no category interference for picture naming tasks with word naming in-between. Besides, the mean RT of translation from Dutch to English (L1-L2) was 119 ms longer than the RT of translation from English to Dutch (L2-L1). Their findings generally confirmed their prediction that backward (L2-L1) translation should be faster than forward (L1-L2) translation as the former only relies on lexical links while the latter, requires the access to concepts. During the process of picture naming and translation, readers used a conceptual representation of the item in the retrieval of a word in the target language. Intense conceptual activity could trigger the activation of several lexical representations. The interference could only happen during the selection process of the best match from those representations. Therefore, category interference found in this study could be seen as indicators of semantic processing.

However, as some items used in this study were low in frequency, one may simply argue that such results might be attributed to the consequence of frequency effect. Kroll and Tokowicz (2005) claimed that instead of invalidating the RHM, different language skill effects could be reflected through the range of difficulty in item processing for a given skilled bilingual in Kroll and Stewart's study in 1994. In the framework of RHM, semantics could be accessed from L2, while the conceptual link from L2 was weaker than that from L1.

The semantic interference reported in RHM could be attributed to the competition process of co-activated semantic representations as discussed in Chapter 1. Damian, Vigliocco, and Levelt (2001) replicated the Kroll and Stewart's (1994) findings among native German speakers. They found that naming pictures in semantic related sets are slower than those in random sets. In the other experiment, they required participants to name words in the same/different semantic sets or produce the word with the determiner. They predicted that no competition could be found in the reading aloud task as the task was at form level. However, an inhibitory effect of semantic category should be found in the "determiner + noun" phrase production task. In German, the determiner is gender specific. Retrieval of such grammatical gender information requires entry to the mental lexicon. Results showed that participants produced the phrase significantly slower for items in the semantic category than items in different semantic categories. No competitive process was observed in the reading aloud task. Their findings supported that semantic interference reflects the competition process during the lexical retrieval.

Learning to Read Chinese as a Second Language

Previous studies suggest that the acquisition of Chinese characters among second language (L2) learners, especially among native alphabet language speakers, is likely to be different compared to their learning of a new alphabet (de Groot & Keijzer, 2000; Liu, Dunlap, et al., 2007; Liu et al., 2006; Liu, Wang, et al., 2007; Perfetti et al., 2007; Stein et al., 2006; M. Wang et al., 2003). The Chinese language is phonologically and visually distinct from alphabet languages. When alphabetic speakers learn a new alphabetic language, the effect of *cognate* takes place. Cognates are words in different languages that share meaning and have similar orthographic or phonological representations, for example *telephone* in English and *teléfono* in Spanish. This effect can only be found when the first language (L1) and L2 share an alphabet. Cognates are "easier to learn and less susceptible to forgetting" (de Groot & Keijzer, 2000). However, not all languages share an alphabet. Much less is known about second language acquisition when scripts are not shared.

Wang, Perfetti and Liu (2003) carried out a longitudinal study of the acquisition of Chinese characters among native English speakers. They demonstrated that after a one-term course of Chinese language, learners were sensitive to the visual-orthographic structures of Chinese characters and had already gained considerable linguistic knowledge about Chinese characters. Learners were able to respond differently to non-characters with illegal radical forms, those with legal radical forms but in illegal position, and those with legal radical forms in legal positions. Using event-related potential (ERP) paradigms, Liu, Perfetti, and Wang's study (2006) found that visual analysis of English was significant easier than that of Chinese characters at the first term; but at the second term, no significant difference was found between them. It means that the visual analysis of stoke and radical processing can be learnt very fast. After two terms of learning of Chinese characters, learners developed a similar pattern of reading to native Chinese readers.

Liu, Wang et al. (2007) provided evidence for the application of lexical constituency model to L2 Chinese learners. In their study, they investigated the learning of orthography and semantics in a new written system. Participants were asked to perform a primed word naming task at the end of the first term as well as in the second term. In this task, a prime character was presented for 500 ms before a target character that participants named. Results showed that at the end of first term, orthographic priming facilitated word naming while homophonic and semantic priming did not. Interestingly, by the end of the second term, the facilitation of naming sped by orthographic priming vanished and was replaced by semantic and phonological priming. Although the SOA stayed the same, these results showed a similar threshold pattern to Perfetti and Tan's study (1998) discussed in Chapter 1 (p. 18). By the end of the second term, there was no facilitation of orthography. Therefore, the author inferred that the orthographic priming effect at 500 ms in the first term reflected the pre-threshold orthographic activation. By the end of the second term, the activation of semantics and phonology suggested the lowering of the threshold of orthographic processing, which enabled those post-threshold activations to take place within 500 ms. Results of this study showed that orthographic threshold could be lower by learning through practice with specific characters. That provided evidence for the acquisition of the language specific threshold processing style among alphabetic readers.

The Building of L2 Lexical Representations

Changes in second language lexical processing can not only be observed longitudinally (Ojima, Nakamura, Matsuba-Kurita, Hoshino, & Hagiwara, 2011; Stein et al., 2006), but within the initial stage of learning (McLaughlin, Osterhout, & Kim, 2004). By comparing the ERP results of English German learners before and after a five-month immersion, Stein et al. (2006)

collected evidence for an increasingly fast processing of L2 vocabulary in the brain (i.e. the latency decreased in N400 and P600). Their results showed that second language word processing was faster and involved shorter frontal activation when the proficiency of second language was higher. A similar pattern was revealed by Ojima et al. (2011). They carried out a three-year longitudinal study among native Japanese school children, who learnt English as a second language. Participants received an ERP session each year. They observed that, with the increase of the L2 proficiency, typical L1 acquisition ERP components can be observed, such as a broad negativity, N400, and late positive component and moreover, in the same order as the developments in the first language. Moreover, the indicator of a faster processing of L2 words could be seen even after 63 hr when native English speakers learn French words (McLaughlin et al., 2004). Therefore, in my thesis, I will investigate the very early changes happened during the initial stage of second language word learning.

Evidence from Initial Learning of L2 Chinese

As mentioned before, most of the existing literature on learning of Chinese as a second language is based on classroom instruction and involves one or two semesters of learning after which some behavioural or brain imaging changes are observed. Obviously, early changes related to the learning of Chinese characters are not observable due to the long time-course of these studies mentioned above.

There are only a few studies focusing on the very first stage of L2 Chinese character learning. Liu, Dunlap et al. (2007) carried out a laboratory designed to investigate the rapid learning of Chinese characters. They recruited native English speakers without knowledge of the Chinese language. The training period consisted of three 2-hr learning sessions across three days.

During each session, they learnt 20 Chinese characters in a computer-based learning programme. Participants saw each character on the screen together a video clip played at the same time. The video clip was either a Chinese speaker producing the Chinese pronunciation of the character or an English speaker reading the translation in English. The training was self-monitored and logged. The pronunciation of characters was taught with or without their meaning in English. After the training period, they participated in a character naming task and a semantic category judgement task to test their learning results. Results of the behavioural test were quite intriguing: 23 of the 29 participants tested reached accuracy levels of at least 92% for the naming task and 92% for the category judgment task. Results for both tasks were high, considering the short time of their training period. Results of the brain imaging study suggested that their reading network had been modified after only such a relatively short learning period. Participants presented a bilateral pattern which is not commonly seen in alphabetic language readers but is usual among native Chinese readers. It is therefore clear that participants adapted to the orthography of Chinese language very quickly. The quick adaption could be observed by brain imaging methods as well as through behavioural tasks.

An ERP study of L2 Chinese character learning also provides an insight into these changes that occur during the initial stage of learning (Yum et al., 2014). Participants of this study were native English speakers who were not fluent in any other languages. There were 10 sessions of vocabulary learning, during which they learnt 2,000 Chinese words, which were either mono-character words or bi-character words. ERPs were measured in four of the 10 training sessions and data from behavioural tests were collected in all 10 sessions. The training sessions contained three tasks which comprised the learning input of the experiment. The three training tasks were a go/no-go N-back word task, a word-word association training and a translation recognition task. During the go/no-go N-back word task, participants were asked to decide whether the stimulus on the screen was shown before. Then participants then completed the word-word association tasks, where they were required to see the character with the English translation together on the screen without responding to the word. Then, in the translation recognition task, participants were asked to judge whether the second word in English presented on the screen was the translation of the first word in Chinese. Feedback of this task was given to the participant after they pressed the button. Behavioural assessments used in this study were a L2 to L1 backward translation task and a go/no-go semantic categorisation task during the ERP session. In the backward translation task, participants were asked to speak out the English translation of the Chinese word shown on the screen. The semantic categorisation task required the participants to judge the semantic category of the Chinese word by pressing the button. Only 15% of the trials needed a response, while remaining trials only asked participants to read the word. Then, participants had another run where all the L2 Chinese words were replaced by English translations while tasks remained the same. Participants were divided into fast and slow groups according to their performance on behavioural tests. Behavioural results of the two assessed tasks clearly presented a developmental change along 10 sessions regardless of learner group: as more learning input was given, the accuracy of both the tasks increased. However, there were significant differences between two groups. Fast learners reached a higher level of accuracy for both tasks at approximately 90%, while slow learners only reached approximately 50% accuracy. For both groups, improvements due to learning input became reduced from Day 7. In the ERP components of the study, they found that fast learners showed patterns for script-specific orthographic processing as well as indicators of higher L2 proficiency, which was not found among slow learners. Those patterns included a left-lateralised increase in N170 and an increased

N400. According to the result above, during the very early learning of Chinese characters, only fast learners could find the orthographic patterns among characters, which led to a better performance. More importantly, this study provided further evidence for the rapid learning of L2 Chinese vocabulary among novice learners. The time window to observe such changes could be within seven days.

Both studies described above suggest that important changes in processing occur during the initial stages of character learning and that the nature of representations built in these early stages of learning have consequences for later performance. However, they both lack behavioural evidence of adaptions of reading networks and the construction of the timeline of the initial stage of L2 Chinese character learning. Those are exactly the gaps which my research aims to fill.

The Sleep Consolidation Effect in Vocabulary Learning

The initial stage of learning examined in studies reviewed in the previous section necessarily occurred over the course of a few days. Results of relevant studies that there was a sleep consolidation effect in L1 novel vocabulary learning (Dumay & Gaskell, 2012, 2007; Gaskell & Dumay, 2003; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010, etc). Dumay and Gaskell (2007) tested the effect of sleep on the mental representation of spoken English words among 64 native English speakers. They firstly heard 24 new words and then 48 disyllabic triplets. Each triplet contained one English word (e.g., *shadow*), a *novel* word to learn (e.g., *shadowks*), and a foil for recognition tasks (e.g., *shadowkt*). After the learning session, they were tested immediately and had the second testing session 12 hr later and the third testing session after another 12 hr. For all participants, the interval between learning and testing sessions were 12 hr. Some participants had an overnight sleep while the others were tested on the same day without sleep. They observed that the newly learnt novel words could slow down latency for the identification of phonologically related words after a certain time. Namely, the competition effects of learnt items were found in the sleeping group after 12 hr with a sleep whereas the non-sleep group only showed the same pattern after 24 hr with a sleep (Dumay & Gaskell, 2007). These findings were further confirmed by their study in 2012: there was no competition right after exposure. A significant inhibitory competition emerged 24 hr later and after seven days (Dumay & Gaskell, 2012). Recently, it was reported such consolation effect also applied to the learning of vocabulary in second languages (Mirkovic & Gaskell, 2016). Using the nap paradigm, they tested the learning of vocabulary of an artificial language among native monolingual English speakers. These words were in an artificial langue rather than pseudo-words in English because it contained grammatical suffixes and determiners (e.g., *tib bisesh,* where *tib* is the determiner, *bis-* is the stem, and *-esh*, the suffix). They were trained for around 35 min with 16 novel words in the language, all of which were pronounceable in English. Before they were tested, some participants had a 90-minute nap and a 30-minute break while the others watched 105-minute DVD programme and a 15-minute break. During the testing period, one of the tests was the forward translation recognition test (L1-L2). It required participants to decide whether the English word they heard at first was the correct meaning of the L2 sound played after that. Participants from the nap group reached significantly higher in accuracy (95%) than those from the non-nap group (84%). These results suggested that the sleep consolidation effect of newly learnt lexical information was not language-specific and could be applicable to languages other than English.

It is worth noting that in this thesis, sleep is neither controlled nor manipulated as they did in studies above. The effect of sleep consolidation is taken it into account and being modelled and tested through these experiments in this thesis.

Grouping: A Novel Perspective to Observe the Initial Acquisition of Vocabularies

Through the discussion in previous sections, it is clear that the initial stage of L2 vocabulary learning provides an opportunity to observe the process of the construction of lexical representations in the mental lexicon. In fact, Henning (1973) pointed out that L2 language learners would code vocabulary acoustically as well as semantically. In his study, he tested participants proficiency in English with a cloze test among native English speakers and native Persian readers of L2 English. Participants completed a 60-item word recognition test. Participants were firstly exposed to five passages. Each passage lasted for 30 s. During the word recognition test, the correct words narrated in the passage were presented with three other words, which could be three semantically related, three phonological related or three unrelated distractors. The correlation results showed that readers of low proficiency made more acoustic errors while high level readers made more semantic errors. This suggested that: low-proficiency learners would prefer phonological clues, whereas semantics is more common among high-proficiency learners.

Suárez-Coalla and Cuetos's study (2017) showed that the pre-existing phonological and semantic information contributed to a fast and robust formation of orthographic representations. According to previous research, a decrease in *length effect* could be seen as the indicator of established orthographic representations (Kwok & Ellis, 2015; Weekes, 1997). Here the length effect refers to the phenomenon that people could spend more time on long and unfamiliar words than short ones when reading alphabetic languages. In their study, 71 adult L1 Spanish speakers learnt 10 unfamiliar concrete Spanish nouns: five long words (7-8 letters, e.g., *dolobre*) and five short ones (4-5 letters, e.g., *duba*). Participants were divided into three groups, each receiving a different training. During the training phase, participants were asked to pay attention to the training and orally repeat the word. Participants of Group 1 saw the picture while hearing the sound (Condition 1: S+P+). The second group only heard the sound (Condition 2: S–P+), while the third without training at all (Condition 3: S–P–). Each word was presented six times. After the training, they were tested with a reading aloud task of six blocks. All the 10 stimuli were tested once in each block. Naming latencies and error rates were recorded during the testing phase. A similar reading aloud test was carried out one month later as a follow-up session. Participants of the S+P+ condition also performed a spoken word-picture matching task. Results showed that the length effect first disappeared among participants of the S+P+ condition in Block 2, then among S–P+ participants in Block 3, and finally in Block 5 among S–P– participants. The fast decrease in length effect among participants who received S+P+ training suggested a crucial role of previous phonological and semantic clues when reading new words. Results of the follow-up reading aloud test showed that orthographic representations still existed in the lexicon. The 52% of correctness of the word-picture matching task only indicated guessing, while participants did significantly better for the long words (63%) compared to the short ones (41%). The authors attributed that to the length of words as they were more discriminable than short ones. Overall this study presented a rather exiting picture: previous phonological and/or semantic information, i.e., what we saw, what we heard and what we knew, could actually helped us when encountering new words during reading. However, compared to participants without training at all, to what extent could the better performance in the reading aloud test among participants with phonological inputs be attributed to the facilitation of phonological background? Or it could be a result of the orthographic transparency of that specific language, Spanish in this case? In other words, for a language like Spanish, the orthographic consistency could probably lead to a confound in a study like this: orthographic representations could have been formed during the training. This suggests a different paradigm is needed to evaluate the facilitation of phonological and semantic influence on vocabulary learning to eliminate such confound.

Since phonological and semantic knowledge could facilitate word learning, one could easily ask: is that the best way and most effective way to present those learning materials? Will that make a difference if theses learning materials were grouped in semantic or phonological sets when learning? As one of the central questions of my thesis, the following sections will focus on existing literature about the effect of grouping words in semantic and phonological sets respectively.

Semantic Grouping

Semantic grouping, according to Tinkham's definition (1997), is a way to present words based on their shared semantic and syntactic similarities. For example, *pig*, *cat*, *monkey*, *dog*, *bird* are all nouns and all belong to the semantic category of *animal*.

The facilitatory effect of semantic grouping in novel vocabulary learning is assumed by many researchers (Gairns & Redman, 1986; Seal, 1991). They claimed that semantic grouping naturally led to a precise distinction among items within the semantic sets and an in-depth learning of the meaning as each item learnt reinforced the learning of other items. In fact, there were many L2 English textbooks practicing this teaching approach (Finkbeiner & Nicol, 2003). Many studies tested the effectiveness of grouping in the acquisition of L2 vocabulary various languages (e.g., artificial languages: Finkbeiner & Nicol, 2003; Tinkham, 1993, 1997; L2 Japanese: Waring, 1997; L2 English: Henning, 1973, etc).

In a classroom environment, Hashemi and Gowdasiaei (2005) tested the effect of learning L2 English vocabulary in semantically related and semantically unrelated sets among 60 intermediate L2 English learners in Iran. Using a between-group design, half of the participants learnt 130 words and expressions in semantically related sets while the other half learnt the same material in semantically unrelated sets. Results of post instruction vocabulary knowledge evaluation showed that learning vocabulary in semantic related groups led to significantly higher levels of knowledge for both lower and upper level students, while upper level students benefited from semantic grouping to a greater extent than lower level students.

For East Asian languages, a study of 119 Japanese readers of L2 English also revealed the effectiveness of semantic grouping (Hoshino, 2010). In this study, the author tested the effect of four learning strategies among Japanese *English as a Second Language* (ESL) learners, namely grouping by synonyms (e.g., *fabric* and *textile*), antonyms (e.g., *dirty* and *clean*), semantic category (e.g., *dish* and *bowl*) and thematic relationship (e.g., *beach* and *sunny*). Results demonstrated that for all participants, regardless of their learning style, the learning of semantically grouped lists led to significantly better results compared to the learning of synonym, antonym, and thematic lists.

However, other researchers challenge the effectiveness of semantic grouping as a learning aid, as they found that semantically related word lists caused difficulties during initial L2 vocabulary learning (Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Higa, 1963; P. Nation, 2000; Tinkham, 1993, 1997; Waring, 1997). Typically, one of Tinkham's studies (1993) used artificial languages to mimic the process of initial L2 vocabulary learning. The author tested how easily participants learnt English-artificial word pairs in semantic related and unrelated sets with a trials-to-criterion paradigm. For each condition of the test (semantically related or unrelated), only after the participant correctly spoke the three artificial words in that condition could the test proceed to the next step. He found that English readers of L2 artificial language needed significantly more trials to learn English-artificial word pairs of the same semantic category than those that were semantically unrelated. These results were consistent across two studies and in the other study, the inhibitory effect of semantic grouping was found in both written and oral modalities. A negative effect of semantic grouping on novice L2 vocabulary learning was also observed in speakers of other languages. Waring (1997) replicated Tinkham's findings in an L1 Japanese population with Japanese-Artificial word pairs.

From the discussion above, it is clear that both positive and negative effects of semantic grouping on vocabulary learning have been observed. However, it is worth noting that a facilitatory effect was generally found among learners with a certain level of L2 proficiency while evidence of a negative effect was observed with first-shot learners.

Phonological Grouping

There are some studies of the learning of vocabulary based on sound similarities. However, the nature of phonological grouping or clustering tested varies from homophones, to shared onsets, or rimes. Some scholars believe that at the early stage of L2 lexical acquisition, the learning of repetitive phonological structure like homophones should be avoided (Henning, 1973; I. S. P. Nation, 1982). By comparing the number of acoustic errors in a vocabulary recognition test across learners of different L2 English proficiency levels, Henning's study (1973) showed that low-proficiency learners relied on acoustic similarity rather than semantic relationships during learning. However, the association between the phonological related newly learnt words and words they had learnt before was more likely to result in interference than facilitation.

For L1 vocabulary learning, Szmalec, Duyck, Vandierendonck, Mata, and Page's study (2009) showed that repeated phonological features among words (e.g., the same syllables) led to better recall for lexical items using Hebb repetition paradigm. The *Hebb repetition effect* is the phenomenon that participants recall repeated sequence significantly better than they do for non-repeated sequences (Hebb, 1949, 1961). This effect was argued to be a laboratory analogue of learning phonological word forms (Page & Norris, 2009). They first tested the Hebb repetition effect with nonsense consonant-vowel syllables (e.g., *fi*, *wa*, *ri*, *mu*, …). Those sequences were visually presented to 42 native Dutch participants for immediate serial recall. They found that repeated sequences of syllables were significantly better recalled than non-repeated sequences. Then they conducted an auditory lexical decision task. The same group of participants were asked to judge whether these sounds they heard were *words* or not. Some of the sounds were nonwords constructed with repeated syllable sequences used in the first experiment while the others were nonwords made from non-repeated filler sequences. Interestingly, they found that participants were significantly slower to reject nonwords built from repeated syllables. This suggested that items learnt during Hebb repetition were able to access and be stored in the mental lexicon, which could be seen as a laboratory analogue of learning new words.

In terms of the acquisition of L2 vocabulary, Wilcox and Medina's study (2013) investigated effects of semantic and phonological grouping on initial learning of L2 Spanish vocabulary among native American English speakers. Participants were instructed with 20 new Spanish words on E-Prime. The 20 words fell into four categories according to whether they were semantically related or not and whether they shared the beginning phoneme or not. Participants learnt the 20 words three times. They were tested immediately after they learnt the five words in one category until they finished learning all the 20 words. Two weeks later, participants came back and attended another test for these words they had learnt. They found that semantically grouped words were the most difficult to learn compared to other categories, which was supported by results of immediate test and the test two weeks later. These findings showed that semantic grouping caused difficulties and therefore led to worse results compared to presenting words randomly. Besides, presenting words in phonological groups contributed to improved performance both in short term and in long term based on results of their experiment.

As discussed in this chapter, both revised hierarchal model (RHM) and lexical constituency model could explain L2 vocabulary learning process. The very initial stage of L2 word learning is a critical time window to observe the formation of lexical representations. Studies like Suárez-Coalla and Cuetos (2017) generally confirm that phonological and semantic information contributes to the formation of orthographic representation when learning new words. Although the effect is largely under debate, phonological and semantic grouping could help to observe the relationship among the newly learnt L2 vocabulary and their representations. In the following chapters, I will design a series of experiments to further investigate the initial stages of learning L2 Chinese characters and explore the effect of grouping. These experiments aim to answer the questions about the building phonological and orthographic representations and the links between existing semantics to those representations.

CHAPTER 3

SELF-PACED LEARNING: EFFECTS OF SEMANTIC AND PHONOLOGICAL GROUPING

Introduction

The aim of my research is to investigate the process of building lexical representation in the initial stage of L2 Chinese learning by native alphabetic readers (English speakers). As reviewed in Chapter 2, it is clear that while the neural network involved in the reading of Chinese characters shares some features with the English one, distinctive characteristics exist. Learners of Chinese as a second language will accommodate their own neural networks to the demand of Chinese language. However, the speed at which this accommodation occurs and critical behavioural landmarks of those changes remain largely unknown.

According to Yum et al.’s study (2014), learning efficiency may reflect different learning strategies affecting the acquisition of phonological, orthographic and semantic links. That means fast learners and slow learners might use different learning strategies when acquiring Chinese characters. The question then arises: if the spatial pattern, for example, the shared semantic radical, or shared phonological pattern (such as homophones), is explicitly presented to learners, does it contribute to a better learning outcome than those who learn characters without highlighting these patterns?

The rationale for such grouping ideas come from the overlap effect in L1 skilled reading. Studies about skilled L1 reading suggest that successful reading involves coactivations of orthographic, semantic as well as phonological neighbours that overlap with the target word. That process was revealed by results from the facilitatory fast-priming effects and the inhibitory masked priming effects (Frisson et al., 2014; H. W. Lee et al., 1999; Y. A. Lee et al., 1999). As reviewed in Chapter 1, the prime-target relationships that could evoke the overlap effects are homophones and orthographic neighbours. Based on the discussion in Chapter 2, phonological neighbours such as words sharing segments could also evoke similar effects. Therefore, in a L2 learning context, we would like to address whether the explicit instruction of such overlaps among characters could result in significant difference in learning outcomes.

In this chapter, I report two experiments in which we manipulated the categorical grouping of characters during training to investigate the learning of links between characters and their phonology and meaning. The aim of the experiment was to address the following questions:

1. Is early learning speed differently affected by categorisation of phonological or semantic characteristics?
2. Does categorical grouping of characters effect the earliest learning stage? If yes, does it facilitate the earliest stages of learning?

As reviewed in Chapter 2, semantic category is believed to have either a positive or a negative effect on learning L2 words, depending on the learner's level of proficiency. Developed learners benefit from semantic category (e.g., Hoshino, 2010) while first-shot learners do not (e.g., Tinkham, 1993, 1997). Effects of phonological similarity are also mixed for L2 vocabulary learning. Studies like Henning's (1973) and Wilcox and Medina's (2013) showed that for beginners of L2 English, the effect of acoustic similarity is inhibitory on learning. In contrast, Szmalec et al's study (2009) pointed out that participants recall better for newly learnt words with repeated phonemes.

The grouping of stimuli by similarity has been acknowledged to be a strategy of vocabulary learning, which is said to be fairly common among learners. Categorising words according to some characteristics by making specific relations among words in a group would help learners to remember them more effectively (Fu, 2005). On one hand, Chinese characters can be distributed into categories according to many criteria, for instance, by radicals. For bi-part characters, there is a radical that suggests the semantic category to which they belong, i.e., the semantic radical. Such radicals provide orthographic clues as well as semantic clues at the same time. In most cases, orthographic and semantic clues overlap with each other and are impossible to separate. The productive semantic radicals make it possible to group characters according to the shared radical. Actually, Chinese dictionaries often use radicals as index. For L1 Chinese learners, the semantically productive radicals are often emphasised by teachers (Jin, Lee, & Lee, 2013). However, by far, no study has been published to confirm the effect of semantic category on L2 Chinese character learning, especially during the initial learning stage. Experiment 1 is designated to test the effect of such groups on the early learning of L2 Chinese characters.

On the other hand, homophones are very common among Chinese characters. It is easy to find a list of characters sharing the same pronunciation and tone, whereas semantics, lexical categories and even their appearances are different. Large number of homophones enable group learning of homophonic characters in Chinese language. However, there is a gap in the existing literature regarding effects of phonological grouping, especially during very beginning of learning L2. Thus, Experiment 2 focuses on the effect of phonological grouping and the initial learning of Chinese characters.

Experiment 1: Effects of Semantic and Orthographic Grouping

This study was designed to explore the effect of orthographic and semantic grouping on the initial learning of Chinese characters. It consisted of two phases: a self-paced learning phase and a testing phase of 160 trials of matching task. The effect of grouping on the early consolidation of learning was investigated by adding a second testing phase 24 hr after the initial learning phase during which participants had slept (Gaskell & Dumay, 2003).

The experiment examined the effect of orthographic and semantic grouping on the learning of Chinese characters. A between subject design was used to allow the same sets of stimuli to be tested in both the grouped and ungrouped learning conditions. The independent variable was the method of learning for orthographic and semantic related Chinese characters. It had two levels: grouped and ungrouped learning. The dependent variables were speed and accuracy of learning, which was measured during test performance. The present study comprised a learning phase and a testing phase. There were two identical sessions during the testing phase: one immediately following the learning phase and another approximately 24 hr after the learning phase.

Method of Experiment 1

Participants.

Forty-six participants (43 women, 3 men) were recruited through the *Research Participation Scheme (RPS)* from the School of Psychology at the University of Birmingham. All participants were adult native British English speakers without any prior learning experience of

East Asian languages, such as Chinese, Japanese, and Korean. They all had normal or corrected to normal vision. No participants reported language impairments such as aphasia or dyslexia.

Materials.

Twenty simplified Chinese characters were selected from *the List of General Standard Chinese Characters* (2013). Each character represents a monosyllabic Chinese word, which is a simple, concrete noun. These 20 words comprised five sets of four characters which share a semantic radical. The semantic radical represents the meaning of a category which can be found across characters within the set, e.g., the shared radical 犭 between 狗 (dog) and 猫 (cat) means *animals*.

The full stimuli list is given in Table 3.1(p. 47). Pronunciations were written in pinyin and *International Phonetic Alphabet* (IPA)[4].

In each set, there were four Chinese characters with minimal phonological overlapping, such as shared phonemes. Each character was translated into English with simple and common English words. No significant difference between sets within the list was found regarding number of strokes, $F(4, 15) = 1.87$, $p = .17$; nor the frequency of the English translation, $F(4, 15) = 1.35$, $p = .30$. The frequency of English translation[5] was gained from the package N-Watch (Version 2007.10.1) (Davis, 2005).

[4] In IPA, ˥, ˧˥, ˨˩˦, and ˥˩ are tone letters, which refer to the four tones in Mandarin Chinese: *high level*, *mid rising*, *low dipping*, and *high falling* respectively. Please note that IPA letters were never presented to participants before, during or after any experiment in this thesis.

[5] The English words frequency here refers to the total CELEX (COBUILD) frequency.

Table 3.1

The Orthographic and Semantic List of Experiment 1.

Semantic radical	Chinese character	Pronunciation: Pinyin	Pronunciation: IPA	Stroke number	English translation
犭	狗	gǒu	[gou˨˩˦]	8	*dog*
	猫	māo	[mɑu˥]	11	*cat*
	猪	zhū	[tʂu˥]	11	*pig*
	猴	hóu	[xou˧˥]	12	*monkey*
氵	瀑	pù	[phu˥˩]	18	*waterfall*
	湖	hú	[xu˧˥]	12	*lake*
	洪	hóng	[xʊŋ˧˥]	9	*flood*
	溪	xī	[ɕi˥]	9	*brook*
木	林	lín	[lin˧˥]	8	*forest*
	树	shù	[ʂu˥˩]	9	*tree*
	枝	zhī	[tʂɿ˥]	8	*branch*
	松	sōng	[sʊŋ˥]	8	*pine*
月	腿	tuǐ	[thueɪ˨˩˦]	13	*leg*
	胸	xiōng	[ɕyʊŋ˥]	10	*chest*
	腰	yāo	[jɑu˥]	13	*waist*
	腹	fù	[fu˥˩]	13	*belly*
衤	衫	shān	[ʂän˥]	8	*shirt*
	裤	kù	[khu˥˩]	12	*trousers*
	袜	wà	[wä˥˩]	10	*socks*
	裙	qún	[tɕhyn˧˥]	12	*skirt*

The reading app TTSApp was used to generate auditory stimuli with the audio base Microsoft Lily (Chinese). Each sound file contained the pronunciation of the character which was read once. The mean length of the sound file was 816 ms (*SD* = 150). Each sound stimulus in the learning task consisted of eight iterations with 500 ms intervals after each repetition, which was created with the programme Cool Edit Pro (Version 2.0). The mean length of the longer version was 10.44 s (*SD* = 1.14).

We used Adobe Photoshop (Version CS5) to create visual stimuli. All visual stimuli were images of Chinese characters and English words in bmp format. The size of each image of the characters was 530 kB, 425 * 425 pixels. Images of English words were 150 kB each of 425 * 122 pixels. Images of words in both languages were in their most commonly used font. For Chinese characters, the font was *Kai* (楷体), whereas for English words, it was *Times New Roman*.

Apparatus.

We used E-Prime (Version 2.0) to conduct the experiment and collect data. The experiment was carried out in a sound attenuated booth. Visual stimuli were displayed by a monitor, while auditory stimuli were played through headphones. For Experiment 1, only one participant was tested at a time.

Design and procedure.

The experiment was formed of one learning session and two testing sessions on two consecutive days. There were also two structured interviews after each testing session (see Figure 3.1, p. 49).

Before the start of the experiment on Day 1, the participant was given a verbal briefing about the experiment purposes, procedures and basic operations. Then s/he read the detailed written instructions and completed the consent form and the questionnaire about their language background. Both documents were stored and then transcribed into electronic form manually. The instructions, the consent form and the questionnaire can be found in Appendices A, B and C (pp. 211-214). Once the paperwork was checked, the learning session started. Once s/he finished the task, the participant could take an optional short break of up to 3 min. Then the first testing session began. In order to collect information about participants' learning strategies, the participant was interviewed after the testing session. These questions for the interview for both days can be found in Appendix D (pp. 215-216). The participant was required to return in approximately 24 hr for the experiment on the following day. For example, if a participant attended the Day 1 experiment from 9:00 a.m., s/he would be asked to take part in the experiment from 9:00 a.m. on Day 2.

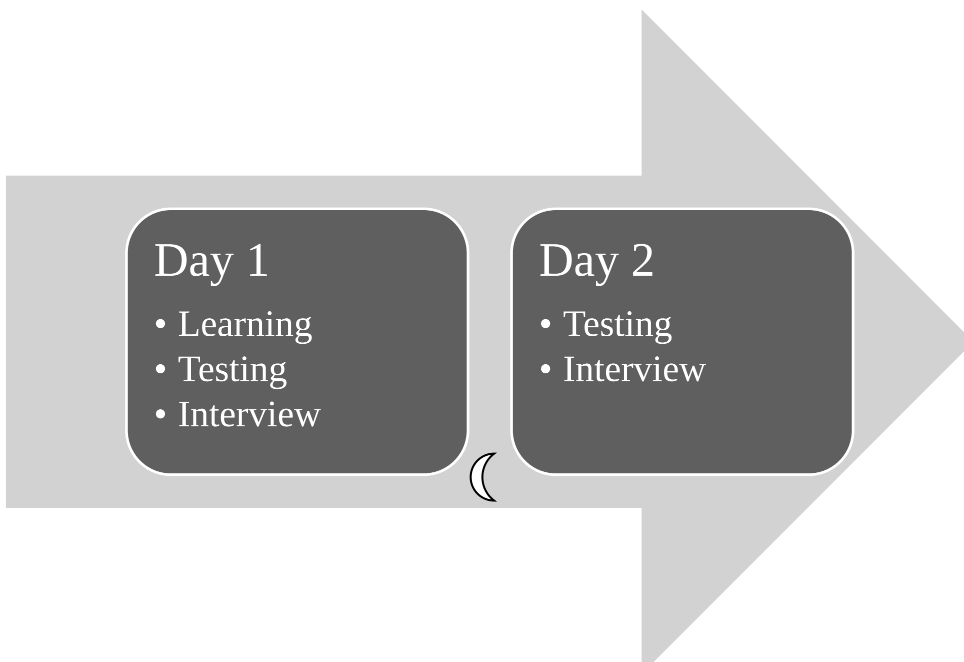


Figure 3.1. Schematic diagram of the structure of Experiment 1.

On Day 2, participants returned to complete another testing session with a different randomisation of trials. The second interview was carried out at the end of the experiment.

The experiments on Day 1 and 2 took about 50 min. Details of the design and procedure for each task are specified below.

Learning phase.

Design.

The learning phase was conducted on Day 1. As learning method was a between subject manipulation, each participant learnt the 20 characters either in semantically grouped sets or in ungrouped sets.

The grouped list consisted of five sets of characters. The four characters within each set all belong to the same semantic category and contained the same semantic radical. The ungrouped version was formed by distributing these 20 characters into five unrelated sets. Items in each ungrouped set had no shared radicals and no underlying links in meaning. We also kept the minimal number of shared phonemes among items within each set. For both the grouped list and the ungrouped list, five versions of rotation of sets were created using Latin square method. For individual sets, the order of presentation of items was randomised by E-Prime.

Procedure.

In the learning phase, we used a self-paced, E-Prime based learning programme. Before the experiment, participants were instructed to learn the Chinese character, its pronunciation and the correspondent English translation. As the present study did not involve language production, they were requested not to produce the sound after hearing them. For each single exposure, they heard the pronunciation of the character repeatedly up to eight times[6] with an interval of 500 ms after each repetition (see Figure 3.2).

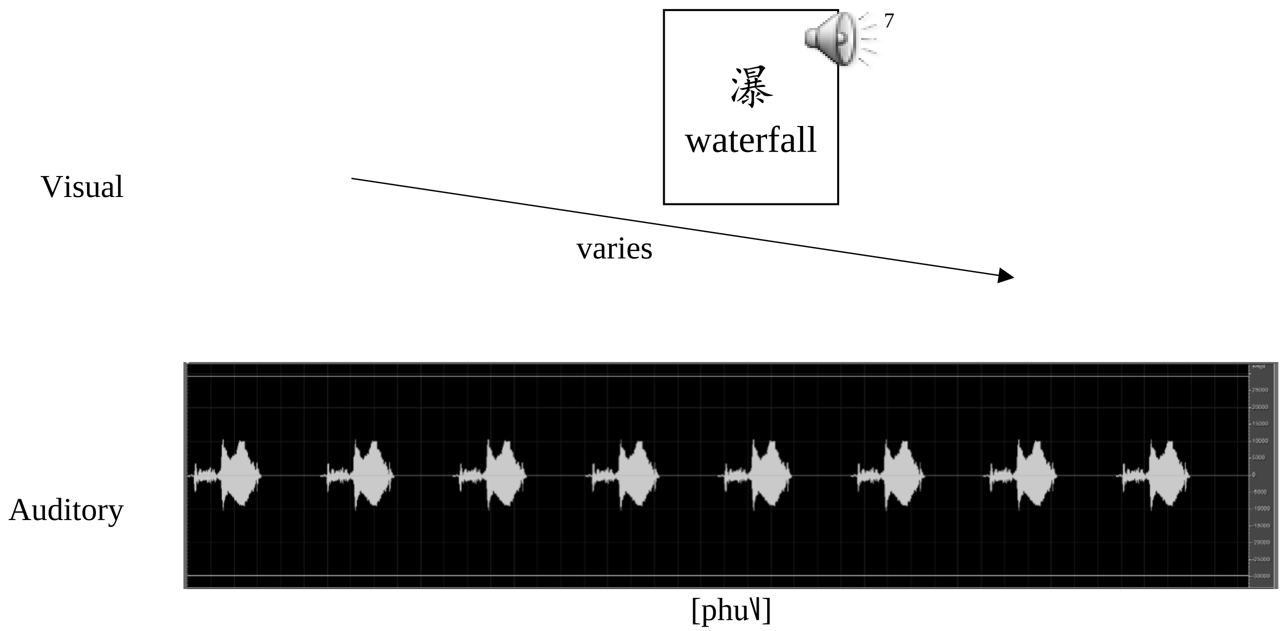


Figure 3.2. Schematic diagram of a single presentation of a character during learning session in Experiment 1.

---

[6] As shown in Figure 3.2 (p. 51), the auditory file of a certain character contained eight iterations of the sound. Although the participant could view the character as long as they like, the sound file would be played once only during a single exposure. Therefore, a participant would hear the sound eight times during a single exposure of a certain character if and only if s/he stopped the presentation of the current character by clicking the button after the auditory file was completely played.

[7] In this thesis, wherever the icon "" appears, it indicates there is a sound played through headphone. The icon was never shown during the experiment.

They were able to choose how long they spent on each character as well as repeat the same set of characters in a different order until they were satisfied. E-Prime automatically collected number of iterations of learning each set as well as the reaction time of each presentation of a certain character.

Testing phase.

Design.

The purpose of the testing phase was to evaluate the learning of the characters in terms of their pronunciation and meaning. The testing session consisted of a lexical judgement task.

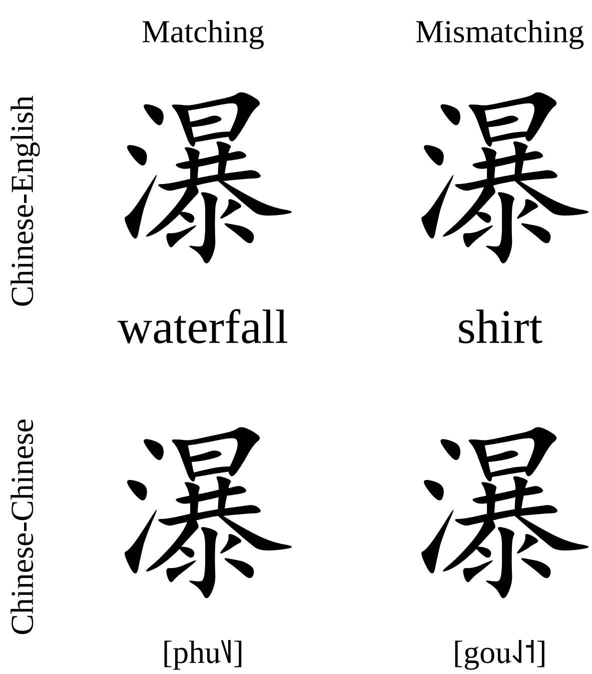


Figure 3.3. Design of the testing phase stimuli of Experiment 1, whereas the character 瀑 means *waterfall*; and [gou˩˧] is the pronunciation of the character 狗, which means *dog*.

Participants were instructed to judge pairings of characters with Chinese sounds or English words. The pairing could be either matching or mismatching. Mismatching pairings were formed with either with an English word or a sound from a different semantic group but within the list.

Consequently, there were in total four combinations of pairings for testing the learning of participants:

Chinese character + the correct English translation;

Chinese character + the correct Chinese pronunciation;

Chinese character + an incorrect English translation;

Chinese character + an incorrect Chinese pronunciation.

These combinations are illustrated in Figure 3.3 (p. 52).

Each testing session consisted of 160 trials. Every 40 trials formed a pair-block. There were four pair-blocks in each testing session. The order of trials was randomised within each pair-block. For every pair-block, one character occurred twice in two conditions: once in a matching condition and once in a mismatching condition. Each pair-block had the same number of Chinese-Chinese (C-C) or Chinese-English (C-E) pairings and the same number of matching and mismatching pairings. Up to four trials of one Matching condition (matching or mismatching) or one Pairing condition (C-C or C-E) could occur sequentially. The number of pairings in each condition was identical. The 40-trial pair-block was divided into two separate blocks of 20 trials when presented to participants.

In order to eliminate possible position effects for a certain trial, we created four unique randomisations of trials orders. This was realised by rotating the order of four pair-blocks by Latin square method and creating different randomisations of trials in each pair-block. Trial order was therefore different for all pair-blocks in the four versions.

Procedure.

Participants were firstly briefed about the task before the testing phase began. Then the E-Prime programme for the testing session started. The programme comprised instructions, a tutorial and formal testing tasks. The instruction about basic operations of the task were presented on the screen. After the instruction, there was a tutorial which was designed to help participants familiarise themselves with the timing of the testing task. In the tutorial, example trials of these four combinations were shown on the screen. Participants were asked to press any button at the end of the trial. After the tutorial, the formal test started following a short break.

A C-C trial contained a pairing of a Chinese character and a Chinese sound. The participant was asked to judge whether the pairing was correct or not. The trial began with a cross (+) as the fixation for 500 ms followed by a 500 ms blank screen. Then a character appeared in the centre of the screen for up to 5000 ms. At the same time, a Chinese sound was played with it once only. During the 5000 ms, the participant was asked to press *Yes* or *No* buttons as quickly as they could. Once a button was pressed or the participant failed to give a response within the 5000 ms, a fixed inter-trial interval of 1500 ms occurred. The following trial initiated after this interval. The time-course of a typical C-C trial is shown in Figure 3.4a (p. 55).

C-E trials (see Figure 3.4b, p. 55) had the same time-course, except that the character was presented simultaneously with an English word underneath it for 1000 ms.

For both types of trails, no feedback was given to participants.

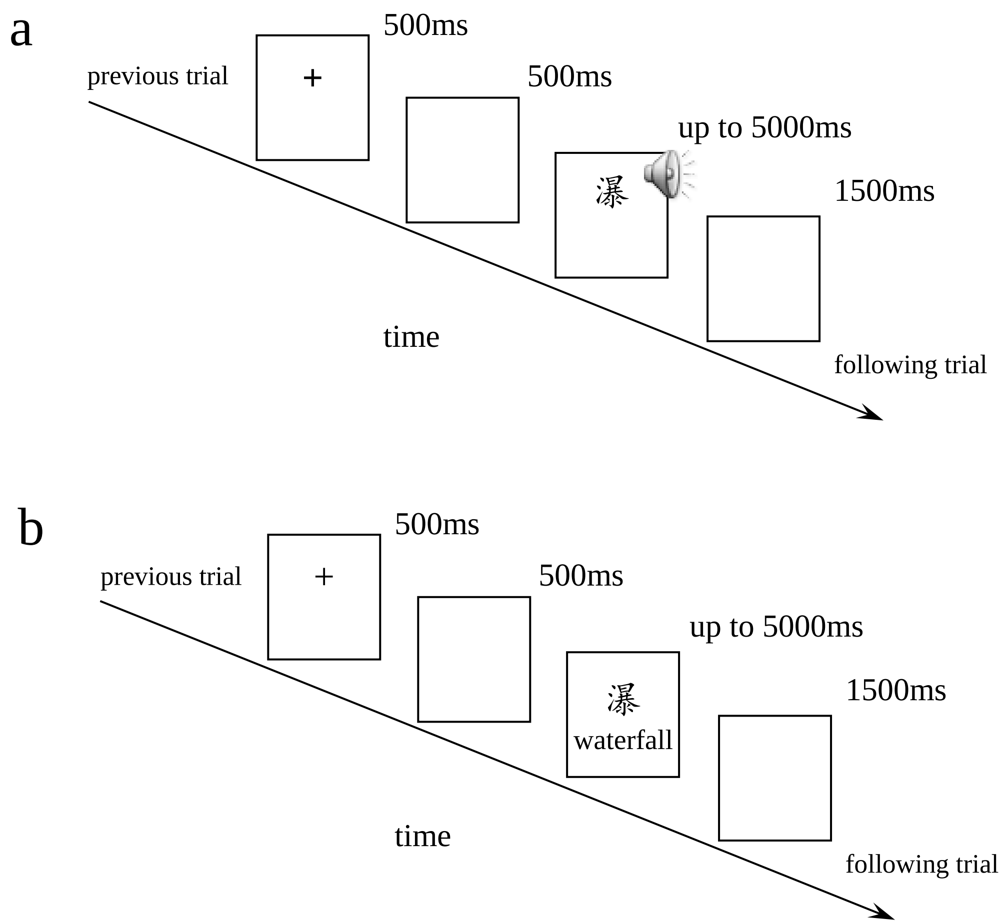


Figure 3.4. Timing of experiment trials of Experiment 1. (a) a Chinese-Chinese trial. (b) a Chinese-English trial.

Participants always pressed the Yes button with their dominant hand[8]. Between blocks, participants could take a short rest until they were ready to move to the next block. Responses and reaction times were automatically recorded by E-Prime. There was no briefing for the testing session on Day 2, but the tutorial was kept as a warm-up.

Interview.

There were two structured interviews across two days to collect information about participants' learning strategies. We used open-ended questions to gather learning strategies. Minutes of the interview were noted down while the whole process of the interview was recorded. The full set of interview questions is given in Appendix D (pp. 215-216).

Results of Experiment 1

Data analysis procedure.

Learning phase.

I used mixed effects linear models to analyse the data from the learning phase (Barr, Levy, Scheepers, & Tily, 2013) in R (R Development Core Team, 2011). There were two factors for this part of experiment: *Grouping* and *Set order*. Grouping, as the only between-participants manipulation, had two levels: grouped and ungrouped. Set order referred to the order that each set appeared in a learning session. As we rotated these five sets to create different randomisations, each set could appear in one of the five positions of sequence during a certain learning session

---

[8] In all experiments reported in this thesis, the Yes button always went with participants' dominant hand.

(first to fifth). Set order was dummy coded with the first position as the reference level. We analysed the effect of these factors on average viewing time for each character and for the number of iterations of each set, i.e., how many times each set was viewed.

Testing phase.

For the testing session, we analysed the data using mixed models in R. The error rate and reaction time data were analysed separately. We used mixed-effect logit models for the analysis of the error rate data (Barr et al., 2013; Jaeger, 2008). Mixed effects linear models were used for the analysis of reaction time (Barr et al., 2013). We excluded 35% (4,666 out of 13,440) of the reaction time data of trials with incorrect responses.

We manipulated four factors in the testing task, which were *Grouping*, *Matching*, *Pairing* and *Sleep*, which were all deviation coded. Grouping was the only between-subject manipulation in Experiment 1. It referred to the way participants learnt the 20 characters, either learning in semantic groups (*grouped learning*) or learning in a randomised group (*ungrouped learning*). The factor Sleep referred to whether testing had taken place *immediately after learning* or on the second day *after sleep*. Factors Pairing (*Chinese-Chinese or Chinese-English*) and Matching (*matching or mismatching*) reflected the nature of the experiment trials.

We started with a model with all factors and a fully specified random effect structure. We applied these following rules to reach convergence of the random effect structure: slopes for items were removed before for subjects; interactions were removed before main effects; interactions which were not included in the main model were removed before those in the model.

Then we reduced the model through *backward model selection* for fixed effects. After each step of reduction, we compared the model with the model before the reduction until the *p* value reached significance using ANOVA. The final model was the more complex one at this point.

For each model, fixed effects are summarised in tables. Random effect structures are presented in footnote along with results.

Results of the learning session.

Results of learning session are summarised in Figure 3.5, Tables 3.2 and 3.3. Generally, learners spent more time for the first set of characters than the other sets, while the number of iterations of all sets did not differ.

Table 3.2

Summary for Fixed Effects in the Mixed Linear Model for Number of Iterations in the Learning Phase of Experiment 1.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1.56	0.05	47.79	28.52	< .001***
Grouping	0.07	0.05	40.68	1.45	.15
Set order 2 vs.1	0.04	0.05	1527.70	0.74	.46
Set order 3 vs.1	0.03	0.06	42.66	0.48	.63
Set order 4 vs.1	-0.01	0.06	104.99	-0.22	.82
Set order 5 vs.1	-0.04	0.08	37.30	-0.48	.63

Note. [a]N = 1,688. In the random effect structure, we included a random intercept for items and subjects as well as a random slope for Set order for subjects, and random slopes of Grouping, and Set order for items. That is the maximal random effect structure for which convergence was reached.

***$p < .001$.

As shown in Table 3.2 (p. 58), there was no significant difference in the number of iterations between grouped learners and ungrouped learners. Besides, no significant main effect was reported among the five Set order positions. The mean number of iterations was 1.67 (*SD* = 0.77).

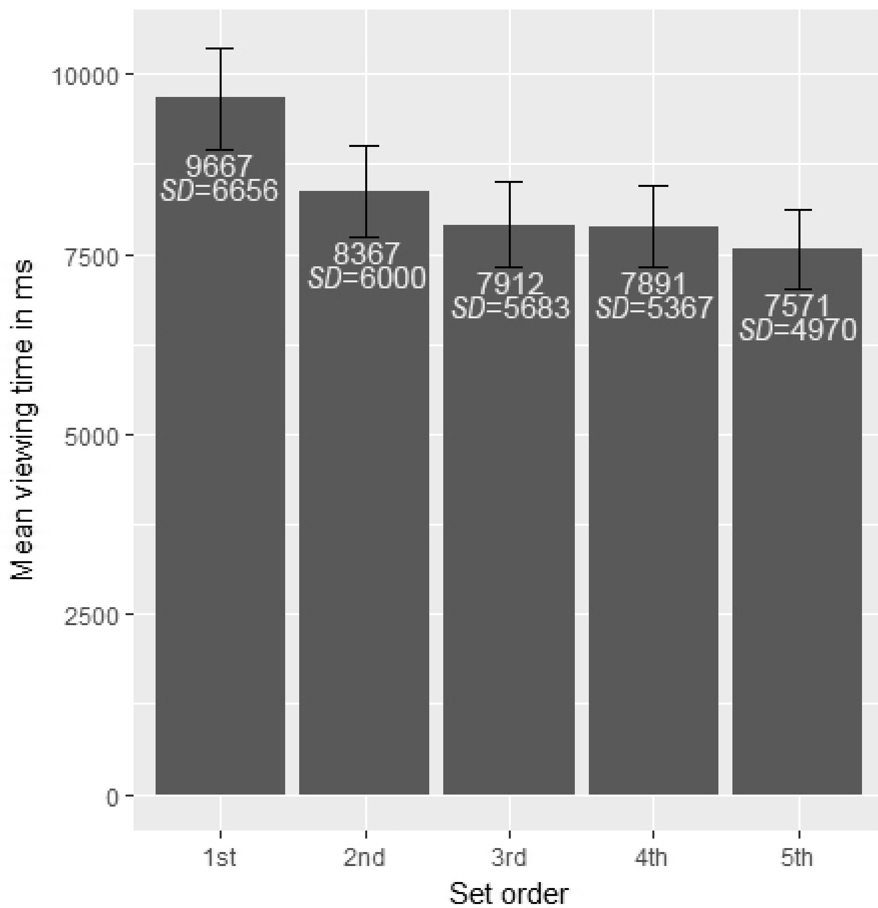


Figure 3.5. Results of viewing time of the main effect of Set order, Learning phase, Experiment 1.

For the average viewing time, we found a significant main effect of Set order ($p < .001$ for Set orders 2, 4 and 5 vs. 1 respectively and $p = .007$ for Set order 3 vs. 1, see Table 3.3, p.

59). Like the number of iterations, we did not find a significant main effect of Grouping for viewing time. As shown in Figure 3.5 (p. 59), the viewing time of items in the first position was significantly higher than the other positions, while the other four positions did not differ much. The average viewing time for each item was 8287 ms (*SD* = 5812).

Table 3.3

Summary for Fixed Effects in the Mixed Linear Model for Viewing Time in the Learning Phase of Experiment 1.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	10501.96	827.09	38.07	12.70	< .001***
Grouping	-758.40	504.91	40.39	-1.50	.13
Set order 2 vs.1	-1848.72	488.01	38.27	-3.79	< .001***
Set order 3 vs.1	-2033.34	709.11	39.14	-2.87	.007**
Set order 4 vs.1	-2296.98	595.98	38.69	-3.86	< .001***
Set order 5 vs.1	-2684.20	599.88	37.69	-4.48	< .001***

Note. [a]*N* = 1,688. In the maximal random effect structure that converged, we included a random intercept for items and random slopes of Grouping and Set order for items, and a random effect intercept for subjects and random slopes of Set order for subjects.

**$p < .01$, ***$p < .001$.

Results of the testing session.

Error rate.

The data for error rates is presented in Table 3.4 and Figures 3.6, 3.7 and 3.8. As shown in Table 3.4, in the complete model, almost every interaction including Matching reached significance. Overall error rates of matching trials were much higher than the error rate of the mismatching trials (*p* < .001). This suggested that the nature of matching and mismatching trials is different. Therefore, in order to understand the data further, we split the data into two subsets, matching trials and mismatching trials. We then designed two separate models accordingly. Each model contained all fixed effects in the model for all trials and excluded the main effect and interactions involving Matching. Results of the two models are also presented in Table 3.4 (p. 65).

Fifty percent of error rates: Complete guessing over matching trials?

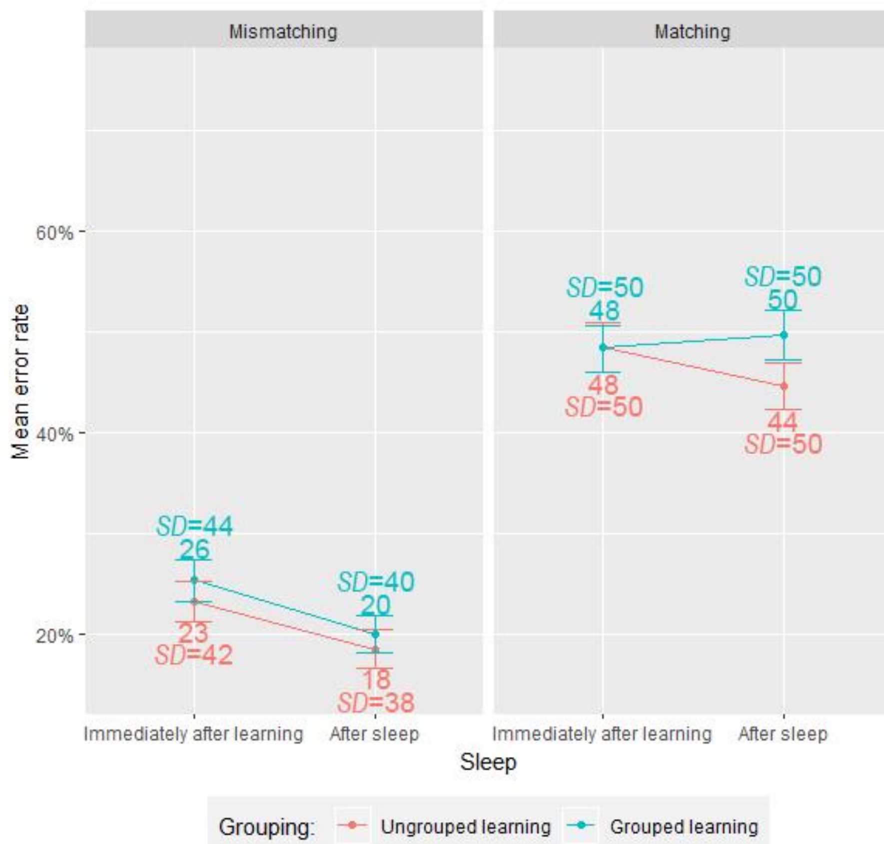


Figure 3.6. Error rates results of interaction Grouping by Matching by Sleep in Experiment 1.

In the model for all trials, the three-way interaction Grouping by Matching by Sleep reached significance ($p < .001$). The result of the matching trials further indicated that the interaction between Grouping by Sleep significantly interacted with Sleep ($p = .03$) for matching trials. However, post hoc Tukey's tests did not show the two groups significantly differed before or after sleep. As shown in Figure 3.6 (p. 61), grouped learners made more (but not significantly more) mistakes in matching trials after sleep than on Day 1, the error rate of matching trials stayed around 50%, which indicates complete guessing of a yes/no question such as trials in the testing phase. This trend was not seen over mismatching trials. This indicated that the nature of the underlying process of matching and mismatching trials could be quite different.

All learners were more error prone for C-C pairings than for C-E pairings while more biased to say "no" for matching C-C pairings.

As shown in Figure 3.7 (p. 63), the three-way interaction of Grouping by Pairing by Matching was significant ($p < .001$). Again, patterns of matching and mismatching trials were quite different. For matching trials, the interaction between Grouping and Pairing approached significance ($p = .07$). It seemed that grouped learners made more mistakes than ungrouped learners for C-C pairings. However, post hoc Tukey tests did not show significant difference between grouped and ungrouped learners among C-C pairings nor C-E pairings. For mismatching trials, there was no difference in error rate for C-C pairings for both groups of participants, while grouped learners made more mistakes than ungrouped learners for C-E pairings although the effect did not reach significance.

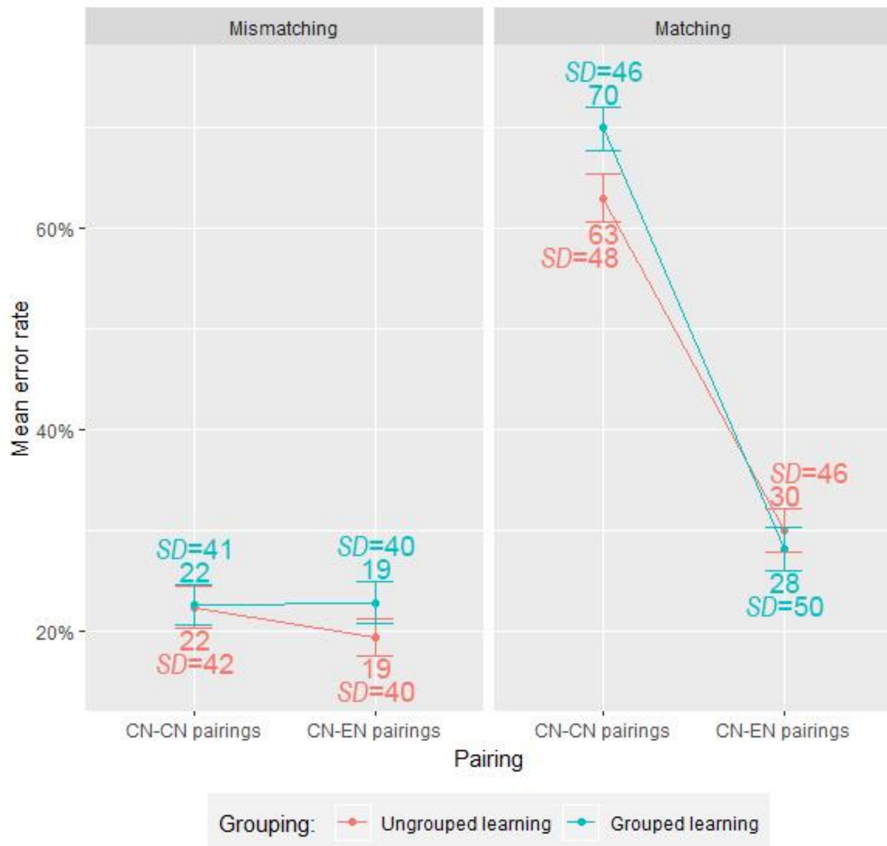


Figure 3.7. Error rates results of interaction Grouping by Pairing by Matching in Experiment 1.

In general, participants were much more likely to make mistakes among C-C pairings than C-E pairings ($p < .001$). That trend could also be seen among matching trails, as all participants had higher error rates for C-C pairings than for C-E pairings ($p < .001$). However, error rates of C-E and C-C pairings for both groups of participants did not significantly differ for mismatching trials. Therefore, it suggested that participants, regardless of groups, were biased to press “no” for matching C-C trials.

All participants were less error prone after sleep overall than Day 1.

It was observed that for all trials, participants made significantly fewer mistakes after sleep (p < .001) than they did on Day 1.

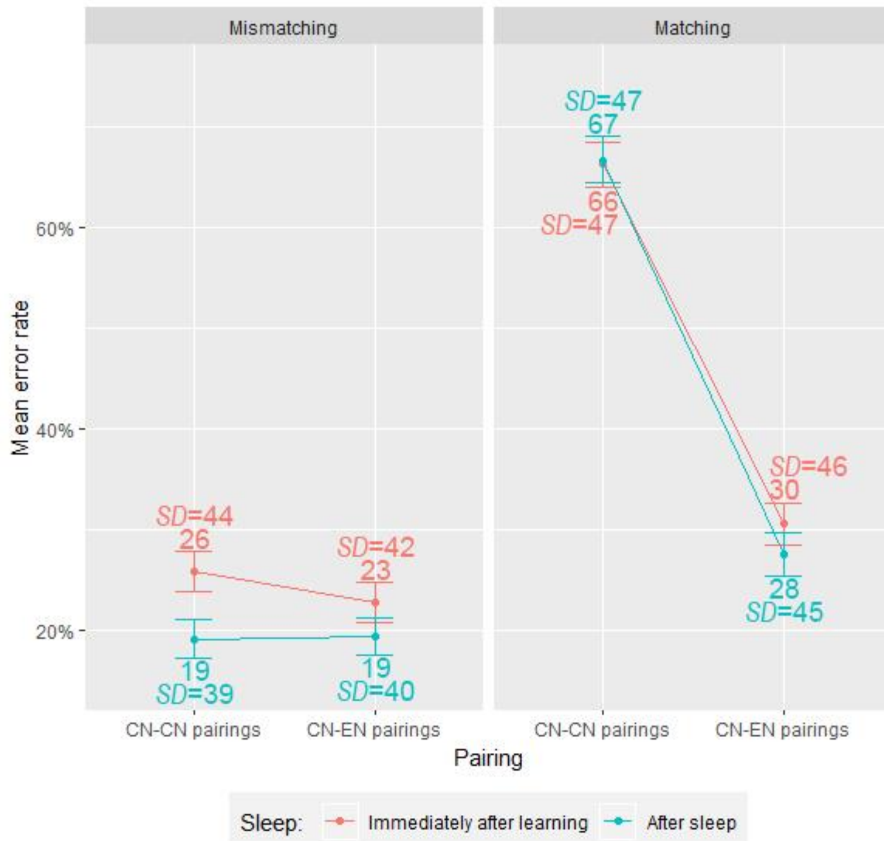


Figure 3.8. Error rates results of the interaction of Pairing by Matching by Sleep of Experiment 1.

As shown in Figure 3.8, the three-way interaction Pairing by Matching by Sleep reached significance (p = .02). For matching trials, the two-way interaction Pairing by Sleep was also significant (p = .01), while post hoc Tukey tests did not show significant difference between Day 1 and Day 2 among trials of either pairing types. For mismatching trials, pairing only marginally interacted with Sleep (*p* = .07). However, post hoc Tukey's HSD tests showed that after sleep, there was a significant decrease in error rate for C-E pairings at the level of .001, while the error rates for C-C pairings showed a marginal drop at *p* = .06.

Table 3.4

Summary of Fixed Effects in the Mixed Logit Model for Error Rates in the Testing Tasks of Experiment 1.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-0.87	0.12	-7.56	< .001***
Grouping	0.07	0.09	0.79	.43
Pairing	0.54	0.09	6.30	< .001***
Matching	-0.70	0.09	-7.45	< .001***
Sleep	0.11	0.03	4.32	< .001***
Grouping by Pairing	0.03	0.06	0.48	.63
Grouping by Matching	0.00	0.08	0.001	.99
Pairing by Matching	-0.40	0.07	-5.89	< .001***
Grouping by Sleep	-0.03	0.02	-1.43	.15
Matching by Sleep	0.07	0.02	3.32	< .001***
Pairing by Sleep	-0.01	0.02	-0.26	.79
Grouping by Pairing by Matching	-0.08	0.02	-3.75	< .001***
Grouping by Matching by Sleep	0.04	0.02	1.74	.08†
Pairing by Matching by Sleep	0.05	0.02	2.39	.02*
Model for matching trials[b]				
Intercept	-0.13	0.14	-0.88	.38
Grouping	0.08	0.10	0.73	.46
Pairing	0.93	0.09	9.97	< .001***
Sleep	0.04	0.04	1.03	.30
Grouping by Pairing	0.11	0.06	1.82	.07†
Grouping by Sleep	-0.07	0.03	-2.20	.03*
Pairing by Sleep	-0.07	0.03	-2.54	.01*
Model for mismatching trials[c]				
Intercept	-1.62	0.16	-10.31	< .001***
Grouping	0.05	0.14	0.40	.69
Pairing	0.11	0.14	0.80	.42

Sleep	0.20	0.04	4.83	< .001***
Grouping by Pairing	-0.03	0.09	-0.37	.70
Grouping by Sleep	0.00	0.04	0.12	.90
Pairing by Sleep	0.06	0.03	1.78	.07†

Notes. [a]*N* = 13,440, log-likehood = -6,975.4. The maximal convergent random effect structure of the model for all trials contained a random intercept for subjects and items and random slopes of Grouping, Sleep, Matching by Pairing for items, and random slopes of Matching, Pairing, and Sleep for subjects.

[b]*N* = 6,720, log-likehood = -3,823.7. The random effect structure of models of matching contained a random intercept for items and subjects and random slopes of Pairing, Grouping, and Sleep for items, and random slopes of Pairing, and Sleep for subjects. That was the maximal random effect structure for which convergence was reached.

[c]*N* = 6,720, log-likehood = -3,044.2. The maximal random effect structure that converged for mismatching trials included a random intercept for items and subjects and random slopes of Pairing, Grouping, and Sleep for items, and random slopes of Pairing, and Sleep for subjects.

†*p*. < .1. **p* < .05. ****p* < .001.

Reaction time.

Results of reaction time data are presented in Figures 3.9 and 3.10, and Table 3.5.

Overall participants were faster for the C-E pairings ($p < .001$). As seen in Table 3.5 (p. 67), significant interactions found among the model for all trials all contained the factor pairing. This indicated C-C and C-E pairings could be of different nature. Therefore, we split the data into two subsets over pairing. Procedures of the separation of subsets and model design were basically the same with the one for the analysis of error rates data (for further details, see pp. 60-61).

Results of the two models for C-C and C-E pairings are shown in Table 3.5 (p. 69).

All participants responded faster after sleep than Day 1.

We found a significant main effect of Sleep in the model for all trials ($p < .001$). As shown in Figure 3.9, for C-C pairing, grouped learners and ungrouped learners were all faster after sleep ($p < .001$). For C-E pairings, sleep helped to reduce reaction time ($p < .001$) as well. Although there was a marginal significant three-way interaction of Grouping by Pairing by Sleep ($p = .08$), there was no significant interaction of Grouping by Sleep neither in the C-C nor in the C-E data.

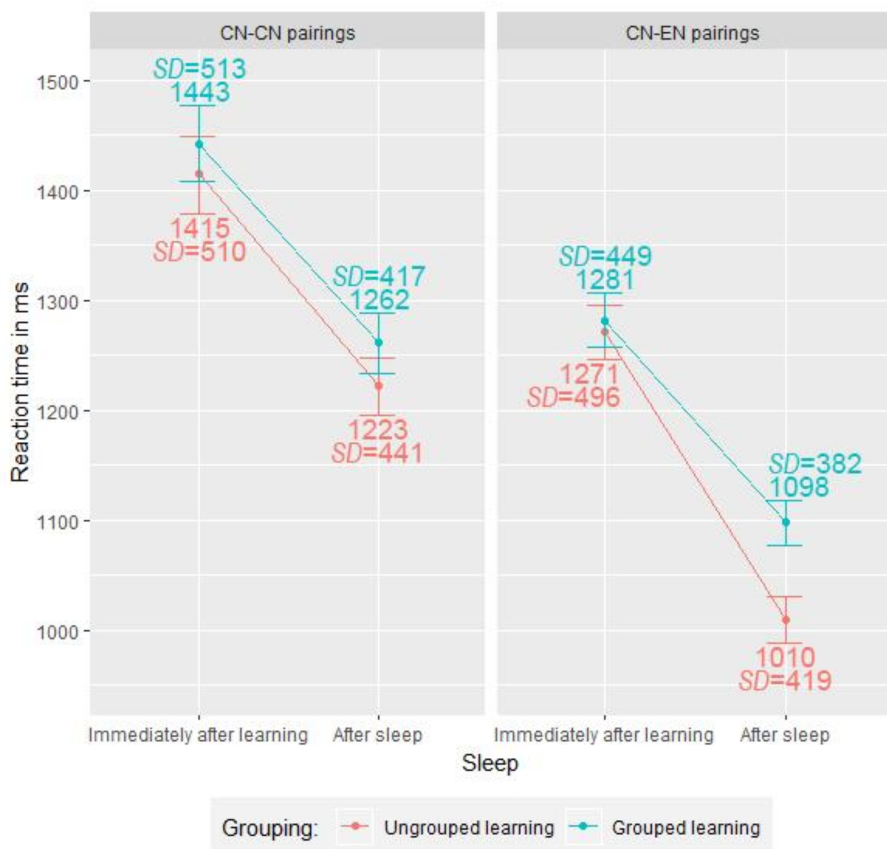


Figure 3.9. Reaction time result of Grouping by Pairing by Sleep, Experiment 1.

Participants were significantly faster for mismatching C-C trials than for matching C-C trials.

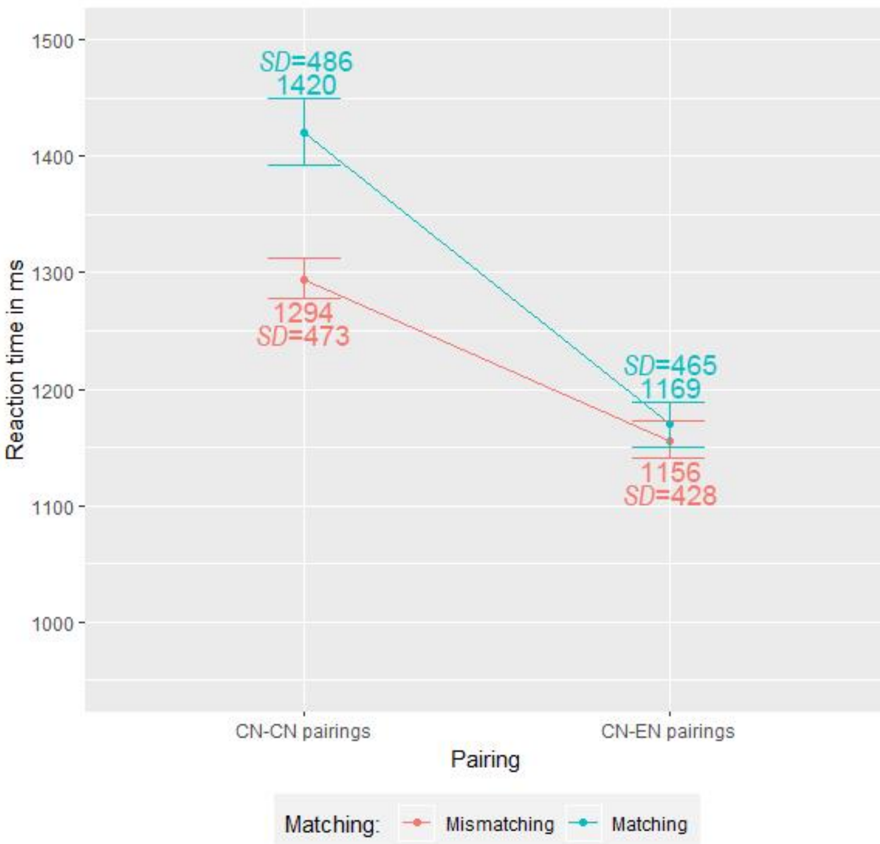


Figure 3.10. Results of Pairing by Matching for all trials, Reaction time, Experiment 1.

We also found that Pairing interacted with Matching ($p < .001$, see Figure 3.10) for the overall data. Generally, participants were slower on matching trials ($p < .001$). For C-C trials, the main effect of matching reached significance ($p < .001$), while for these two matching conditions did not differ significantly among the C-E trials., and they were faster for C-E pairings compared to C-C pairings ($p = .04$).

Table 3.5

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time in the Testing Tasks of Experiment 1.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1272.56	39.02	46.47	32.61	< .001***
Grouping	33.14	37.28	10.52	0.89	.38
Pairing	105.52	15.69	46.60	6.72	< .001***
Sleep	102.03	13.05	40.27	7.81	< .001***
Matching	-41.32	10.11	35.84	-4.09	< .001***
Grouping by Pairing	-0.72	12.16	39.75	-0.06	.95
Grouping by Sleep	-10.38	12.45	40.17	-0.83	.41
Pairing by Sleep	-6.86	3.96	8561.15	-1.73	.08†
Pairing by Matching	-30.90	5.76	33.12	-5.36	< .001***
Grouping by Pairing by Sleep	6.93	3.96	8563.76	1.75	.08†
Model for Chinese-Chinese pairings[b]					
Intercept	1379.75	43.24	46.48	31.91	< .001***
Grouping	30.67	40.88	39.98	0.75	.46
Sleep	95.21	15.79	41.33	6.03	< .001***
Matching	-74.49	17.53	29.97	-4.25	< .001***
Model for Chinese-English pairings[c]					
Intercept	1168.82	40.95	53.62	28.55	< .001***
Grouping	28.37	36.46	41.06	0.78	.44
Sleep	110.00	13.80	41.67	7.97	< .001***
Matching	-7.97	11.19	24.32	-0.71	.48

Note. [a]*N* = 8,774. In the random effect structure, we included a random intercept for items and subjects as well as random slopes of Sleep, and Pairing by Matching for subjects, and random slopes of Pairing, Matching, Grouping, and Sleep for items. That was the maximal random effect structure for which convergence was reached.

[b]*N* = 3,736. For the models of C-C pairings, the maximal random effect structure that could converge consisted of an intercept for items and subjects, and random slopes of Grouping, Sleep, and Matching for items and random slopes of Matching and Sleep for subjects.

[c]*N* = 5,038. The random effect structure for C-E pairings contained a random intercept for items and subjects, and random slopes of Grouping, Sleep, and Matching for items and random slopes of Matching and Sleep for subjects. That was the maximal random effect structure that converged.
[†]*p*. < .1. ****p* < .001.

Before moving into the next experiment, here is a brief short summary of the findings of Experiment 1.

For the learning session, all participants spent significantly more time for the first set they learnt. No difference was found between participants of these two groups. Analysis results of testing session showed that regardless of groups, participants made more mistakes for matching trials than mismatching trials. Specifically, they were more error prone for C-C trials than for C-E trials. When responding matching trials, they were probably doing a complete guess. They also presented a bias to say "no" among matching C-C trials. Among C-C trials, they were faster for mismatching ones than matching ones. Moreover, participants were all less error prone and faster after sleep. However, although the difference was not significant, grouped learners were more error prone after sleep than they did on Day 1, especially for C-C pairings.

Now I will switch the topic into learning Chinese characters in groups of homophones.

Experiment 2: Effects of Homophone Grouping

Experiment 2 examined the effect of phonological grouping on the initial learning of Chinese characters. During skilled L1 reading, phonological neighbours of the target word like homophones could be activated simultaneously. Repeated features, such as syllables, across newly learnt words were reported to be better recalled according to Szmalec et al’s study (2009). Therefore, as learning in homophone sets reduces the number of representations, we predict that grouping by homophone leads to higher accuracy and/or shorter latency during the testing session.

Method of Experiment 2

Participants.

Participants were 44 adult British English speakers from the University of Birmingham recruited via RPS. Among them, 40 were female and four were male. Recruitment and exclusion criteria were the same as in Experiment 1 of this chapter.

Materials.

There were 20 simplified Chinese characters each of which represents a monosyllabic Chinese word. They were all simple concrete nouns in Simplified Chinese also selected from *the List of General Standard Chinese Characters* (2013). Those 20 words formed five sets of homophones which shared segmental structure and tone, for example, 店 (shop) and 淀 (pool) are homophones which share the pronunciation *dian* [diɛn] and tone ˥˩ 51. Full details of the homophone list can be found in Table 3.6 on Page 72.

Within each set of homophones, there were four characters for which semantic relationship and orthographic overlap were minimised. Translations of these Chinese words were simple and common English words.

Table 3.6

The Homophone List of Experiment 2.

	Pronunciation			
Chinese character	Pinyin	IPA	Stroke number	English translation
金	jīn	[tɕin˥]	8	*gold*
襟			18	*collar*
津			9	*ferry*
筋			12	*muscle*
颌	hé	[xɤ˧˥]	12	*jaw*
核			10	*seed*
河			8	*river*
盒			11	*box*
零	líng	[liŋ˧˥]	13	*zero*
铃			10	*bell*
陵			10	*hill*
棂			11	*frame*
店	diàn	[diɛn˥˩]	8	*shop*
淀			11	*pool*
殿			13	*palace*
垫			9	*cushion*
宴	yàn	[jan˥˩]	10	*banquet*
雁			12	*goose*
焰			12	*flame*
堰			12	*dam*

There was no significant difference between sets in regard to number of strokes, $F(4, 15) = .31$, $p = .87$, nor the frequency of English word, $F(4, 15) = 1.28$, $p = .32$. Word frequency data came from N-Watch (Version 2007.10.1) package for English frequency.

The method for the generation of visual and auditory stimuli was the same as in Experiment 1. We also used TTSApp to generate sounds with the same audio base Microsoft Lily (Chinese). Mean length of the sound was 728 ms ($SD = 139$). Each individual sound then went through the similar procedures to create stimuli for the learning task, using Cool Edit Pro (Version 2.0). The average length of the extended sound was 9.82 s ($SD = 1.10$). Visual stimuli were created with Adobe Photoshop (Version 12.0). We applied the same settings of each stimuli, which was 530 kB, 425 * 425 pixels for Chinese and 150 kB, 425 * 122 pixels for English. We set Kai as the Chinese font and Times New Roman as the English one.

Apparatus.

Apparatus and settings of the experiment were all identical to Experiment 1 (see p. 48 for further details).

Design and procedure.

The overall experiment structure of Experiment 2 was identical to that of Experiment 1, comprising one learning session and two testing sessions occurring over two days (see Figure 3.1, p. 49).

The design and procedure of the learning and testing phases were in most part identical to Experiment 1 of this chapter. Any differences are specified below.

Learning phase.

We again manipulated the factor Grouping, with either grouped or ungrouped learning. The grouped list of the present experiment was formed of five sets of homophones. The four characters in each set shared the same pronunciation and tone. The ungrouped list was constructed by distributing these 20 characters into five unrelated sets. Items of each set had no shared radicals.

The time-course of each trial was identical to Experiment 1 of the same chapter (see Figure 3.2, p. 51).

Testing phase.

The testing phase was the same as for Experiment 1. There were again four pairing combinations (see Figure 3.3, p. 52). However, now, the mismatching trials were formed with a Chinese sound or and English word from a different homophone set but within the character list. The block structure, randomisation and testing procedure were identical to Experiment 1 of this chapter (see Figure 3.4, p. 55 for the time-course of trials).

Interview.

Like Experiment 1, we also held a structured interview by the end of each testing session. Questions for the interviews remained the same to Experiment 1 (see Appendix D, pp. 215-216 for full set of questions).

Results of Experiment 2

Data analysis procedure.

Learning phase.

For this experiment, we followed the same procedure of data analysis for the learning phase as we did in Experiment 1. We again used mixed effects linear models. We included two factors in the learning phase: the between subject factor *Grouping* (grouped/ungrouped) and *Set order* (1-5). We again dummy coded Set order with the reference level as the first. Grouping was deviation coded.

Testing phase.

As with Experiment 1, we also used mixed models in R to analyse data: mixed-effect logit models for error rate data and mixed effects linear models for reaction time data. 36% of trials (4,861 out of 13,440) were again excluded for the analysis of reaction time, as they were with incorrect responses.

We again manipulated 4 deviation coded factors, which were *Grouping* (Grouped/ungrouped learning), *Matching* (matching/mismatching), *Pairing* (Chinese-Chinese/Chinese-English) and *Sleep* (immediate after learning/after sleep). We followed the same method for model reduction and random slope removal as in Experiment 1 (see pp. 57-58 for details).

Results for the learning session.

Tables 3.7 and 3.8 summarise results of the learning phase, Experiment 2.

For the number of iterations, we did not find any difference between grouped and ungrouped learners. Neither did we find difference among Set order for number of iterations. The average number of iterations was 1.63 (*SD* = 0.74).

Table 3.7

Summary for Fixed Effects in the Mixed Linear Model for Number of Iterations in the Learning Phase of Experiment 2.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1.52	0.04	0.002	41.51	< .001***
Grouping	0.01	0.03	0.81	0.20	.84
Set order 2 vs.1	0.09	0.08	0.60	1.06	.29
Set order 3 vs.1	0.03	0.07	0.64	0.37	.71
Set order 4 vs.1	0.02	0.06	0.83	0.40	.69
Set order 5 vs.1	0.01	0.07	0.65	0.16	.88

Note. [a]*N* = 1,731. The maximal random effect structure that converged comprised a random intercept for both items and subjects, together with random slopes of Grouping, and Set order for items, and a random slope for Set order for subjects.

***$p < .001$.

For the viewing time, it seemed that participants spent longer time for the first set than the other sets. However, as shown in Table 3.8 (p. 77), there was no significant main effect of Grouping nor Set order. For each set, participants regardless of the way they learnt, spent statistically the same time for learning each set. The average viewing time for an item was 12,229 ms (*SD* = 30,666).

Table 3.8

Summary for Fixed Effects in the Mixed Linear Model for Viewing Time in the Learning Phase of Experiment 2.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	15385.37	3515.04	40.49	4.38	< .001***
Grouping	-1060.78	981.27	45.79	-1.08	.28
Set order 2 vs.1	-1432.19	1779.43	36.93	-0.80	.42
Set order 3 vs.1	-647.92	3089.98	39.40	-0.21	.83
Set order 4 vs.1	515.49	4400.41	40.10	0.12	.91
Set order 5 vs.1	2568.14	5888.24	40.54	0.44	.66

Note. [a]*N* = 1,731. For the random effect structure, there was a random intercept for items and subjects. It also contained a random slope for Set order for subjects and slopes of Grouping and Set order for items. That was the maximal structure which converged.

***$p < .001$.

Results of the testing session.

Error rate.

Table 3.9 and Figures 3.11 and 3.12 summarise the data of error rate. The four-way interaction Grouping by Pairing by Sleep by Matching reached significance ($p = .03$). To understand the complexity of the pattern presented by the data, we split the data over Pairing as all the significant interactions found in the model for all trials contained that factor. The model for each subset included all fixed effects and interactions from the model for all trials except for fixed effect and interactions containing Pairings.

After sleep, all learners made fewer mistakes, especially they were less error prone for mismatching C-C trials.

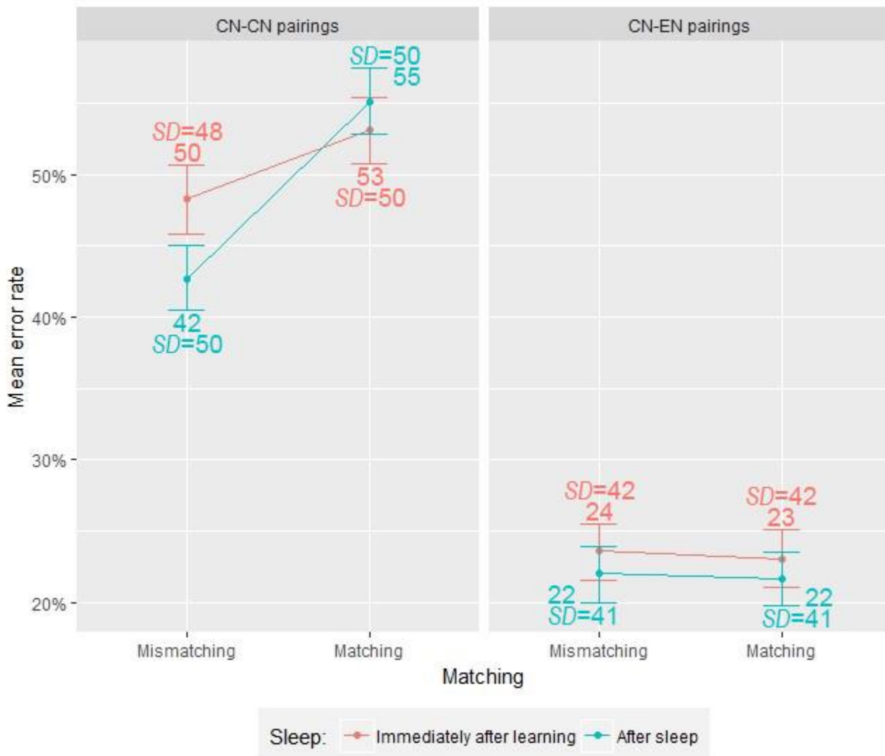


Figure 3.11. Error rates results of three-way interaction Pairing by Sleep by Matching of all trials, Experiment 2.

There was a significant main effect of Matching (p = .02), showing that in general, participants did better for mismatching C-C trials than C-C matching trials. There was a significant main effect of Sleep (*p* = .05). As shown in Figure 3.11, for C-C pairings, Sleep affected error rates differently for matching and mismatching trials (*p* = .05). Results of post hoc Tukey's HSD tests showed that for matching trials, regardless of grouping or not, error rates were not significantly affected by Sleep. However, for mismatching trials, the error rate significantly decreased after sleep at $p < .05$. Moreover, on Day 1, the average error rates for matching and mismatching C-C trials stayed around 50%, which suggest potential guessing over them. However, on Day 2, the effect of sleep on matching and mismatching C-C trials emerged at < .05.

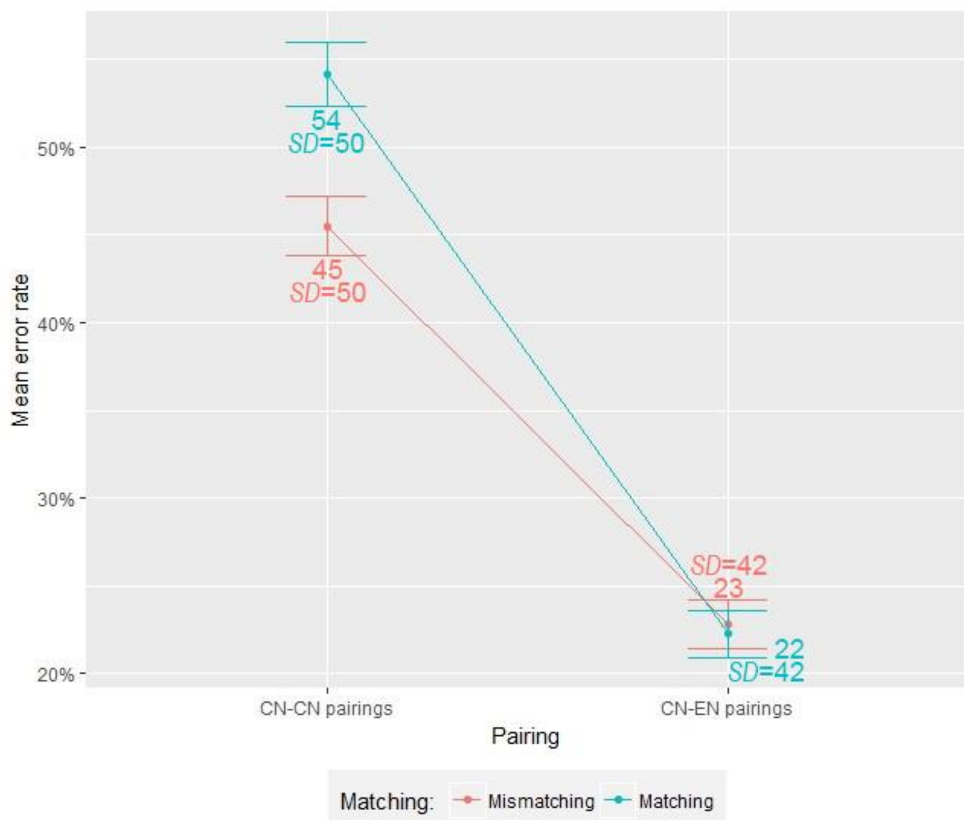


Figure 3.12. Error rates results of Pairing by Matching of all trials, Experiment 2.

Results of C-E pairings only showed that the error rate decreased as the main effect of Sleep reached marginal significance ($p = .09$). In order to further analyse this pattern, the data containing C-E trials data was further split into two subsets over Matching. As shown in Table 3.9 (pp. 80-82), only among mismatching trials the two-way interaction Grouping by Sleep reached significance (*p* = .02). However, post hoc Tukey tests showed that for mismatching trials, there was no significant difference between grouped learners and ungrouped learners after sleep.

Moreover, Pairing also interacted with Matching (*p* = .04). As shown in Figure 3.12 (p. 79), the difference was between pairings. While for C-E pairings error rates for matching and mismatching trials were quite alike, participants made many more errors for matching trials compared to mismatching trials for C-C pairings (*p* = .02).

Table 3.9

Summary for Fixed Effects in the Mixed Logit Model for Error Rates in the Testing Tasks of Experiment 2.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-0.80	0.10	-7.59	< .001***
Grouping	-0.02	0.10	-0.25	.80
Pairing	0.78	0.085998	9.09	< .001***
Sleep	0.06	0.03	1.99	.05*
Matching	-0.08	0.08	-1.10	.27
Grouping by Pairing	0.05	0.07	0.72	.47
Grouping by Sleep	0.02	0.02	0.71	.48
Pairing by Sleep	-0.02	0.03	-0.60	.55
Grouping by Matching	-0.04	0.06	-0.63	.53
Pairing by Matching	-0.11	0.05	-2.10	.04*
Sleep by Matching	0.04	0.03	1.23	.22
Grouping by Pairing by Sleep	-0.01	0.02	-0.24	.81
Grouping by Pairing by Matching	0.06	0.05	1.24	.21
Grouping by Sleep by Matching	0.03	0.03	1.04	.30
Pairing by Sleep by Matching	0.06	0.03	1.67	.09†
Grouping by Pairing by Sleep by Matching	-0.06	0.03	-2.11	.03*
Model for Chinese-Chinese pairings[b]				
Intercept	-0.01	0.06	-0.25	.80
Grouping	0.03	0.05	0.60	.55
Sleep	0.04	0.03	1.38	.17
Matching	-0.20	0.08	-2.40	.02*

Grouping by Sleep	0.01	0.03	0.42	.68
Grouping by Matching	0.02	0.07	0.28	.78
Sleep by Matching	0.09	0.05	1.99	.05*
Grouping by Sleep by Matching	-0.03	0.04	-0.71	.48
Model for Chinese-English pairings[c]				
Intercept	-1.58	0.18	-8.62	< .001***
Grouping	-0.07	0.16	-0.47	.64
Sleep	0.07	0.04	1.69	.09[†]
Matching	0.03	0.10	0.27	.79
Grouping by Sleep	0.02	0.03	0.68	.50
Grouping by Matching	-0.09	0.08	-1.14	.25
Sleep by Matching	-0.02	0.04	-0.50	.62
Grouping by Sleep by Matching	0.08	0.03	2.56	.01*
Model for matching Chinese-English pairings[d]				
Intercept	-1.60	0.20	-7.62	< .001***
Grouping	0.00	0.18	0.02	.98
Sleep	0.09	0.06	1.49	.14
Grouping by Sleep	-0.06	0.05	-1.26	.21
Model for mismatching Chinese-English pairings[e]				
Intercept	-.154	0.21	-7.42	< .001***
Grouping	-0.16	0.17	-0.92	.36
Sleep	0.05	0.05	0.86	.39
Grouping by Sleep	0.11	0.04	2.32	.02*

Notes. [a]N = 13,440, log-likehood = -7,565.2. The maximal random effect structure of the model for all trials which reached convergence included a random intercept for subjects and items, and random slopes, and random slopes of Matching by Pairing and Pairing by Sleep by Matching for subjects, and slopes of Matching by Grouping by Pairing, and Pairing by Sleep by Matching for items

[b]N = 6,720, log-likehood =- 4,486.7. In random effect structure, we included a random intercept for items and subjects, and random slopes of Matching by Grouping, and Sleep by Matching for items, and slopes of Matching by Sleep by subjects. That was the maximal structure in which convergence was reached.

[c]N = 6,720, log-likehood = -3,102.1. The random effect structure which reached convergence consisted of a random intercept for subjects and items, as well as random slopes of Matching by Sleep for subjects, and slopes of Matching by Grouping, and Matching by Sleep for items.

[d]N = 3,360, log-likehood = -1,553.6. The maximal random effect structure that converged contained a random intercept for subjects and items, and random slopes of Grouping and Sleep for subjects.

[e]N = 3,360, log-likehood = -1,562.3. The random effect structure in which convergence was reached included random intercept and a random slope of Sleep for items and subjects, and a random slope of Grouping for item.

[†]p. < .1. *p < .05. ***p < .001.

Reaction time.

The reaction time data are summarised in Figures 3.13 and 3.16, and Table 3.10.

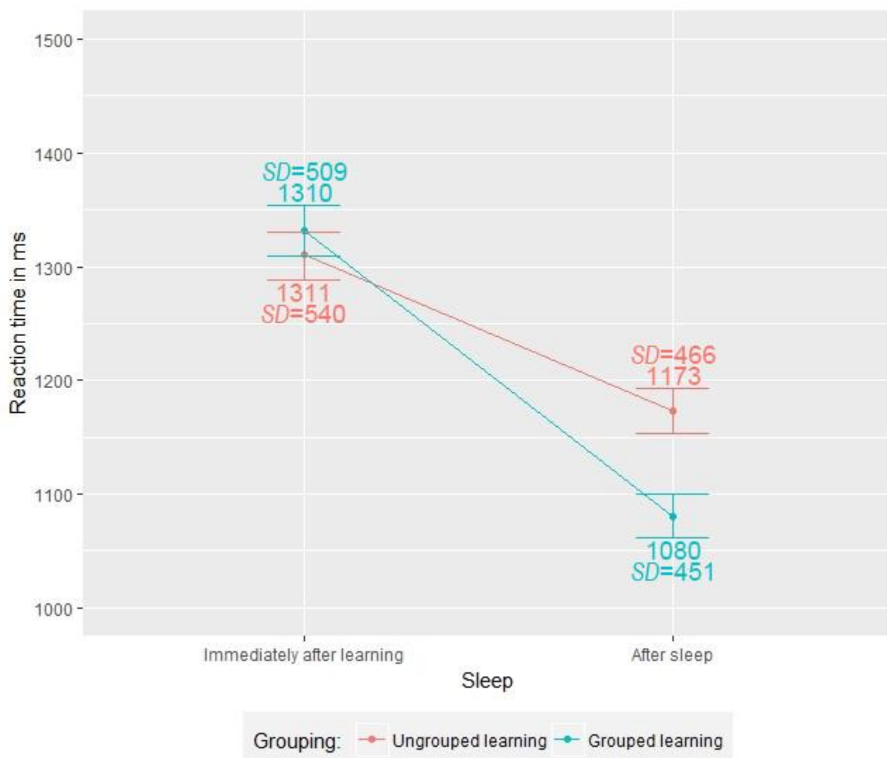


Figure 3.13. Results of the interaction Grouping by Sleep, Reaction time, Experiment 2.

Grouped and ungrouped learners all benefited from the consolidation of sleep while grouped learners benefited more than ungrouped learners.

Figure 3.13 (p. 82) illustrates the significant two-way interaction Grouping by Sleep ($p = .03$). Post hoc Tukey's HSD tests only showed that after sleep, both grouped learners and ungrouped learners became significantly faster than they did on the previous day ($p < .001$ for grouped learners and $p = .05$ for ungrouped learners respectively). It indicated that the sleep consolidation effect worked effectively on grouped learners and ungrouped learners, while benefited grouped learner more.

Participants were significantly faster for C-E trials than C-C trials, while they were marginally significantly faster for mismatching trials than matching trials.

Figures 3.14 to 3.16 shows reaction time differed regarding to factors of the design of trials.

Sleep significantly interacted with Matching ($p = .007$). Results of post hoc Tukey tests suggested that on Day 1, the difference in reaction time between matching and mismatching trials was significant at the level of .05. Namely during the first testing session, participants were significantly slower for mismatching trials than matching trials. However, after sleep such difference became insignificant (see Figure 3.14, p. 84).

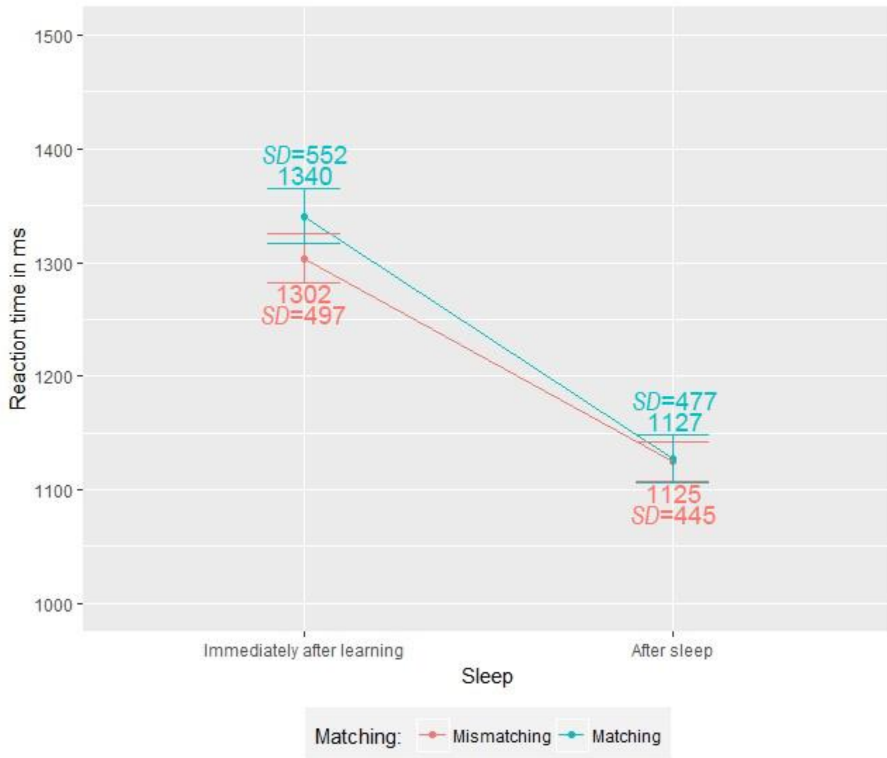


Figure 3.14. Results of the two-way interaction Matching by Sleep, Reaction time, Experiment 2.

There was a significant main effect of Sleep ($p < .001$). The drop of reaction time could be found over both kinds of pairings. As shown in Figure 3.15 (p. 85), the two-way interaction Sleep by Pairing also reached marginal significance ($p = .07$). Results of post hoc Tukey's tests showed that after sleep, drops in reaction time for C-C and C-E pairings were significant at the level of .001 respectively.

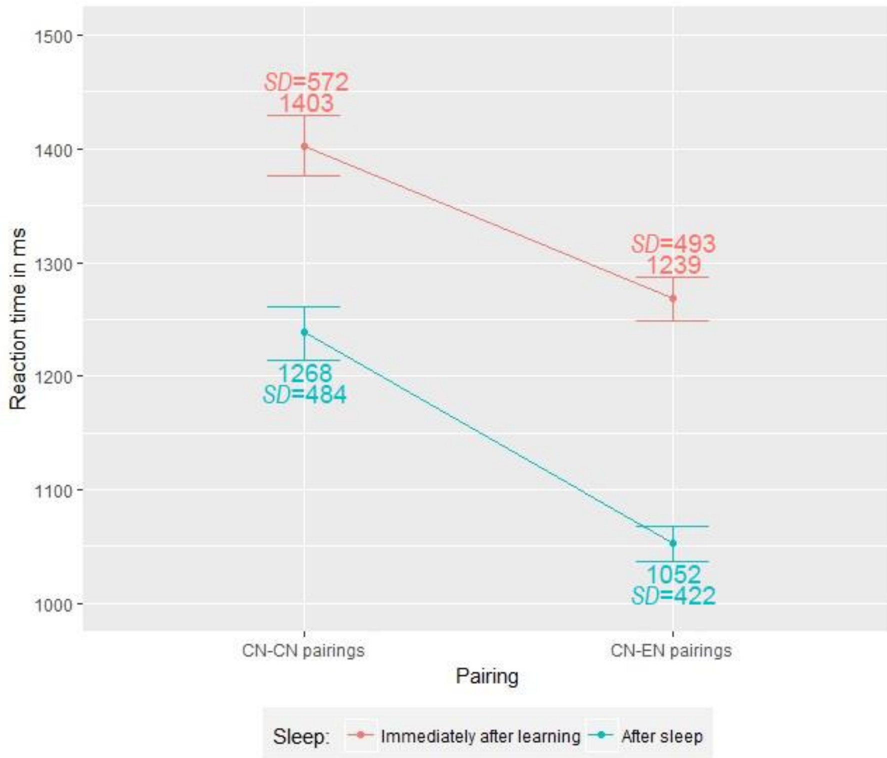


Figure 3.15. Reaction time results of the interaction Pairing by Sleep, Experiment 2.

The two-way interaction Pairing by Matching was significant (p < .001, see Figure 3.16, p. 86). According to results of post hoc Tukey tests, for C-C pairings, the reaction time of matching trials was longer than mismatching trials at .01 level of significance, while for C-E pairings the trend became only marginal significant and also reversed. Generally, participants spent longer time for matching trials than mismatching trials (p = .08).

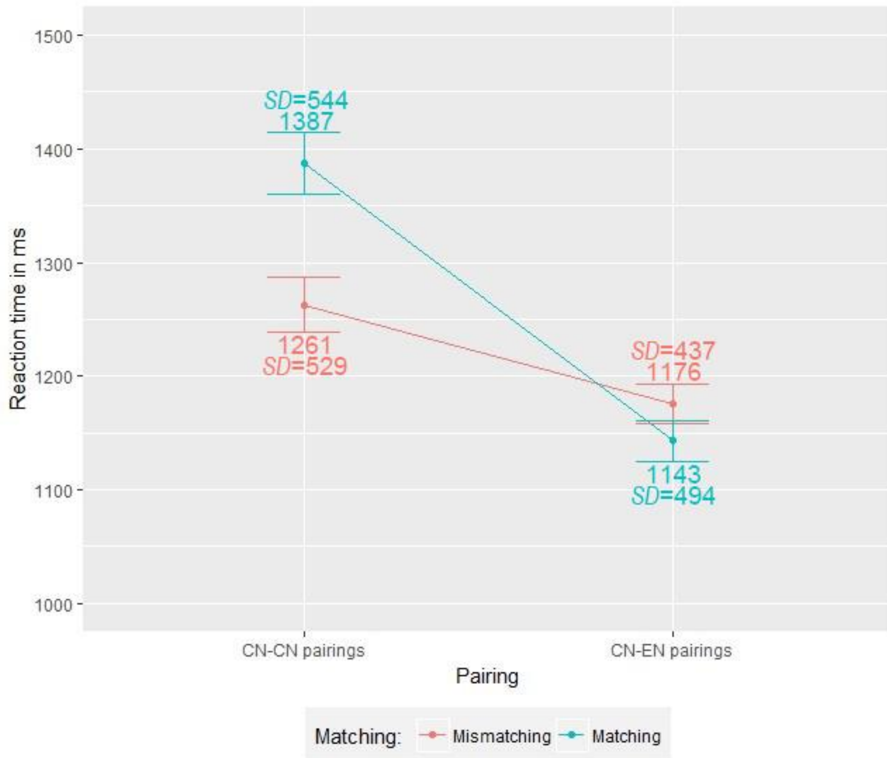


Figure 3.16. Results of the interaction Pairing by Matching, Reaction time, Experiment 2.

Table 3.10

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time in the Testing Tasks of Experiment 2.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1248.75	46.44	43.66	26.89	< .001***
Pairing	87.38	19.08	52.35	4.58	< .001***
Sleep	95.93	16.71	40.64	5.74	< .001***
Grouping	-8.64	37.94	39.40	-0.23	.82
Matching	-17.34	9.71	26.80	-1.78	.08[†]
Pairing by Sleep	-7.61	4.21	8359.87	-1.81	.07[†]
Sleep by Grouping	34.47	15.71	38.81	2.20	.03*
Sleep by Matching	-13.67	4.92	60.16	-2.78	.007**
Pairing by Matching	-26.50	6.07	40.76	-4.36	< .001***

Note. [a]N = 8,579. For the random effect structure of the model of all trials, we included a random intercept for items and a random intercept for subjects, together with random slopes of Grouping, Pairing, Sleep, and Matching for items, and slopes of Sleep by Matching, and Pairing by Matching for subjects.

[†]p. < .1. *p < .05. **p < .01. ***p < .001.

In Experiment 2, we found that sleep helped to consolidate the learning of Chinese characters, namely, in reducing error rates and reaction time. Grouped learners benefited from such consolidation more than ungrouped learners. Specifically, error rates significantly drop among mismatching C-C trials after sleep. However, all participants were more error prone for C-C trials than for C-E trials. They were faster in reacting to C-E trials than they did for C-C trials. Participants spent statistically the same time for each set during the learning session. In the next section I will discuss these findings in the two experiments.

Discussion

Experiments 1 and 2 were designed to answer the following questions. First, does categorical grouping of characters effect the earliest learning stages? If yes, is it facilitatory or inhibitory during the earliest stages of learning? Second, is early learning speed differently affected by categorisation of phonological or semantic characteristics? Each of these questions is addressed in the sections below.

Is Early Learning Speed Differently Affected by Categorisation of Phonological or Semantic Characteristics?

Learning speed was not affected by the way participants learnt the character for both category methods. During the learning phase, all participants spent almost the same average time for each character and the same mean number of iterations of each set. For both experiments, there was no difference between grouped learners and ungrouped learners on the number of iterations of each set. However, the viewing time differed for the order of sets. For Experiment 1, they always spent more time on the first semantic set, while the viewing time of the remaining semantic sets stayed the same. While for the homophone grouping experiment, the viewing time did not differ much. The significantly longer viewing time among the first semantic set could be attributed to their unfamiliarity with the Chinese language as well as the effort in finding features. Participants could spend more time on picking up the shared semantic radicals within each semantic set. Distinguishing radicals and linking them with semantics could take time, especially when they were facing a completely unfamiliar writing system for the first time. Once they found the existence of shared semantic radicals, they could easily pick up the next one and viewing time was reduced.

Does Categorical Grouping of Characters Effect the Earliest Learning Stages? Is It Facilitatory or Inhibitory then?

Both experiments revealed effects of grouping on learning. In general, from the tendencies presented in the data we could infer that phonological grouping by homophone could be facilitatory while the effect of semantic-orthographic grouping by radical could be inhibitory. The effects of such grouping methods emerged only after sleep.

The inhibitory effect of semantic radical grouping.

For grouping by semantic radicals, we found a complex pattern of results in Experiment 1. Firstly, after sleep, grouped learners were more error prone for matching trials than Day 1. Grouped learners were slower for C-E pairings than ungrouped learners. This suggested that the link from orthographic representations to semantics was not well formed for grouped learners after sleep. Interestingly, participants, regardless of grouped learning or not, could pick up the orthographic information from the semantic radical and then linked it with the semantic category of the group. Visual clues from the semantic radicals seemed to benefit grouped learners as well as ungrouped learners. That was reflected through the interview, five of 22 grouped learners and seven of 24 ungrouped learners mentioned that they noticed that there were parts of that character appeared repetitively which indicated a semantic group. All these results above clearly suggested that the formation of stable links to the Chinese sounds and semantic representations was incomplete.

Secondly, compared to ungrouped learners, grouped learners were more error prone for C-C pairings but less error prone for C-E pairings. Although grouped learners did not form the link from orthography to semantics well, here is a piece of evidence that visual clues of semantic radicals benefit the link from orthography to semantics more than building phonological representation in Chinese. Participants could rely on the visual clues and pay more attention to link the character to the meaning in English. Their reliance on visual clues might facilitate categorical grouping but hinder the building of individual links between the characters and its particular meaning. As the Chinese sounds were unfamiliar and not accompanied by any visual clues. During the learning phase, participants might have struggled to build a phonological representation to link with meaning. This is supported by responses in the interview. Regardless of learner group, most participants (43 out of 46) claimed that C-C trials were more difficult than C-E trials.

The inhibitory effect of semantic grouping was in line with previous studies (Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Higa, 1963; P. Nation, 2000; Tinkham, 1993, 1997; Waring, 1997). In Tinkham's studies (1993, 1997) and Waring's study (1997), participants needed more time to distinguish words in the same semantic category. Here in my study, the difficulty caused by semantic category was more obvious for grouped learners than ungrouped learners. Even results of studies in favour of the facilitatory effect of semantic grouping like Hashemi and Gowdasiaei (2005) showed that semantic grouping benefit more for advanced learners than first-shot learners. Through my experiments, the inhibitory effect of semantic grouping was proven to be applicable to the initial learning of L2 Chinese among alphabetic readers with several empirical evidence.

The inhibitory effect of semantic radical grouping could also be a result of the following reasons:

Firstly, these results of the radical grouping could be explained by the competing effect of learning category information versus learning individual character identity. The inhibitory effect of semantic and orthographic grouping might be due to the learning of category information represented by the semantic radical across items in the set. If categorical information was easier to learn than identity information, participants could quickly link the orthography with the meaning of the categorical radical. However, because the words in the same category overlapped in meaning, this would render the process of distinguishing individual characters more difficult. The transition from categorical learning to individual learning may need more time. This was reflected in comments during the post experimental interview, as some participants mentioned that the visual clue was easier to remember.

Secondly, the inhibitory effects could be attributed to the confounded semantic and orthographic information due to semantic radicals. Chinese characters often contain a semantic radical as part of the character. This makes it very difficult to separate the orthography and semantics of Chinese characters as the radical is part of the orthography as well as having an inherent meaning. There are some characters which do not comply with the rule involving semantic radicals, however, those characters are either very simple in spatial configuration (e.g., 手 *hand* and 足 *foot*) or very low frequency in meaning (e.g., 龟 *turtle*), which makes them unsuitable as learning materials for testing new learners.

Therefore, the next step is to eliminate the interference brought by orthography and investigate the effect of semantic grouping. This will be addressed in Chapter 4.

The facilitatory effect of homophone grouping.

When stimuli were phonologically grouped as homophones, grouped learners were less error prone for mismatching C-E trials and were faster than ungrouped learners for all trials after sleep. It is interesting that ungrouped participants became more error prone after sleep for mismatching C-E trials than Day 1. This suggested that phonological grouping helped participants build better phonological representations and at the same time, also facilitated the link between the existing semantic representations in English with the Chinese character. Learning limited number of phonological representations in sets could be the reason to that phenomenon. At the initial stage of learning, participants were not familiar with Chinese phonology at all. As reflected through the interview, all participants believe the C-C trials were more difficult than C-E trials as the common reason was all pronunciations “sound the same”. Learning in homophone groups give participants more time to familiar with shared sounds. It could also help participants focus on building the link from the orthography to semantics. In addition, the lower error rates and shorter reaction time for grouped learners compared to ungrouped learners indicated that the effect of phonological grouping was facilitatory without any evidence of a speed- accuracy trade-off (SAT).

In Experiment 2, we found that learning in homophone groups could contribute to a more robust link from orthography to semantics. One possible explanation for such facilitation could be the limited number of phonological representations. Although that applied to both groups, it is quite possible that explicit presentation of homophones helps participants focus more on the difference among characters of the same set. As reviewed in Chapter 1, the way of co-activating representations works differently in skilled reading in the first languages for rhyme and homophones. Would facilitatory effects could be found among other phonological grouping methods, such as in shared onsets or rimes? Experiment 2 only contained one learning session and two testing sessions. What is the effect of phonological grouping in a longer time span? Moreover, there were only five pronunciations to learn in Experiment 2. However, if a study is designed to testify the effect of rhyme grouped learning, the number of sounds would definitely grow to 20. Would the shared rhyme also help participants to build a solid phonological representation? Would the increase in representations overweight the facilitation? Those questions will be further explored in Chapter 5.

The consolidation effect of sleep.

The effect of sleep consolidation is further confirmed. From results of both experiments, sleep generally helped to consolidate the outcome of learning. For orthographic and semantic grouping, the error rates dropped greatly for both groups of learners. We also observed a shorter reaction time for matching C-E pairings for the phonological grouping experiment (Experiment 2). However, we also found that after sleep, the difference between levels of a certain factor became move obvious and significant. For example, the inhibitory effects of semantic grouped learning were only discovered during the second testing phase on Day 2. Here, it is clear that what sleep consolidates lies in line of previous literature: the integration into lexical competition (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). This is reflected by the finding that sleep contributed more to the link from orthographic representation to semantics in the mental lexicon.

Conclusion

In these two experiments in this chapter, we found preliminary inhibitory effects of semantic grouping and facilitatory effects of phonological grouping. These effects of grouping were observed in both reaction time and error rates. Generally higher error rates and longer reaction time were found among semantically grouped learners while phonological grouped learners were faster and more accurate than ungrouped learners. We also found the consolidation effect of sleep on results of learning. That consolidation effect was not only limited to improvements of learning result but including the enlargement of difference after sleep. However, the overall level of learning is not high enough. Participants need more inputs to distinguish between effects of building category links as opposed to individual representations.

The confound of orthography and semantics should be separated. For the phonological grouping, the next step is to test these effects of grouped learning of characters sharing onsets or rimes.

CHAPTER 4

EFFECTS OF SEMANTIC GROUPING

Introduction

The semantic radical grouping experiment (Experiment 1) showed an inhibitory effect of semantic and orthographic grouping in the initial learning of Chinese characters. Grouped learners made significantly more mistakes on Chinese-Chinese matching trials than ungrouped learners and had longer response times for Chinese-English pairings than ungrouped learners. During post experimental questions, participants reported that they remembered the category better than individual characters, especially for the semantic-orthographic experiment. However, does this necessarily suggest that grouping is not helpful for the building of early representations? Is the early stage of category building necessarily needed to be gone through which grouping could facilitate later or does grouping simply interfere with laying down strong long-term individual representation?

As discussed in the previous chapter, Experiment 3 in this chapter was designed to separate of the confounding effects from semantics and orthography brought by shared radicals in the same semantic category. While doing so, we need to keep word frequency and complexity constant. Characters fitting those features are rare to find. The creation of pseudo-character is a possible way to separate semantics with orthography by substituting the semantic radical with another radical, which has no semantic overlap with the original radical. Therefore, instead of using real characters, the present study used pseudo-characters to realise that idea. For example, the character 猴 means *monkey*, while the radical 犭 (4 strokes) indicates *animal*. If 犭 is replaced by 土 (4 strokes, means *soil*), the pseudo-character becomes 堠 , and it is still assigned the meaning of *monkey* in the experiment. In this way, a pseudo-character without a radical suggesting the meaning of animal is created, while the number of strokes stays the same.

The aim of the study reported in this chapter was to further examine the effect of semantic grouping on learning over a longer time-course. Specifically, we hoped to investigate the building of individual lexical representations. It is clear in Experiment 1 that participants required more learning input to build such representations, especially the recognition of Chinese sounds (less than 30%). As indicated by Gaskell and Dumay's study (2003), the integration of the newly learnt words into lexical competition happened over five days. Following the studies conducted by Gaskell's team (Dumay & Gaskell, 2012; Tamminen et al., 2010), the formation of lexical representation was therefore tested over a seven-day time span incorporating extra learning and testing sessions on the second and the third day.

Another key change from Experiment 1 is that the timing of the learning sessions was fixed instead of self-paced. Based on the result of the Experiment 1, it suggested that the viewing time and the number of iterations of each set did not significantly vary across participants. In order to keep consistent to Experiment 1 and allow further comparison of results between Experiment 3 and 1, we decided to fix the viewing time and number of iterations across all participants. The uniformed learning input also enables multiple participants to be tested at one time. According the result of learning session of Experiment 1, the average viewing time for each character was approximately 17 s while the mean number of iterations was two. Therefore, in the current study we present each character for 9 s while each set was shown twice, which means the viewing time for each character was 18 s.

Two additional tasks were included in order to provide a direct test of effects of grouping on the development of categorical and identity representations. Categorical relationships were evaluated using a categorisation task in which participants were asked to identify whether the two pseudo-characters appearing on the screen belonged to the same semantic category or not. This task occurred on the second and seventh day to measure effects of short-term as well as long-term learning. In addition, a priming test was performed on the seventh day to test identity learning. In this task, a target lexical judgement task which was identical to the testing task was primed by a pseudo-character. The prime could be either semantically related, semantically unrelated or the target itself. Therefore, we hypothesise that in the time span of seven days, a developmental change could be observed. In the testing tasks, learning in semantic groups could help to reach higher accuracy or shorter latency compared to the performance of ungrouped learners. A competition effect could be seen during the priming task. We also would like to discover further evidence of the semantic interference in the categorisation task for further evidence of the well formation of the link from orthography to semantics.

Methods

Participants

Participants were recruited through the Research Participation Scheme (RPS) from the School of Psychology at the University of Birmingham. There were 34 participants in this study. Among them, 33 were female and one was male. They were all native British English speakers with normal or corrected to normal vision. No one reported any language impairments such as aphasia or dyslexia. No participants had any prior learning experience of East Asian languages, such as Chinese, Japanese, and Korean.

Materials

Twenty pseudo-characters were all created by replacing these semantic radicals of items in the character list of Experiment 1 (see Table 3.1 in Chapter 3, p. 46). For a certain character, the semantic radical was substituted with another real semantic radical with the same number of strokes. Each replacement of semantic radicals was orthographically unique and visually distinct from each other. All radicals used for replacement appeared in the position where they form a real character. There were in total 20 substitution radicals to ensure that no pseudo-characters shared a semantic radical. A full list of pseudo-characters can be found in Table 4.1 (p. 99).

Each pseudo-character was assigned the meaning and pronunciation of the character from which it originated, which were all monosyllabic Chinese words: high frequency, simple concrete nouns. The number of strokes was the same as for the real characters tested in Experiment 1 of Chapter 3. The pseudo-characters were categorised into the same five semantic groups. In addition, two characters, 洪 (flood) and 腹 (belly) were replaced by 海 (sea) and 脸 (face) respectively. Both were more frequently used than the original in both languages.

There was no significant difference between sets within the character list with regard to number of strokes, $F(4, 15) = 1.88$, $p = .17$. There was no significant difference in the frequency of English translation, $F(4, 15) = 0.91$, $p = .48$. The source of word frequencies was identical to Experiment 1: N-Watch (Version 2007.10.1) for English frequency.

Table 4.1

The Semantic Grouping Pseudo-Character List of Experiment 3.

Semantic radical		Chinese Character		Pronunciation			
Original	Substitutional	Original	Pseudo	Pinyin	IPA	Stroke number	English Translation
犭	大	狗	⿰火句	gǒu	[gou˨˩˦]	8	*dog*
	巾	猫	⿰巾苗	māo	[mɑu˥]	11	*cat*
	彳	猪	⿰彳者	zhū	[tʂu˥]	11	*pig*
	土	猴	堠	hóu	[xou˧˥]	12	*monkey*
氵	女	瀑	⿰女暴	pù	[phu˥˩]	18	*waterfall*
	山	湖	⿰山胡	hú	[xu˧˥]	12	*lake*
	扌	海	挴	hǎi	[xai˨˩˦]	10	*sea*
	马	溪	⿰马奚	xī	[ɕi˥]	9	*brook*
木	王	林	⿰王木	lín	[lin˧˥]	8	*forest*
	车	树	⿰车对	shù	[ʂu˥˩]	9	*tree*
	方	枝	⿰方支	zhī	[tʂɿ˥]	8	*branch*
	月	松	⿰月公	sōng	[sʊŋ˥]	8	*pine*
月	贝	腿	⿰贝退	tuǐ	[thueɪ˨˩˦]	13	*leg*
	牛	胸	⿰牛匈	xiōng	[ɕyʊŋ˥]	10	*chest*
	火	腰	⿰火要	yāo	[jɑu˥]	13	*waist*
	礻	脸	⿰礻佥	liǎn	[liɛn˨˩˦]	11	*face*
衤	立	衫	⿰立彡	shān	[ʂän˥]	8	*shirt*
	石	裤	⿰石库	kù	[khu˥˩]	12	*trousers*
	目	袜	眜	wà	[wä˥˩]	10	*socks*
	田	裙	⿰田君	qún	[tɕhyn˧˥]	12	*skirt*

We used the same method and settings for creating visual stimuli as we did in Experiment 1 of Chapter 3 with Adobe Photoshop (Version 12.0). The same settings of each stimuli applied, which was 530 kB, 425 * 425 pixels for Chinese and 150 kB, 425 * 122 pixels for English. They were in the font of Kai for Chinese and Times New Roman for English.

As in Experiment 1, the reading app TTSApp was used to generate sound files for the Chinese characters. To improve the sound quality compared to the stimuli of Experiment 1, a more dedicated audio base VW Hui (Chinese) was used instead of Microsoft Lily (Chinese). In order to control the exposure duration for all stimuli, a uniform length for each individual sound file was required. The length of each single sound file was set to 1000 ms. The unified length was realised by either compressing or stretching the sound file and then smoothing it using MATLAB (Version 8.1). Care was taken to minimise any loss of sound quality by smoothing the noise generated from compression or stretch. A sound file for each stimulus was then created comprising six iterations with 500 ms intervals between iterations: 9 s total duration. As each stimulus was presented twice, the total exposure time for each pseudo-character was 18 s, the longer version was created with the programme Cool Edit Pro (Version 2.0).

Apparatus

This study used E-Prime (Version 2.0) for programming, carrying out the experiment, and collecting the data. The whole experiment was delivered in a sound attenuated room. Up to four participants could take part in the experiment at the same time. They were allocated to separate cubicles. Visual stimuli were presented through monitors. Auditory stimuli were played by headphones.

Design and procedure

The experiment consisted of three learning sessions, five testing sessions, two categorisation tasks and one priming task across seven days. The structure of the experiment is illustrated in Figure 4.1. Detailed descriptions of the individual task components follow this overview of the structure of the study.

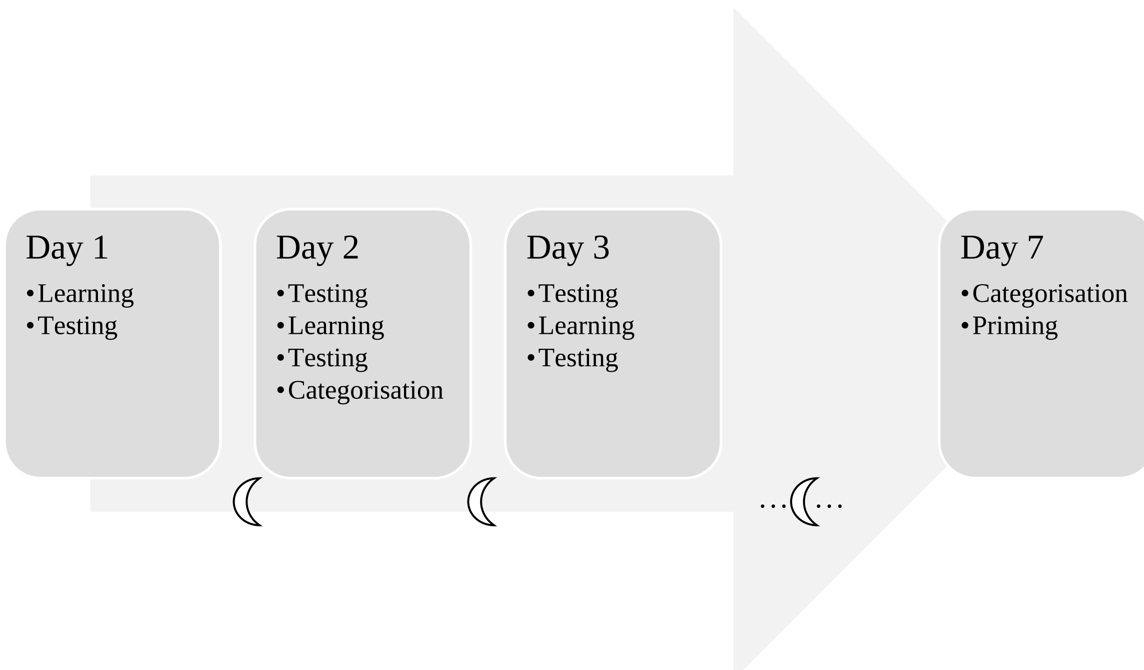


Figure 4.1. Schematic diagram of the structure of Experiment 3.

On Day 1, participants were briefed about the purpose, procedures, and basic operations of the experimental programme. A detailed set of written instructions, the consent form and questionnaire about physical and linguistic background were presented to these participants following the briefing. The consent form and questionnaire were the same as those used in

Experiment 1. The consent form, questionnaire and instructions are listed in Appendices A, B and E respectively (pp. 211-217). After the signed form and the questionnaire were checked, participants started the first learning session. With a short break after the learning session up to 3 min, they had their first testing session. After the testing session, they were asked to come back in 24 hr.

On Day 2, participants firstly completed a testing session in which the order of trials differed. Then they learnt the pseudo-characters again in a different order. After the learning session, they were tested again. The last task on Day 2 was a categorisation task. They were required to return in 24 hr once they had completed all the tasks on Day 2. On the third day, they had a testing session, a learning session, and another testing session. There was a 96-hr gap between Day 3 and Day 7. On the final day, they were asked to perform a categorisation task and a priming task. For a given participant, all tasks occurred with a different random trial order. Details about the randomisation procedures are specified in the following sections.

The whole set of experiments took around 200 min to complete in four days across a week. Participants were debriefed at the end of the experiment and told that what they learnt were pseudo-characters. The procedures for each task are described in detail below.

Learning phase.

Design.

The learning phase took place on Days 1, 2 and 3. For the learning of the 20 pseudo-characters, *Grouping*, was a between subject factor with two levels: grouped and ungrouped. One group of participants learnt the 20 pseudo-characters in semantic groups while the other group learnt them in ungrouped sets.

The grouped list was formed of five sets of pseudo-characters, which shared a semantic category, i.e., there were four pseudo-characters in each set. Due to the replacement of semantic radicals, items in a set shared only a semantic relation without any repetition of orthographic features. Similar to Experiment 1, the ungrouped learning version was created by distributing the same 20 pseudo-characters into five unrelated sets without shared semantics between the items in each set. Randomisation of sets was realised by Latin square method. The presentation orders of items in each set was done by E-Prime. The number of iterations of each set and the viewing time of a certain pseudo-character were all fixed within and across groups.

Procedure.

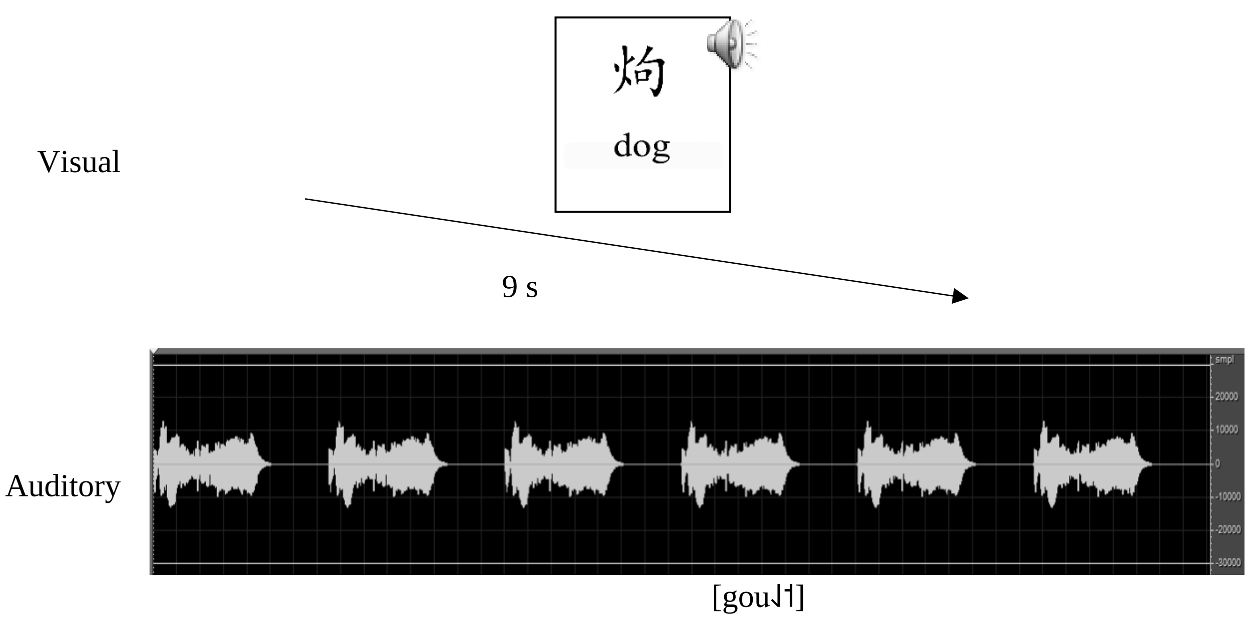


Figure 4.2. Schematic diagram of a single presentation of a pseudo-character during learning session of Experiment 3, where 炣 originates from the character 狗, which means *dog*.

Participant were instructed to try to learn the relationship of the pseudo-character with the pronunciation and the English translation. As they did not need to produce the sound during the test, they were asked not to read aloud or silently after hearing the sound. Every participant saw each pseudo-character and its English meaning for 9 s while hearing the Chinese sound repeated six times with an interval of 500 ms between iterations during a single exposure (see Figure 4.2, p. 103). Before moving into the next set, they were presented with the same set again in a different random order.

Testing phase.

Design.

The testing phase aimed to evaluate the learning outcome of individual pseudo-characters for their meaning and pronunciation. It consisted of five sessions across the study. The first followed the learning session on Day 1. Days 2 and 3 both had one testing session before and after the learning session. Each testing session was formed of a series of lexical matching judgement tasks.

The design of the task was identical to the testing session of Experiment 1, except that those pairings were pseudo-characters with Chinese sounds or English words. There were matching and mismatching trials too. Mismatching items were formed with a sound or an English word from outside of the semantic set but from the character list. Similarly, there were four different combinations for each character:

Pseudo-Chinese character + the assigned English translation;

Pseudo-Chinese character + the assigned Chinese pronunciation;

Pseudo-Chinese character + another English word which was not the assigned one;

Pseudo-Chinese character + another Chinese sound which was not the assigned one.

These combinations are illustrated in Figure 4.3 below.

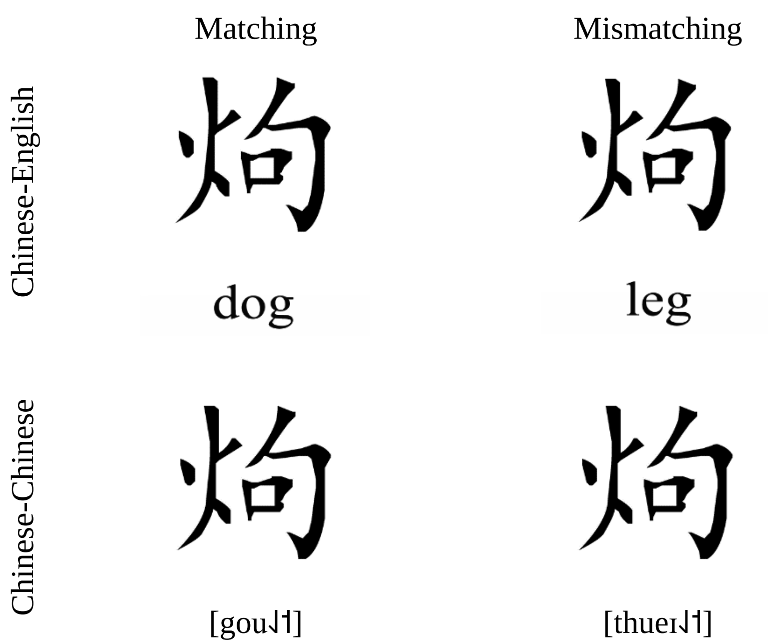


Figure 4.3. Design of the testing phase stimuli of Experiment 3, where ⿰火句 originates from the character 狗, which means *dog*; and [t^{h}ueɪ˩˥] is the pronunciation of the character 腿 (the origin of the pseudo-character ⿰足退), which means *leg*.

Testing sessions of Experiment 3 also contained 160 trials. We followed the same logic of blocking trials as in Experiment 1. One pseudo-character occurred once in matching condition and once in mismatching condition in a pair-block. There were equal number of Chinese-Chinese (C-C) or Chinese-English (C-E) pairings and the same number of matching and mismatching trials in each pair-block. There were identical number of trials in each condition. The order of trials in a pair-block was randomised with the constraint that there were at most four trials of one Pairing condition (C-C or C-E) or one Matching condition (matching or mismatching) in a row. The order of presentation of pair-blocks was counterbalanced by Latin square method. When being presented to participants, each pair-block was divided into two blocks of 20 trials.

Each participant completed five testing sessions over the course of the experiment. Therefore, five unique randomisations of trial order were prepared to eliminate any position effects on a given trial. This was accomplished by rotating the order of four pair-blocks as well as creating different randomisation of trials within each pair-block such that no trial order was repeated in any pair-block across these five versions. The same set of testing programmes was used for both grouped and ungrouped learners.

Procedure.

Participants were briefed about their task before the testing phase started. Before the formal testing began, instructions were given on the screen with a practice session included. The purpose of the practice session was to help participants to familiarise themselves with the timing of testing trials. Pseudo-characters were substituted by images of objects while sounds of Chinese characters were replaced by English pronunciations. The aim was to prevent repetition effects of items in the test. Similar to the formal testing, the task was to judge the image match with the English word or sound or not. In total, there were 10 trials in the practice session, which included all the four combinations of stimuli described above. After the practice, the formal test began after a short break.

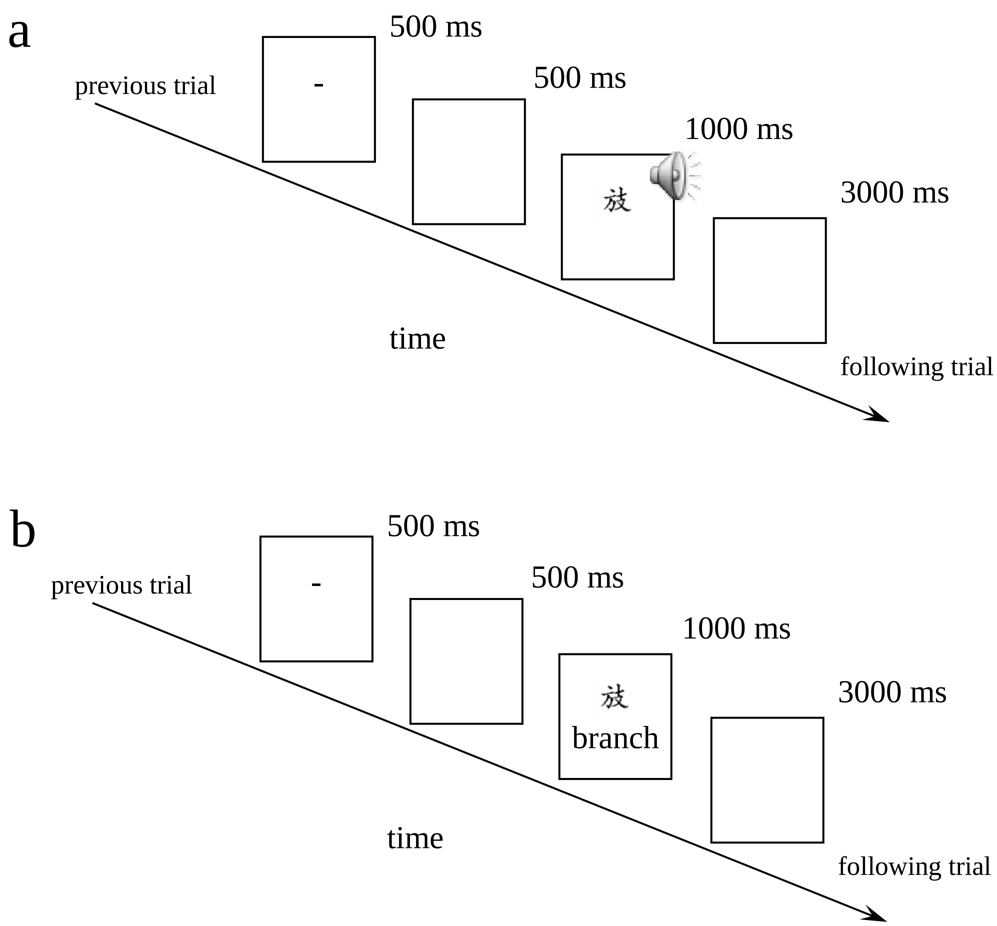


Figure 4.4. Timing of experiment trials of Experiment 3. (a) a Chinese-Chinese trial. (b) a Chinese-English trial.

During a *Chinese-Chinese* trial, participants were asked to judge whether the pairing of a pseudo-character and a Chinese sound was correct or not. The trial began with a dash (-) as the fixation for 500 ms followed by a 500 ms blank screen. Then a pseudo-character appeared in the centre of the screen for 1000 ms with a sound file of a Chinese word played concurrently with it once only, followed by a fixed interval of 3000 ms. From the time when the pseudo-character firstly appeared, participants were required to press Yes or No buttons within 4000 ms. That design aimed to force participants to make a decision as quickly as possible without overthinking. The following trial started automatically after the 3000 ms blank. Figure 4.4a shows the time-course of a typical C-C trial.

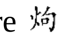
A similar time-course occurred for the *Chinese-English* trials (see Figure 4.4b, p. 107), except that a pseudo-character was displayed together with an English word below it for 1000 ms without any sound.

A gap of 10.5 s was available for participants between blocks. For testing sessions on other days, the procedures were identical but without briefing. The practice session was still included as a warm-up.

Categorisation test.

Design.

There were two categorisation tests, occurring on Day 2 and Day 7 of the experiment to examine the result of short-term and long-term category learning respectively. The design of the categorisation phases was the same for both groups of participants. The categorisation test involved paired judgement tasks. On each trial participant saw two pseudo-characters and were asked to decide whether they belonged to the same semantic category or not. The two pseudo-characters could be either semantically related (from the same set) or semantically unrelated (from a different set). Examples are shown below (see Figure 4.5, p. 109).

Figure 4.5. Design of the categorisation test stimuli of Experiment 3, where  originates from the character 狗, which means *dog*;  originates from the character 猫, which means *cat*; and  originates from the character 枝, which means *branch*.

There were 120 trials in each categorisation session. They were organised into six blocks of 20 trials. Every pseudo-character occurred 12 times during the test: six times in semantic related pairings, six times in semantic unrelated pairings. There were four testing versions. Each occurred with a different randomisation of the testing order. The maximum number of consecutive trials with two pseudo-characters of the same semantic relation (related or unrelated) was restricted to four. The number of times a certain character appeared on the left position or the right position was identical.

In order to eliminate trial order effects, we created four different randomisations. That was realised by creating four random order of all trials and then dividing the 120 trials into six blocks. However, each participant only went through any two of the four randomisations.

Procedure.

The categorisation task had similar procedures to the testing phase. There was a short briefing about the categorisation task. Following the briefing, participants read the on-screen instructions. The following a 10-trial practice session was a chance for participants to become familiar with the timing of the task, with stimuli replaced by English words. Participants were asked to judge whether the English words rhyme or not. The formal testing started after the practice.

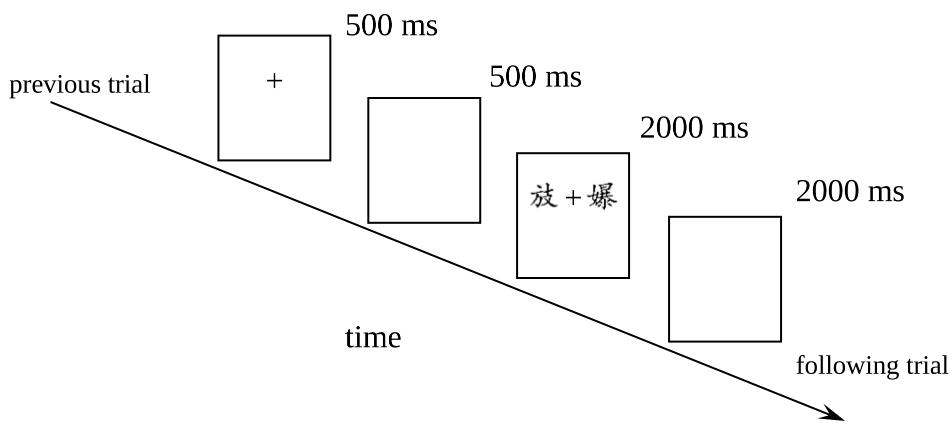


Figure 4.6. Timing of a categorisation trial of Experiment 3, where [illegible] originates from [illegible], which means *branch*; while [illegible] originates from [illegible], which means *waterfall*.

Figure 4.6 illustrates the sequence of events in a categorisation trial. A fixation (+) was presented in the centre on the screen for 500 ms followed by a 500 ms interval. After that two pseudo-characters were shown together horizontally with the fixation in the middle of the screen. One pseudo-character was on the left-hand side and the other one on the right-hand side for 2000 ms. Once they disappeared, a 2000 ms blank followed. Participants were asked to press the Yes or No button to make a judgement during the presentation of the stimuli and the blank, the length of which was fixed to 4000 ms in total. The next trial resumed after the blank. Similar to the testing session, participant could take a break of up to 10.5 s between blocks. Response error rate and reaction time were automatically recorded by E-Prime.

Priming test.

Design.

The priming phase was designated to evaluate the quality of these representations that participants had developed of the characters during the experiment. It occurred only at the end of Day 7. The structure of a priming test trial resembled the testing trial. However, the target pseudo-character was primed by another pseudo-character. The priming pseudo-character could be the same pseudo-character, a semantically related pseudo-character, or a semantically unrelated character. In order to balance the number of prime types, four additional filler conditions were introduced. The judgement task still required the matching of target pseudo-characters with other items (English words or Chinese sounds). Together there were 16 conditions as illustrated in Figure 4.7 (p. 112).

There were 320 trials in each test run. They were divided into four pair-blocks, which were designed to prevent participants from anticipating trials or the design pattern of the experiment. Every pseudo-character appeared in 16 times as a prime as well as 16 times as a target. Each pair-block had 60 experiment trials and 20 semantically unrelated fillers. Half of the 80 trials and fillers were matching while the other half were mismatching. There were four different randomisations of trial order. A Latin square method was used to counterbalance the order of pair-blocks. The sequence of trials in each pair-block was distinct across the four randomisations. No more than four trials of the same conditions or of the same type of tasks (i.e. C-E or C-C) could appear consecutively.

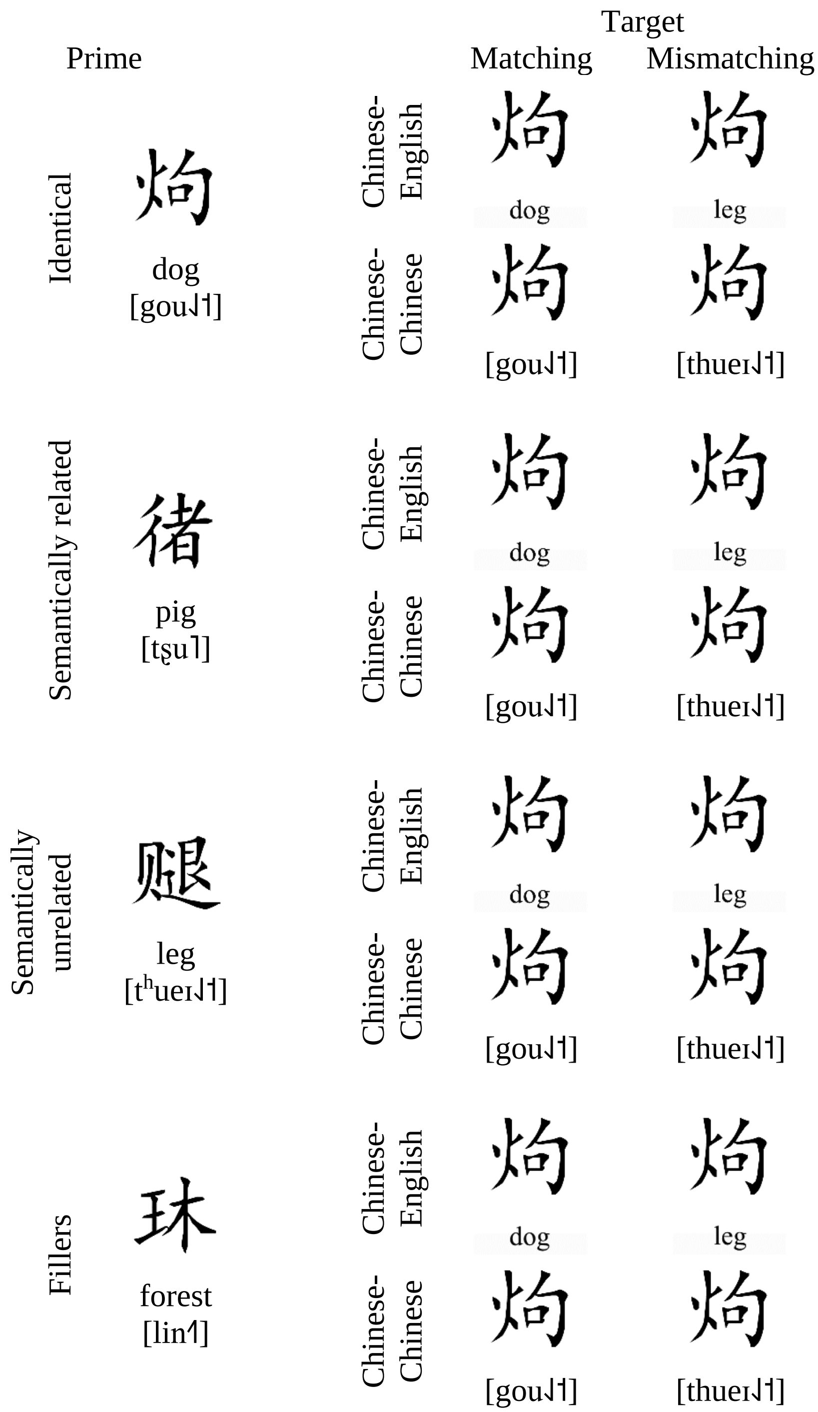


Figure 4.7. Examples of priming conditions in Experiment 3, where [illegible] originates from the character 狗, which means *dog*; [illegible] originates from the character 猪, which means *pig*; [illegible] originates from the character 腿, which means *leg*; and [illegible] originates from the character 林, which means *forest*.

The maximum number of the same response (i.e., Yes or No) shown consecutively was four. During the experiment, each pair block was split into two blocks of 40 trials.

In this task, we chose a longer SOA of 1500 ms for the prime. As indicated by Liu et al.'s study (2007), a longer SOA would enable potential priming effect by allowing more time to process the prime. In their study, they chose a SOA of 500 ms to observe the change of orthographic threshold for L2 Chinese characters learners across two semesters. It could be inferred from results of their study that the improvement of proficiency could shorten the threshold time for orthographic processing. However, in the present study only three learning sessions was presented and therefore a much longer SOA, 1500 ms in this case, was used as the start point.

Procedure.

Similar to other tasks of the experiment, participants were briefed about the procedure of the task. They then read instructions presented on the screen and completed a 10-trial practice. The practice mimicked the real trials with the pseudo-characters replaced by images of concrete objects and Chinese pronunciation by sounds of English words. Once they finished the practice, the priming test began, which comprised eight blocks of 40 trials.

Figure 4.8 (p. 114) illustrates the timing of events during a priming trial. It started with a fixation (-) for 500 ms, followed by a blank for 500 ms. The prime then appeared on the screen for 1000 ms followed by another 500 ms blank. After that, a matching task was presented on the screen (and played through the headphone for trials of C-C pairings) for 1000 ms. During the 1000 ms of presentation and the following 3000 ms blank, participants decided whether the pairing was correct or incorrect and respond with an appropriate button press. The next trial began after the blank.

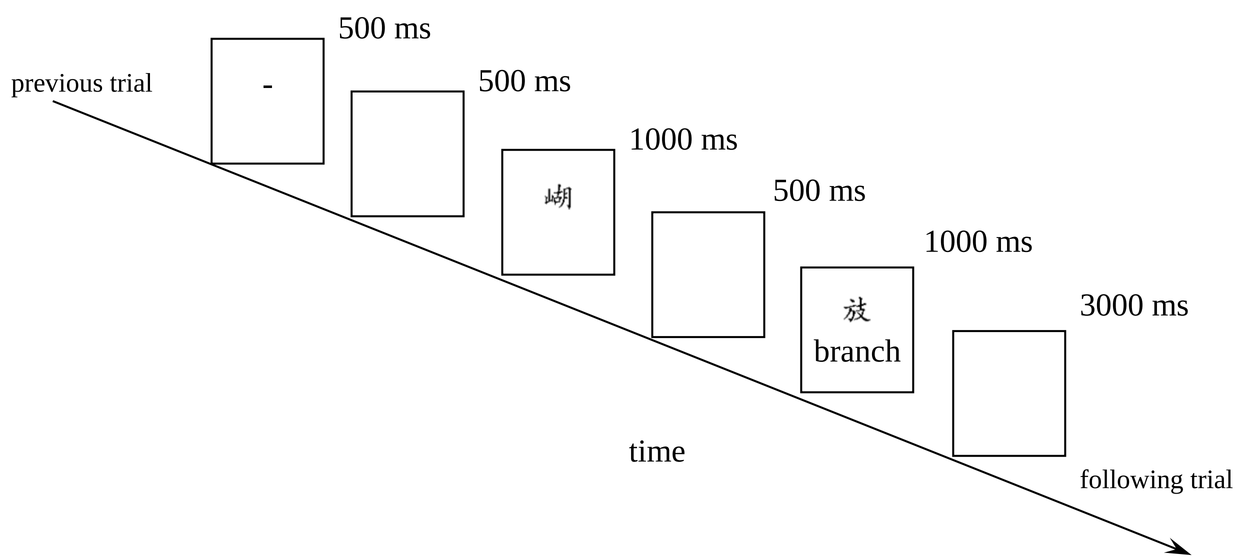


Figure 4.8. Timing of a priming task of Experiment 3, where 𡹬 originates from 湖, which means *lake*; while 𣎴 originates from 枝, which means *branch*.

Results

Data analysis procedure

The data were analysed using mixed models in R. Reaction time and error rate were analysed separately for all tasks of the experiment. Reaction time data were analysed using mixed effects linear models (Barr, Levy, Scheepers, & Tily, 2013). The data from trials with incorrect responses were excluded from reaction time analyses, namely 8,562 out of 31,200 (27%) for testing tasks, 2,866 out of 9,360 (31%) trials for categorisation tasks, and 2,012 out of 9,120

(22%) for priming task. For the analysis of error rate data, we used mixed-effect logit models (Barr et al., 2013; Jaeger, 2008).

There were many factors which were manipulated in the experiment. They will be specified in the next few paragraphs.

In this experiment, *Grouping* was the only predictor that affected all the tasks performed. It referred to the learning method, either learning in semantic groups or learning in a randomised group. It was also the only between subject manipulation in the experiment.

For the testing task, the following factors were motivated from the design. *Learning* was the number of times that the participants took the learning session, which could be one, two or three. *Sleep* meant the task was taken place *immediately after learning* (of each day) or *after sleep*. There were two more factors for the design of trials: *Pairing* (*Chinese-Chinese* or *Chinese-English*) and *Matching* (*matching* or *mismatching*).

For the categorisation task there were two other factors besides Grouping: *Day* (Day 2 or Day 7) and *Category* (*semantically related* or *semantically unrelated*). While for the priming task, *Priming* was the type of the prime, which could be *identical to the target*, *semantically related* and *semantically unrelated*. The priming task also contained the factors of *Matching* and *Pairing*, which were both identical to the testing task.

All the factors above were deviation coded except for *Learning* and *Priming*, which were both dummy coded. For models of all tasks, the start point was a prototypical model which contained all factors and a fully specified random effect structure. Every model went through a similar model comparison process for fixed effects. In this study, we used the backward model selection to compare a model and the one which was one step more simplified until the *p* value was significant. We followed the same rules for reducing random effect structure as in

Experiment 1 (p. 57). In the following sections, main models are summarised in tables. Random effects are presented in a table note along with the results.

Results of the Testing Tasks

The dependent variables of the testing phase are error rate and reaction times. Results of these variables are reported in separate sections below.

Error rate.

The error rate data for all trials are presented in Figures 4.9 and 4.10 and Table 4.2. The variable Learning was dummy coded with the reference level set as Learning 1.

In the full model, as two of 3 two-way interactions containing matching are significant: Pairing by Matching (p = .009), and Learning 2 vs.1 by Matching (p = .04). We therefore split the data by Matching and carried out analyses separately. Each subset was analysed with a model which contained all the fixed effects of the model for all trials, while the fixed effect and interactions including Matching were ruled out from the subset model. In order to keep consistency of the data subsets, the fixed effects structure for matching and mismatching trials stayed the same while the random effect structure differed in order to reach convergence for each model respectively.

All participants were more error prone after sleep while grouped learners tended to be less error prone than ungrouped learners.

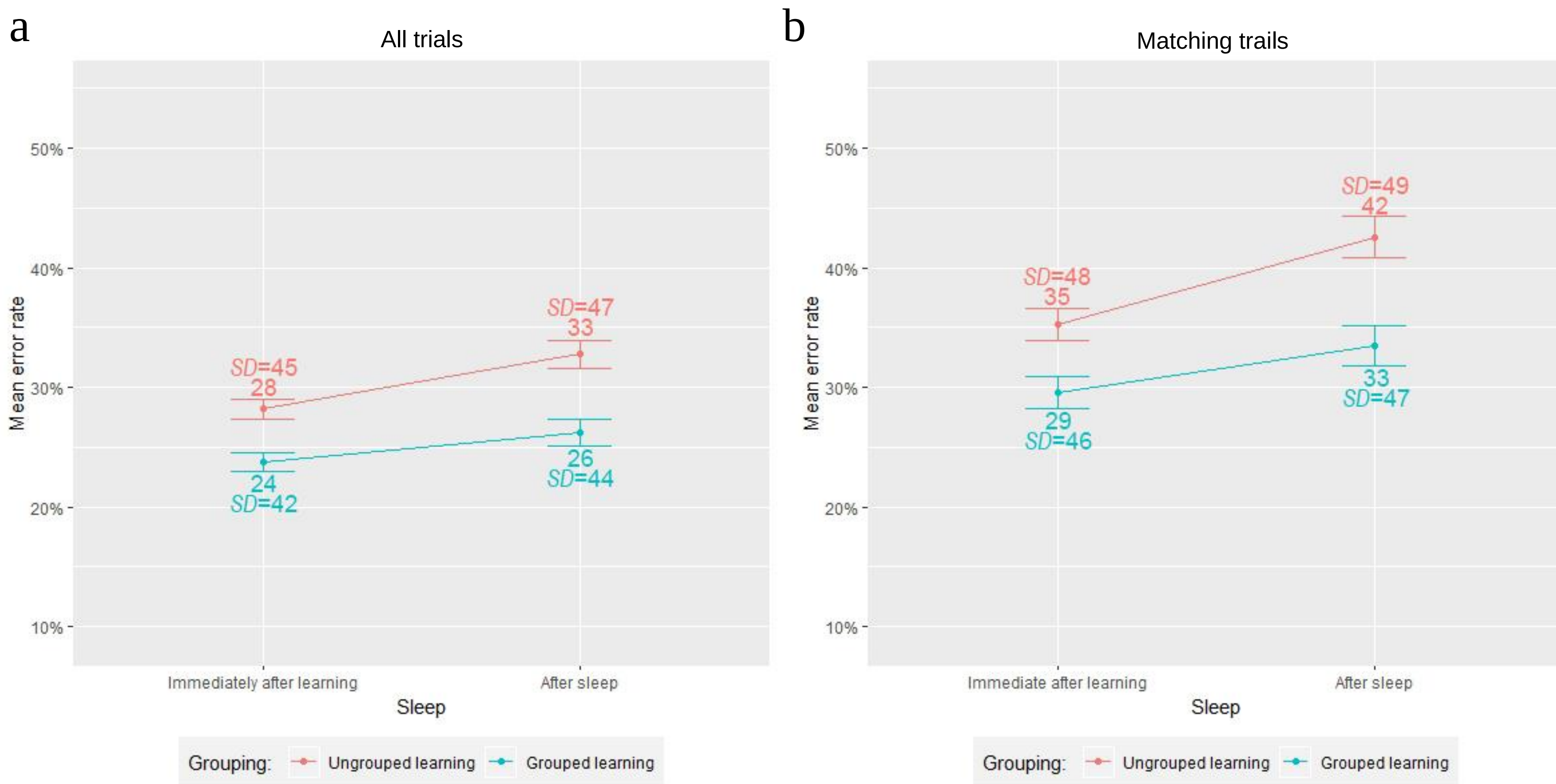


Figure 4.9. Results of error rates in terms of Grouping by Sleep (a) of all trials, and (b) of matching trials, Testing tasks, Experiment 3.

Error rates increased after sleep overall, as can be seen by the significant main effect of Sleep (p = .04). The two-way interaction of Grouping by Sleep reached significance (p = .04, see Figure 4.9a, p. 117). Similar to the pattern found in the complete data set, there was also a significant two-way interaction Grouping by Sleep among matching trials (p = .03, see Figure 4.9b, p. 117). However, no further significant difference was found through post hoc Tukey tests. However, it seemed that for the overall data, after sleep, the increase in error rate for the ungrouped learners was significantly larger than the increase for the grouped learners, while the mean error rate of the matching trials was higher than the overall data. Therefore, it could be inferred that after sleep, the difference between grouped learners and ungrouped learners was enlarged.

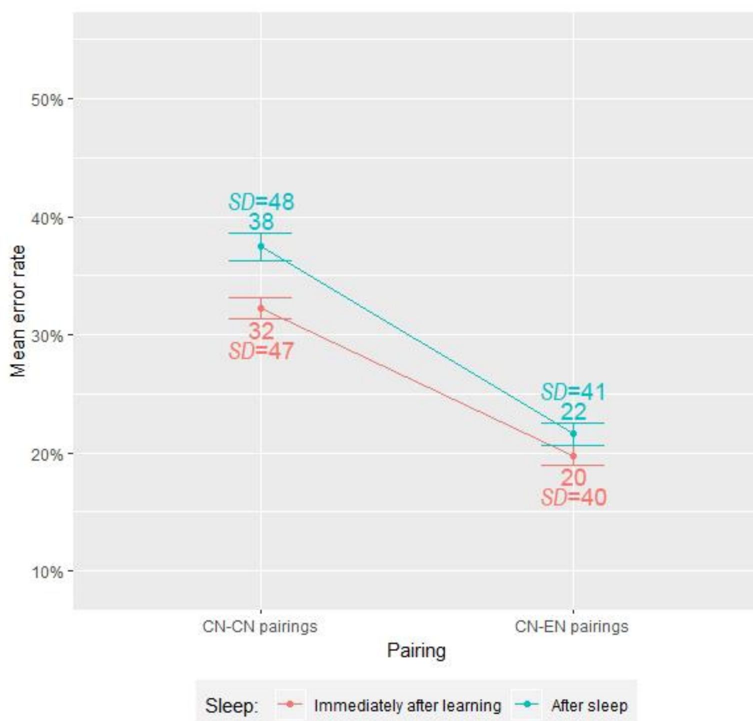


Figure 4.10. Results of the interaction Sleep by Pairing on error rates of testing tasks, Experiment 3.

According to Figure 4.10 (p. 118), after sleep, the error rate for C-C pairings and C-E pairings all increased. The increase was larger for C-C pairings was significantly larger than C-E pairings in the main model (p = .02).

A significant drop in error rates for matching C-E trials on Day 2, while a similar drop for matching C-C trials occurred on Day 3 among all participants.

Error rates on Day 2 vs. Day 1

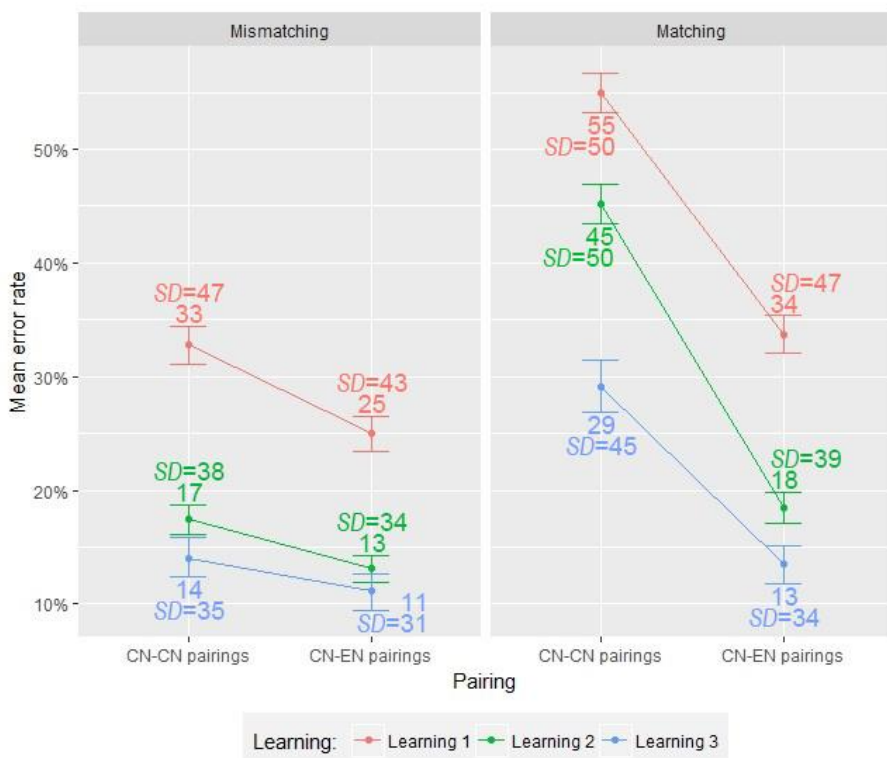


Figure 4.11. Results of the effect of Learning by Pairing on error rates of testing tasks, Experiment 3.

Figure 4.11 (p. 119) summarises the decline in error rates across three learning sessions regarding different Matching and Pairing conditions.

In the full model, we found a significant three-way interaction Pairing by Learning 2 vs. 1 by Matching (p < .001). Based from results of post hoc Tukey tests, although Pairing significantly interacted with Learning 2 vs.1 for all trials (p < .001), there was only insignificant difference between decrease in error rates of C-C pairings and that of C-E pairings for mismatching trials. However, for matching trials, we observed a significant drop in error rate among C-E pairings at p < .001 which was significantly larger than the decrease among C-C pairings at p < .01. That was confirmed by the interaction Pairing by Learning 2 vs. 1 which was only significant among matching trials (p < .001).

Error rates on Day 3 vs. Day 1

Learning 3 vs. 1 significantly interacted with Pairing for all trials (p = .05) and also for matching trials (p = .04). There was a significant main effect of Learning 3 vs. 1 (p < .001). Post hoc Tukey's test of the interaction Pairing by Learning among matching trails showed that the significant drop error rates was observed for matching C-C pairings on Day 3 at p < .001. The effects of Learning and Pairing above indicated participants required more time to form representations of Chinese sounds than linking existing semantic representations with Chinese characters.

Moreover, from Learning 1 to Learning 2, there was a significant drop in error rate overall, which is confirmed by the main effect of Learning 2 vs. 1 for all trials (p < .001), matching trials (p < .001) and mismatching trials (p < .001). The main effect of Learning 3 vs. 1

was found among all trials, matching and mismatching trials (*p* value of which were all < .001). As one would expect therefore, more learning input resulted in higher accuracy.

Matching trials had significantly higher error rates than mismatching trials ($p < .001$). In addition, participants were more accurate for C-E pairings than C-C pairings, which applied to all trials ($p < .001$), matching trials ($p < .001$), as well as mismatching trials ($p = .005$).

Table 4.2

Summary for Fixed Effects in the Mixed Logit Model for Error Rates in the Testing Tasks of Experiment 3.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-0.68	0.10	-6.46	< .001***
Pairing	0.40	0.07	5.46	< .001***
Learning 2 vs. 1	-0.99	0.10	-9.24	< .001***
Learning 3 vs. 1	-1.67	0.18	-8.94	< .001***
Grouping	0.10	0.07	1.47	.14
Matching	-0.38	0.08	-4.76	< .001***
Sleep	0.05	0.02	2.08	.04*
Pairing by Learning 2 vs. 1	0.25	0.06	3.88	< .001***
Pairing by Learning 3 vs. 1	0.21	0.11	1.96	.05*
Pairing by Grouping	-0.01	0.05	-0.22	.82
Grouping by Matching	-0.06	0.05	-1.18	.24
Pairing by Matching	-0.12	0.05	-2.62	.009**
Learning 2 vs. 1 by Matching	-0.15	0.08	-2.04	.04*
Learning 3 vs. 1 by Matching	0.11	0.10	1.18	.24
Grouping by Sleep	-0.03	0.02	-2.02	.04*
Pairing by Sleep	-0.04	0.02	-2.25	.02*
Pairing by Grouping by Matching	0.04	0.04	1.00	.32
Pairing by Learning 2 vs. 1 by Matching	-0.12	0.04	-3.40	< .001***
Pairing by Learning 3 vs. 1 by Matching	-0.04	0.05	-0.89	.38
Model for matching trials[b]				
Intercept	-0.30	0.14	-2.09	.04*
Grouping	0.12	0.08	1.53	.13
Pairing	0.51	0.08	6.73	< .001***

Learning 2 vs. 1	-0.85	0.13	-6.31	< .001***
Learning 3 vs. 1	-1.79	0.24	-7.40	< .001***
Sleep	0.01	0.03	0.46	.65
Pairing by Learning 2 vs. 1	0.36	0.05	6.41	< .001***
Pairing by Learning 3 vs. 1	0.23	0.11	2.08	.04*
Grouping by Sleep	-0.05	0.02	-2.31	.02*
Model for mismatching trials[c]				
Intercept	-1.07	0.12	-8.69	< .001***
Grouping	0.10	0.09	1.16	.25
Pairing	0.30	0.11	2.82	.005**
Learning 2 vs. 1	-1.15	0.13	-8.79	< .001***
Learning 3 vs. 1	-1.65	0.20	-8.15	< .001***
Sleep	0.08	0.03	2.40	.02*
Pairing by Learning 2 vs. 1	0.10	0.09	1.14	.26
Pairing by Learning 3 vs. 1	0.15	0.17	0.89	.38
Grouping by Sleep	-0.01	0.02	-0.60	.55

Notes. [a]*N* = 31,200, log-likehood = -14,712.9. The random effect structure included an intercept and slopes of Grouping, Matching, Sleep, Pairing by Learning, and Learning by Matching for items, and an intercept and slopes of Sleep, Pairing by Learning, Pairing by Matching, and Learning by Matching for subjects, which is the maximal random effect structure that reached convergence.

[b]*N* = 15,600, log-likehood = -7,985.4. For matching trials, the random effect structure contained a random intercept for items and subjects. It also included the random slopes of Grouping, Pairing, Learning, and Sleep for items and random slopes of Pairing, Learning and Sleep for subjects, which was the maximal random effect structure that converged.

[c]*N* = 15,600, log-likehood = -6,693.1. For mismatching trials, the random intercept and slopes of Sleep, Grouping and the interaction Pairing by Learning for items and the random intercept and slopes of Sleep and the interaction Pairing by Learning for subjects were included in the random effect structure. This was the maximal random effect structure to converge.

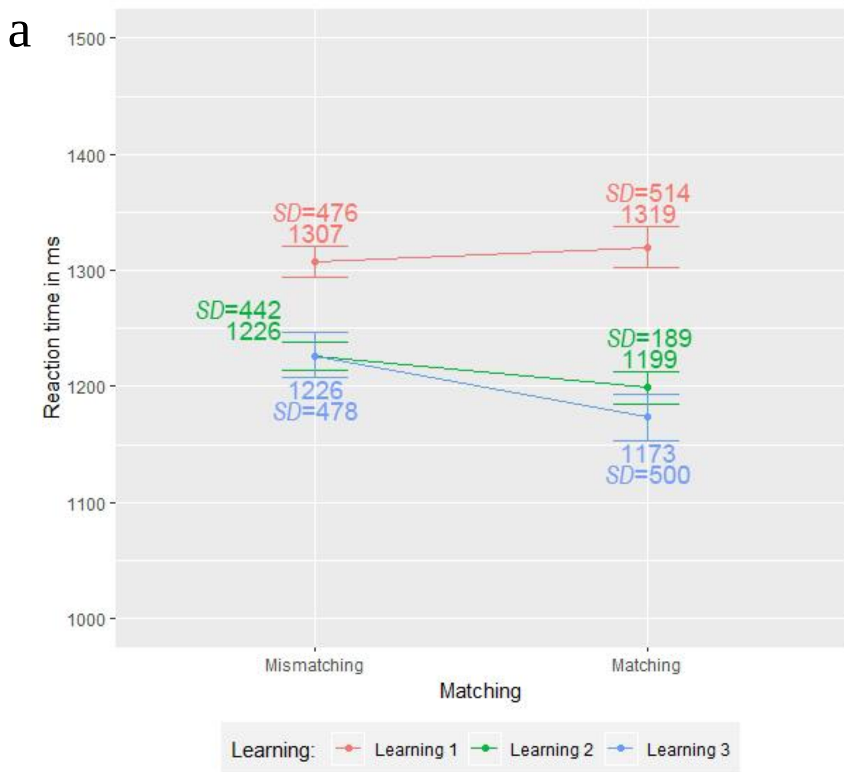
*$p < .05$. **$p < .01$. ***$p < .001$.

Reaction time.

Figures 4.12, 4.13 and Table 4.3 summarise the reaction time data for all correct trials. The independent variable of Learning was dummy coded where the reference level is Learning 1.

Learning input helped to reduce the reaction time, especially after the second learning input.

Reaction times decreased from Learning 1 to Learning 3, showing that participants were faster to response for both kinds of matching trials after more learning ($p < .001$ for both Learning 2 vs. 1 and Learning 3 vs. 1), as seen from Figure 4.12a (p. 124). Based on post hoc Tukey HSD tests for the interaction of Pairing by Matching, participants were slower to respond to matching trials than mismatching trials at $p < .01$. However, after the second learning session, the reaction time for matching and mismatching did not differ much. This trend still existed on Day 3. Note that there was a difference from the situation presented in the picture. According to Figure 4.12a, originally participants were slower to respond to matching trials than mismatching trials. However, after Learning 2, the pattern reversed with the significant drop in reaction time found in matching trials. Then, on Day 3, the reaction time for mismatching trials stayed while the reaction time for matching trials dropped further. Difference could be attributed to the unbalanced nature of the present study. The least square means were the means which were adjusted according to other factor in the model, while the arithmetical means presented in the figures were not adjusted.



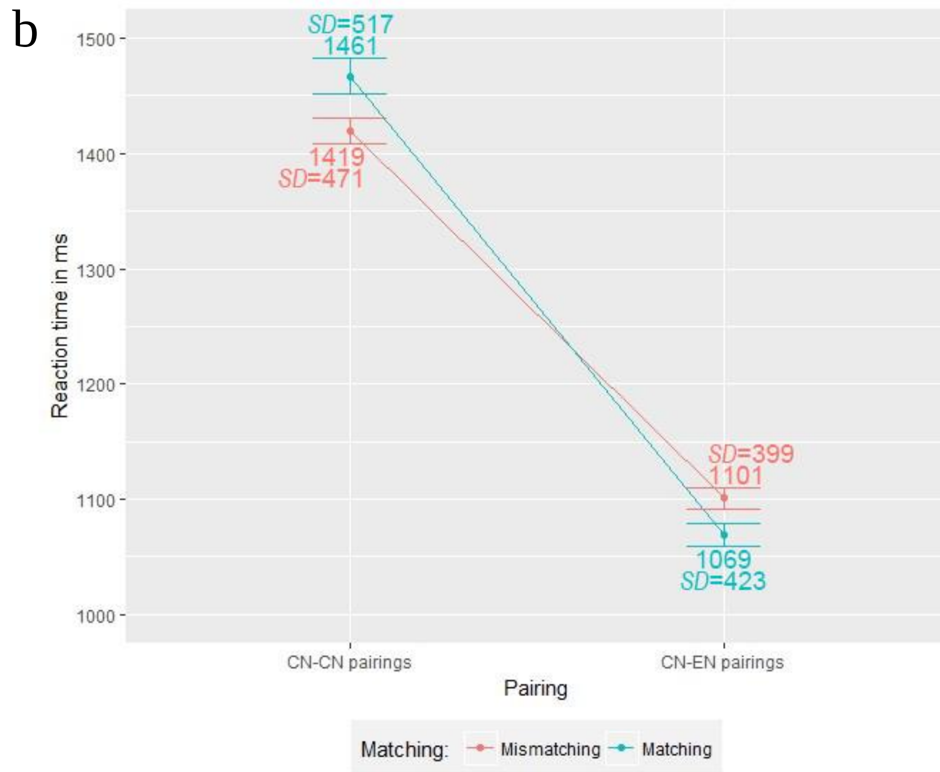


Figure 4.12. Reaction time results of interactions (a) Learning by Matching, and (b) Pairing by Matching, Testing tasks, Experiment 3.

Participants were faster for matching trials than mismatching trials overall; while they were fastest to matching C-E trials.

The two-way interaction Matching by Pairing ($p < .001$) reached significance. Further post hoc Tukey tests showed for C-C pairings and C-E pairings, the difference of RTs between matching and mismatching trials was both significant at the level of .001. That suggested participants identified matching C-E pairings significantly faster than the other pairings, although they were generally quicker at C-E pairings overall (Figure 4.12b, see p. 124).

Speed-accuracy trade-off before and after sleep.

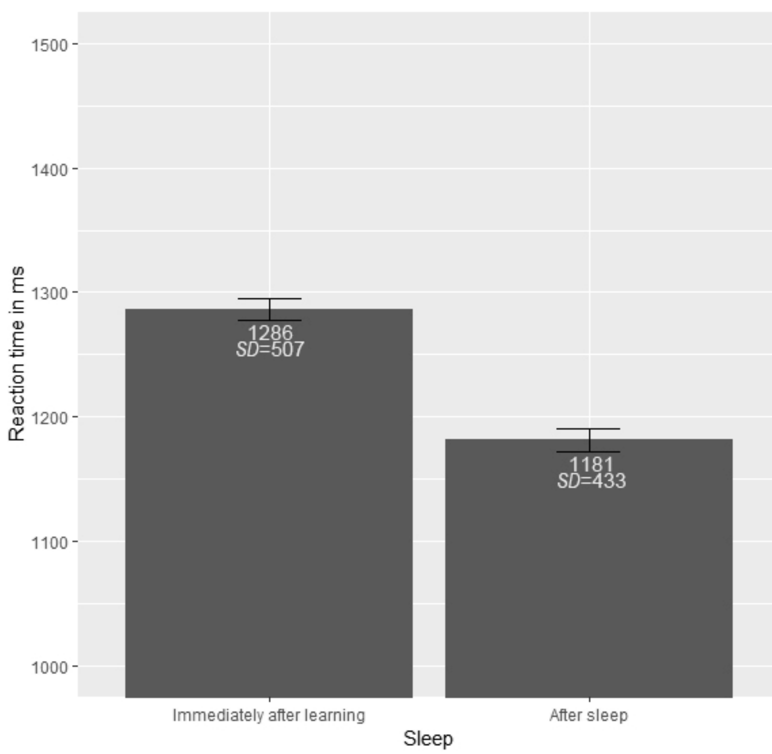


Figure 4.13. Results of fixed effect of Sleep for reaction time, Testing tasks, Experiment 3.

As shown in Figure 4.13 (p. 125), there is a significant decrease in reaction time after sleep ($p < .001$). Together with results of error rates, we found a speed-accuracy trade-off over sleep as error rates increased after sleep overall.

Table 4.3

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time in the Testing Tasks of Experiment 3.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1337.19	38.39	43.01	34.83	< .001***
Grouping	18.32	25.58	37.21	0.72	.47
Learning 2 vs. 1	-92.10	25.34	43.54	-3.63	< .001***
Learning 3 vs. 1	-191.80	34.87	41.51	-5.50	< .001***
Matching	-14.81	4.30	22063.58	-3.45	< .001***
Pairing	177.22	12.84	51.84	13.81	< .001***
Sleep	76.11	6.36	39.18	11.97	< .001***
Learning 2 vs.1 by Matching	12.37	5.76	22177.57	2.15	.03*
Learning 3 vs.1 by Matching	31.36	6.74	23888.02	4.65	< .001***
Matching by Pairing	-19.87	2.55	22407.28	-7.79	< .001***

Note. [a]*N* = 22,638. In the maximal random effect structure for which the convergence was realised, we included a random intercept and slopes of Grouping, Pairing, Learning and Sleep for items and a random intercept and slope of Pairing, Learning, and Sleep for subjects.

$*p < .05$. $***p < .001$.

Results of the Categorisation Tasks

Error rate.

Table 4.4 and Figure 4.14 summarise the error rate data of the categorisation task.

Grouped learners were more error prone for semantically unrelated trials but less error prone for semantically related trials compared to ungrouped learners.

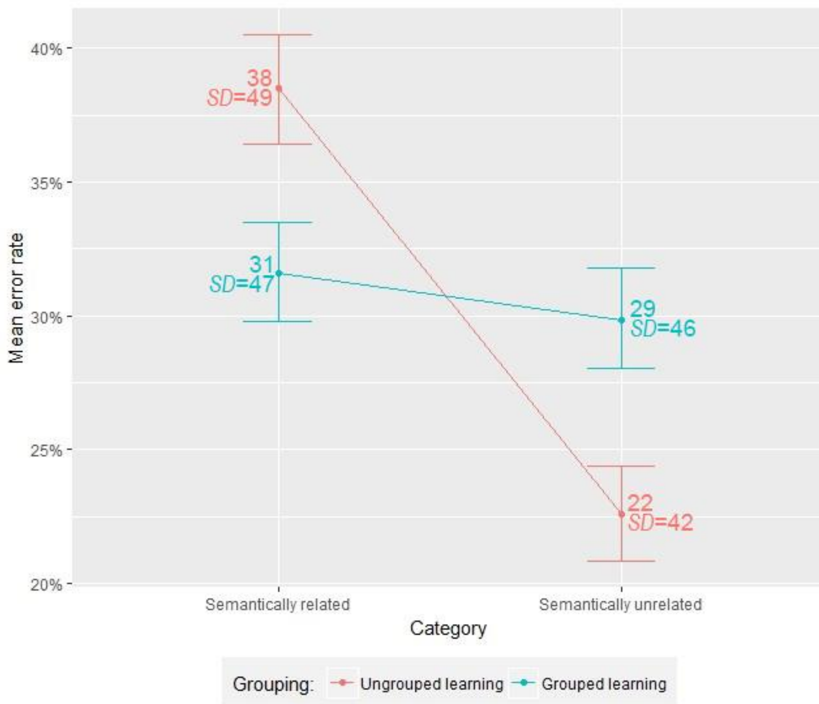
Figure 4.14a (p. 128) presents a very interesting picture: as the performance of grouped learners did not differ much between semantically related and unrelated pairings. In contrast, ungrouped learners did very well on picking up pairings which were not semantically related while they made more mistakes when dealing with pairings which were related in meaning ($p = .004$).

In addition, error rates of semantically related trials were significantly higher than error rates of semantically unrelated trials ($p = .001$).

Evidence of the established link between characters and semantics.

We also found a significant main effect of Time ($p < .001$) (Figure 4.14b, p. 128), which suggested from Day 2 to Day 7, consolidation still happened along and after the learning of those characters. The improved performance resulted from two factors: more learning and the consolidation effect. On one hand, there was another learning session on Day 3, which happened after the first categorisation task. On the other hand, there was no more input after Day 3. The consolidation after learning was obvious. Therefore, this improvement could still be seen as evidence of the establishment of long-term lexical representations of category information.

a



b

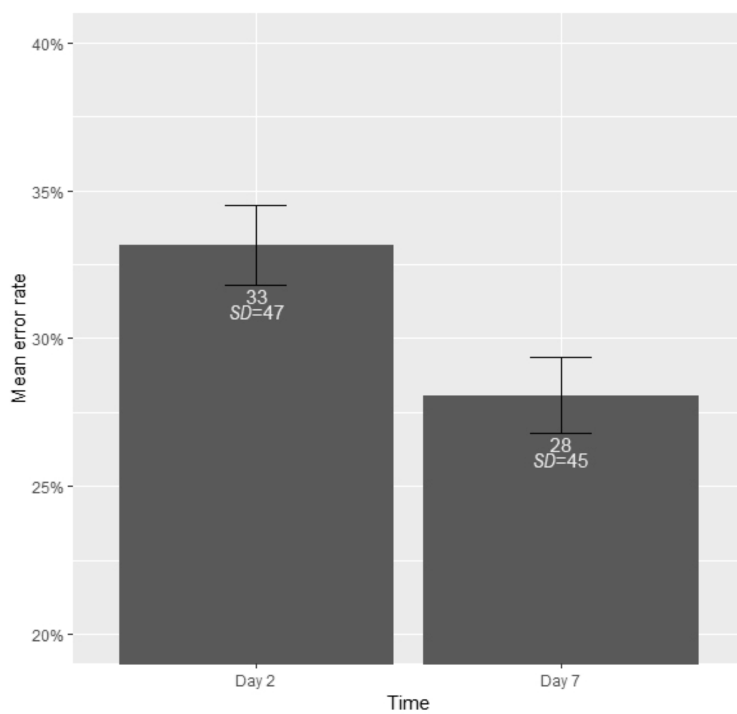


Figure 4.14. Results of error rates for (a) Category by Grouping, and (b) the main effect of Time, Categorisation tasks, Experiment 3.

Table 4.4

Summary for Fixed Effects in the Mixed Logit Model for Error Rates in the Categorisation Tasks of Experiment 3.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-1.07	0.15	-6.82	< .001***
Grouping	0.00	0.15	0.02	.98
Time	0.18	0.05	3.97	< .001***
Category	0.27	0.08	3.28	.001**
Grouping by Time	0.03	0.04	0.73	.14
Grouping by Category	-0.20	0.07	-2.90	.004**

Note. [a]N = 9,360, log-likehood =-4,978.5. The maximal random effect structure which converged included a random intercept for subjects and items and random slopes of Grouping, Time, Category and Grouping by Time for items, and random slopes of Time and Category for subjects.

**$p < .01$. ***$p < .001$.

Reaction time.

The reaction time data are shown in Table 4.5 and Figure 4.15.

Speed-accuracy trade-off for semantically related trials.

There was a significant main effect of Category (p = .002) in the data of reaction time (see Figure 4.15a, p. 130), with faster RTs for related trials. Together with results of error rates, it showed a speed-accuracy trade off: participants were significantly faster for semantically related trials, but at the same time, they made more mistakes.

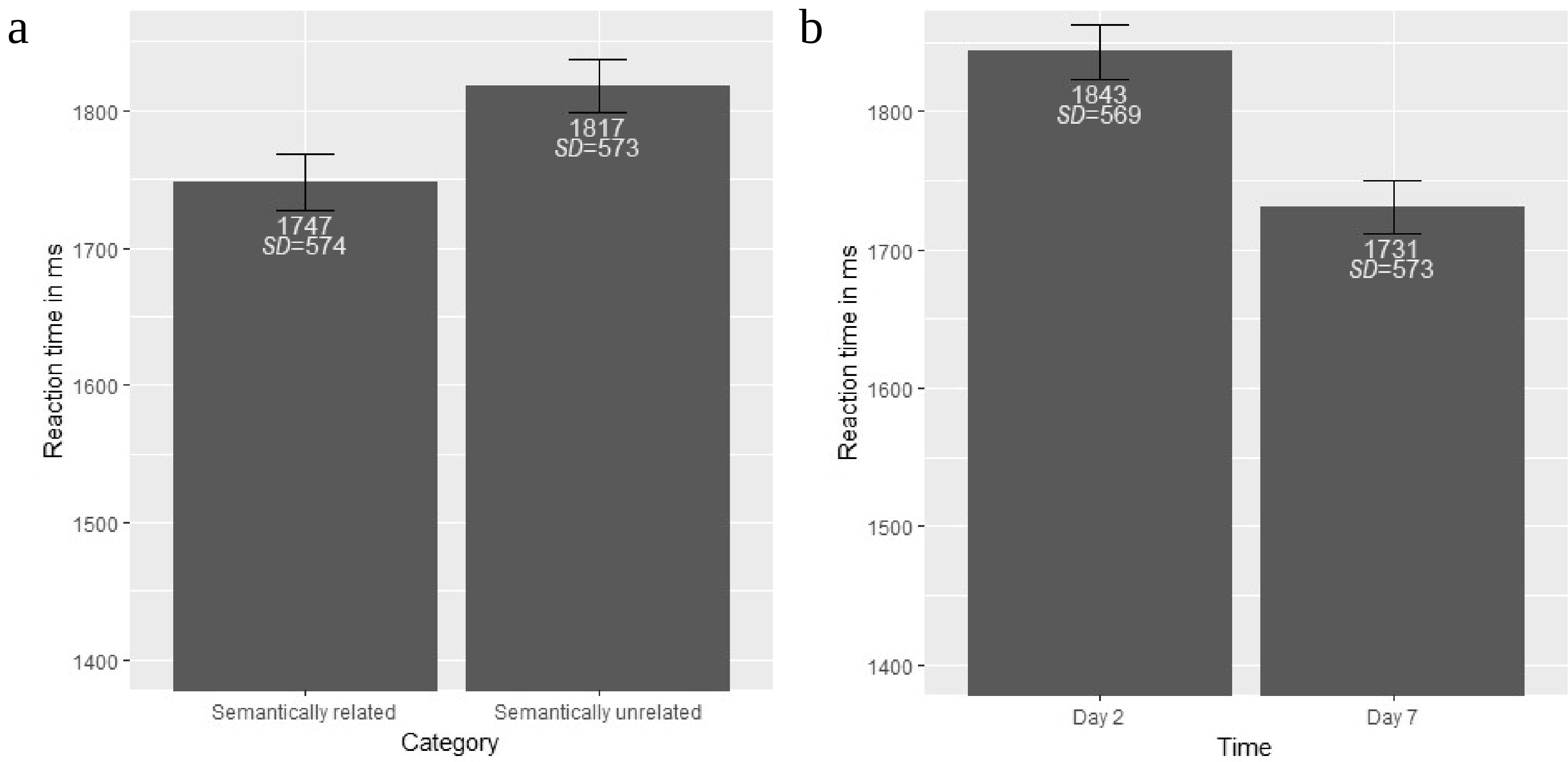


Figure 4.15. Reaction time results (a) for the main effect of Category and (b) for the main effect of Time, Categorisation tasks, Experiment 3.

Another piece of evidence for the establishment of the link between characters and semantics.

We also found a significant drop in reaction time from Day 2 to Day 7 (p = .007) (see Figure 4.15b, p. 130). Together with the result of error rate, it was clear that participants were faster and much more accurate on Day 7 than Day 2. That is another piece of evidence that the categorical representations were better on Day 7 despite the fact that no more input was delivered from Day 3.

Table 4.5

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time in the Categorisation Tasks of Experiment 3.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1774.40	50.57	39.17	35.09	< .001***
Grouping	15.63	46.64	36.27	0.34	.74
Time	52.52	17.78	35.90	2.95	.006**
Category	-37.60	11.87	87.76	-3.17	.002**

Note. [a]N = 6,494. The random effect structure contained a random intercept for subjects and items respectively, and random slopes of Grouping, Time, and Category for items, and random slopes of Category and Time for subjects. That was the maximal random effect structure which could converge.

**p < .01. ***p < .001.

Results of the Priming Tasks

Error rate.

Results of error rate data are summarised in Table 4.6 and Figure 4.16. The factor Priming was dummy coded with the reference level set as Priming 1.

As shown in Figure 4.16a (p. 133), in the full model, the error rates of matching and mismatching trials differed significantly (p = .015). Since the three-way interaction of Priming 3 vs. 1 by Matching by Pairing in the full model reached marginal significance (p = .08), we investigated it further using separate models for the data split by Matching. However, neither the results of the matching trials nor the results of mismatching trials contained a significant two-way interaction of Pairing by Priming. Therefore, the significance of the three-way interaction Priming 3 vs. 1 by Matching by Pairing was mainly due to the main effect of Matching.

Less error prone in C-C pairings than C-E pairings: Evidence of established phonological representation.

Error rates of C-E pairings were significantly higher than error rates of C-C pairings, as shown in the main effect of Pairing in the full model (p < .001, see Figure 4.16b, p. 133). That provided evidence for the establishment of the phonological representation.

Moreover, in the full model error rates were significantly higher for matching trials than mismatching trials, showing that participants were more accurate in rejecting mismatching trials (p = .015).

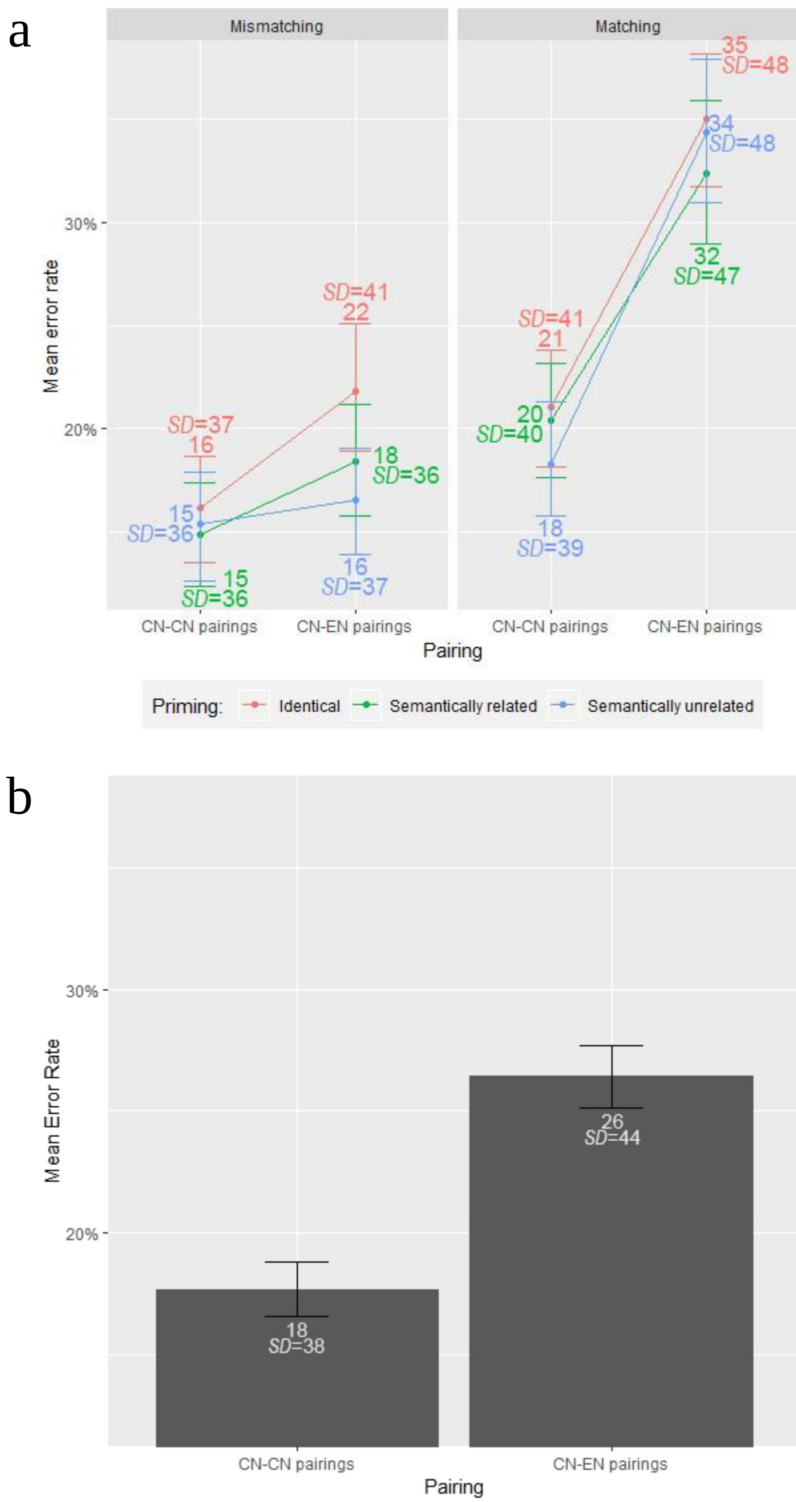


Figure 4.16. Error rates results of all trials. (a) the three-way interaction Matching by Priming by Pairing, (b) the main effect of Pairing, Priming task, Experiment 3.

Table 4.6

Summary for Fixed Effects in the Mixed Logit Model for Error Rates in the Priming Task of Experiment 3.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-2.04	0.29	-6.99	< .001***
Priming 2 vs. 1	-0.18	0.13	-1.39	.16
Priming 3 vs. 1	-0.13	0.17	-0.79	.43
Matching	-0.31	0.13	-2.43	.015*
Pairing	-0.73	0.16	-4.68	< .001***
Grouping	-0.12	0.24	-0.52	.60
Priming 2 vs.1 by Matching	-0.07	0.14	-0.52	.60
Priming 3 vs.1 by Matching	-0.08	0.17	-0.48	.63
Priming 2 vs.1 by Pairing	0.11	0.17	0.67	.50
Priming 3 vs.1 by Pairing	0.23	0.17	1.39	.16
Matching by Pairing	0.09	0.10	0.87	.38
Matching by Grouping	0.04	0.07	0.57	.57
Pairing by Grouping	-0.06	0.08	-0.77	.44
Priming 2 vs.1 by Matching by Pairing	0.03	0.08	0.37	.71
Priming 3 vs.1 by Matching by Pairing	0.14	0.08	1.74	0.08[†]
Model for matching trials[b]				
Intercept	-1.81	0.33	-5.43	< .001***
Priming 2 vs.1	0.02	0.17	0.15	.88
Priming 3 vs.1	0.05	0.22	0.24	.81
Pairing	-0.90	0.23	-3.99	< .001***
Grouping	-0.21	0.20	-1.02	.31
Priming 2 vs.1 by Pairing	0.20	0.19	1.08	.28
Priming 3 vs.1 by Pairing	0.20	0.23	0.87	.38
Model for mismatching trials[c]				
Intercept	-2.44	0.34	-7.20	< .001***
Priming 2 vs.1	-0.24	0.24	-1.00	.32
Priming 3 vs.1	-0.20	0.29	-0.68	.49
Pairing	-0.72	0.23	-3.06	< .001***
Grouping	0.06	0.18	0.33	.74
Priming 2 vs.1 by Pairing	0.17	0.35	0.49	.62
Priming 3 vs.1 by Pairing	0.26	0.29	0.91	.36

Note. [a]*N* = 9,120, log-likehood = -3,522.2. The maximal random effect structure which converged consisted a random intercept for subjects and a random intercept for items, together with random slopes of

Priming by Matching, Priming by Pairing, Matching and Pairing and Grouping for items, and random slopes of Priming by matching, Priming by Pairing, and Matching by Pairing for subjects.

[b]*N* = 4,560, log-likehood = -1,965.1. The random effect structure for matching trials included a random intercept for subject and a random intercept for item. It also contained random slopes of Grouping and Pairing by Priming for item, and random slopes of Priming by Pairing for subject. That is the maximal structure which converged.

[c]*N* = 4,560, log-likehood = -1,573.5. For the random effect structure for mismatching trials, there was a random intercept for subject and a random intercept for item. It also contained random slopes of Grouping and Pairing by Priming for item, and random slopes of Priming by Pairing for subject. It was the maximal random effect structure for which convergence was reached.

[†]*p*. < .1. **p* < .05. ****p* < .001.

Reaction time.

The reaction time data for all trials in the priming tasks are summarised in Table 4.7 and Figures 4.17 and 4.18. Here we also split the data over Matching, as most of the two-way and three-way interactions contained Matching. Also, the significant main effect of Matching (p = .001) showed that participants took longer to respond to mismatching trials than matching trials. Therefore, we designed models for matching and mismatching trials and analysed them separately. The model for these subsets contained all the interactions and fixed effects in the model for all trials, while excluding all the interactions and fixed effects containing Matching.

The potential facilitatory effect of Grouped learning.

Figure 4.17 shows the marginally significant two-way interaction Matching by Grouping for all trials (p = .09). Post hoc Tukey's test showed that only grouped learners showed difference between matching and mismatching trials at p< .05 level of significance, while for ungrouped learners, the difference was only marginal (p = .08). That suggested grouped learners were not significantly faster than ungrouped learners for matching trials as well as for mismatching trials. However, it should be pointed out that this effect was not conclusive.

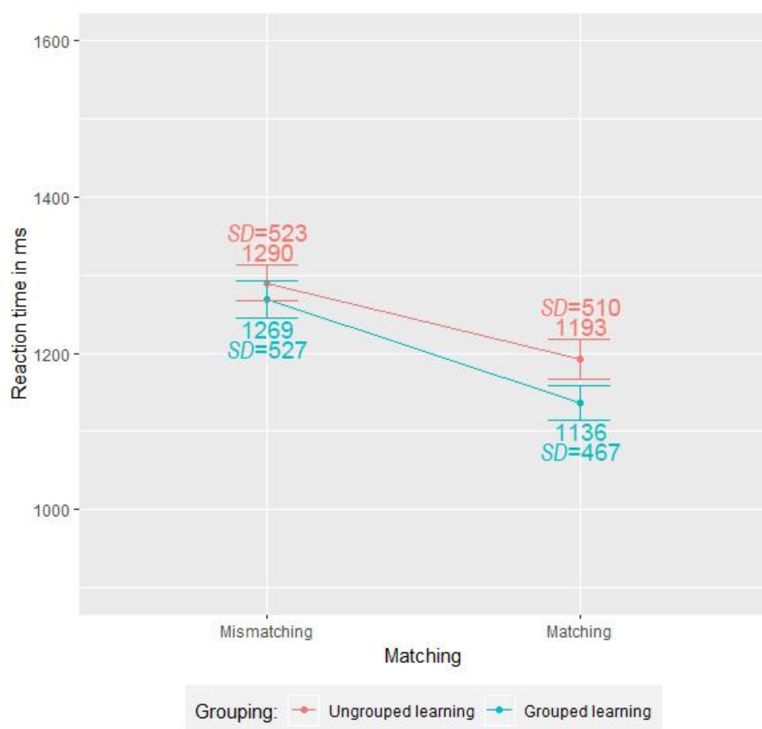


Figure 4.17. Reaction time results of the two-way interaction Grouping by Matching of all trials, Priming task, Experiment 3.

In the full model, the three-way interaction Grouping by Priming 2 vs.1 by Matching regardless of priming condition ($p = .01$) reached significance. However, there was no interaction between Grouping by Priming in neither of the split models.

The effect of priming: Priming by semantically related pseudo-characters increased the reaction time for matching C-C pairings compared to other priming conditions.

We found a significant interaction of Pairing 3 vs. 1 by Matching by Priming in the main model ($p = .02$, see Figure 4.18a, p. 138). It could be seen that shortest mean reaction time was found among trials primed by identical characters.

According to results of post hoc Tukey's tests, for matching C-E trials, there was no significant difference among the three priming conditions. For matching C-C trials, the reaction time differed: reaction time for identical trials and semantic related trials differed significantly at $p < .01$, reaction time for semantically unrelated trials also significantly differed from identical trials at $p < .01$. That was confirmed by the significant two-way interaction Pairing by Priming for matching trials ($p < .001$ for Priming 2 vs.1 and $p = .004$ for priming 3 vs. 1 respectively).

For mismatching trials, a two-way interaction of Pairing 2 vs. 1 and Priming was found ($p = .03$). However, results of post hoc Tukey's tests showed no significant difference among the three priming conditions, neither for C-C pairings nor for C-E pairings.

There was a significant main effect of Priming ($p = .004$ for Priming 3 vs. 1 only). As shown in Figure 4.18b (p. 138), participants were significantly faster for identical priming trials than semantically unrelated priming trials, while latencies of semantically related priming trials and identical priming trials did not significantly differ.

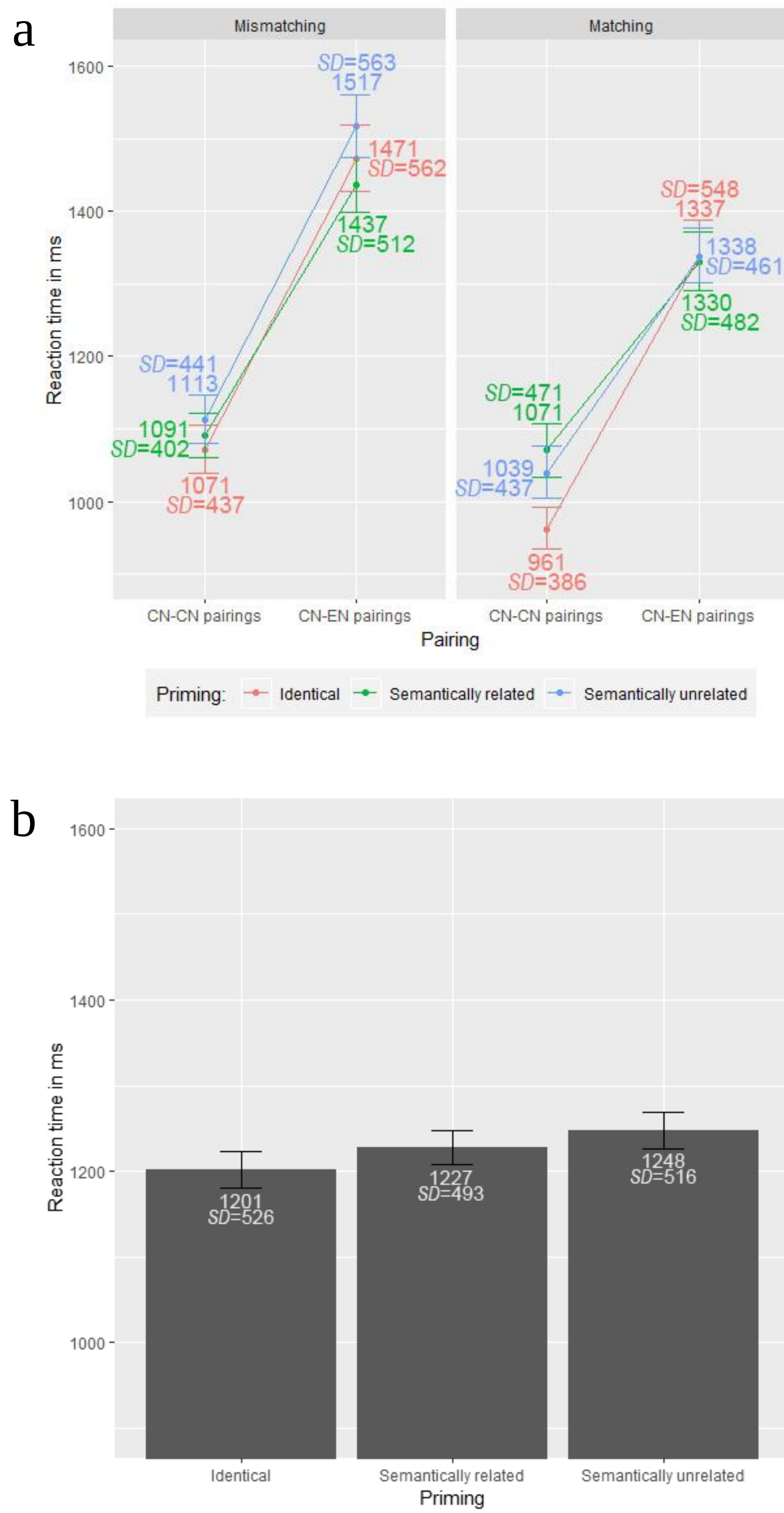


Figure 4.18. Reaction time results of all trials containing the factor of Priming, Priming task, Experiment 3. (a) Priming by Matching by Pairing, (b) main effect of Priming.

Table 4.7

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time in the Priming Task of Experiment 3.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1218.36	48.73	45.44	25.00	< .001***
Grouping	-21.18	43.80	36.09	-0.48	.63
Priming 2 vs.1	19.74	15.61	23.41	1.27	.21
Priming 3 vs.1	44.43	14.22	23.89	3.13	.004**
Matching	47.11	14.58	49.06	3.23	.002**
Pairing	-205.56	14.10	70.08	-14.57	< .001***
Grouping by Priming 2 vs.1	-4.97	11.27	1407.96	-0.44	.66
Grouping by Priming 3 vs.1	-9.80	11.21	4615.61	-0.87	.38
Grouping by Matching	17.00	9.94	96.63	1.71	.09†
Priming 2 vs.1 by Matching	-25.83	11.24	6927.23	-2.30	.02*
Priming 3 vs.1 by Matching	4.44	11.25	6933.92	0.39	.69
Priming 2 vs.1 by Pairing	49.28	11.25	6929.20	4.38	< .001***
Priming 3 vs.1 by Pairing	22.58	11.25	6935.65	2.01	.04*
Matching by Pairing	-0.75	8.04	6942.81	-0.09	.92
Grouping by Priming 2 vs.1 by Matching	-28.09	11.22	6917.82	-2.50	.01*
Grouping by Priming 3 vs.1 by Matching	-12.49	11.21	6922.48	-1.11	.26
Priming 2 vs.1 by Matching by Pairing	-15.95	11.25	6929.62	-1.42	.15
Priming 3 vs.1 by Matching by Pairing	-25.00	11.25	6934.02	-2.22	.02*
Model for matching trials[b]					
Intercept	1176.81	50.52	48.20	23.29	< .001***
Grouping	-35.82	42.98	35.79	-0.88	.41
Priming 2 vs. 1	47.82	21.13	18.16	2.26	.04*
Priming 3 vs. 1	41.80	17.80	28.92	2.35	.02*
Pairing	-208.38	15.80	62.72	-13.19	< .001***
Priming 2 vs.1 by Pairing	62.85	15.79	3198.43	3.98	< .001***
Priming 3 vs.1 by Pairing	45.89	15.79	3170.01	2.91	.004**
Model for mismatching trials[c]					
Intercept	1265.03	52.10	40.08	24.28	< .001***
Grouping	-49.20	41.95	36.89	-1.17	.25
Priming 2 vs. 1	-5.60	20.88	23.81	-0.27	.79
Priming 3 vs. 1	48.46	25.36	16.20	1.91	.07†
Pairing	-207.61	17.54	75.55	-11.84	< .001***

Priming 2 vs. 1 by Pairing	33.43	15.75	3616.96	2.12	.03*
Priming 3 vs. 1 by Pairing	-1.37	15.72	3622.02	-0.09	.93

Note. [a]*N* = 9,421. The maximal convergent random effect structure had a random intercept of item and subject respectively, while it contained random slopes of Priming, Pairing, Grouping and Matching for item, and random slopes of Priming, Pairing and Matching for subject.

[b]*N* = 3,333. The random effect structure for matching trials contained a random intercept for item and a random intercept subject, and random slopes of Grouping, Priming, and Pairing for items, and slopes of Priming, and Pairing for subjects. That was the maximal structure for which convergence was reached.

[b]*N* = 3,775. For the random effect structure of mismatching trials, we included random slopes of Grouping, Priming and Pairing for items and slopes of Priming and Pairing for subjects, as well as a random intercept for items and subjects. That was the maximal random effect which converged.

[†]$p. < .1$. $*p < .05$. $**p < .01$. $***p < .001$.

Discussion

As mentioned in the introduction section, this experiment aimed to answer the following questions: what is the effect of semantic grouping on the initial learning of L2 Chinese characters after eliminating the effect from shared radicals? What is the time-course during the initial days of learning? A quick answer to those questions is that based on these results, the overall effect of semantic grouping is facilitatory. In a time span of seven days, links from forms (phonological and orthographic) to meanings were well formed. I will specify each aspect in the following sections.

Effects of Semantic Grouping

In testing tasks, we found that grouped learners tended to be less error prone than ungrouped learners, while all participants were more error prone after sleep. This finding was in contrast to what we found in Experiment 1 which showed an inhibitory effect of grouping by semantic radicals. It is worth pointing out that the comparison between semantic grouping experiment (Experiment 3) and the experiment of shared semantic radical grouping (Experiment 1) was not over the same period. An analysis of results of the first two testing sessions of Experiment 3 will be carried out in Chapter 6 in order to see whether the facilitatory effect was a result of the longer period of learning time.

In categorisation tasks, grouped learners were more error prone for semantically unrelated trials but less error prone for semantically related trials compared to ungrouped learners. As reviewed in the L1 skilled reading process, semantic related words, like phonological and orthographic neighbours, could also result in coactivations of words in the same semantic category (H. W. Lee et al., 1999). During the categorisation task, each of the two characters on the screen could activate representations of characters of the same semantic category. A higher accuracy for correct identification of semantically related pairs could lie in the competition process during word identification. The competition effects in naming tasks usually caused delay in reaction time. Such delay was for the retrieval of the best match, resulting from the coactivation of related representations in the lexicon. When the two characters belong to the same semantic category, there is a good chance that their semantic representations are mutually activated by each other. Alternatively, this could be explained under Kroll's Revised Hierarchical Model. The evidence for the completed conceptual link is the category interference for words within the same semantic category, suggesting the coactivation of representations in the lexicon.

In the categorisation task, matching trials contained two semantically related character. Those two characters activated each other's representations. Instead of resulting in longer latencies or higher error rates in tasks of picture naming and translation tasks, the error rates were reduced in the categorisation task. This could be attributed to the nature of the present task. The present task only required a weaker conceptual link between the character and the semantic categorisation while for those competition resulting in inhibitory effects, the link was between the word and the exact semantic representation. These results suggested that learning characters in semantic group was more likely to trigger the coactivation of representations of the same semantic category. This could be seen as another piece of evidence for the facilitatory effect of semantic grouping.

Moreover, the facilitatory effect of learning was also seen in the priming task, while participants of grouped learning reacted faster than ungrouped learners for all trials on Day 7.

Reconstructing the Timeline of Initial L2 Chinese Character Learning: Critical Points

In testing tasks, regardless of participant group, there was a significant drop in error rates for matching trials, as the drop among C-E trials happened on Day 2, while another drop in error rates for C-C trials on Day 3. This suggested the formation of phonological representation took longer than the establishment of the link between the character and the English meaning. As participants were not familiar with Chinese phonology at all, they probably need more time to form the phonological representation and link it to the existed semantics. There was a significant decrease in reaction time on Day 2 for matching trials for both groups.

On Day 7, we found evidence for completion of the link from orthography to semantics from the categorisation task. There was a drop in error rate as well as in reaction time on Day 7 compared to the results on Day 2. In the priming task, we found phonological representations had been well-formed by Day 7. In that task, all participants were less error prone for in C-C trials than for C-E trials, although the error rates both conditions were low.

We discovered an inhibitory priming effect for semantically related primes among matching C-C pairings at the SOA of 1500 ms. This suggested that even at such a long SOA, semantic related primes could activate not only representations at semantic level but also at phonological level.

The Sleep "Consolidation" Effects: What is Consolidated and What is not?

In testing phase, we found that all participants were faster after sleep. However, they became more error prone after sleep, regardless of learner group. In this sense, we did not find a closer integration into lexical competition as suggested in Gaskell's studies (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). Moreover, difference in error rates between grouped and ungrouped learners emerged after sleep. In other words, after sleep, the difference between the two groups became explicit after sleep. One possible explanation for these results could be that the effect of grouped learning could be seen in longer term, after the information fully integrated into long-term memory.

However, in the categorisation task, we found that from Day 2 to Day 7, the consolidation effect was found in terms of reducing error rates. This improved performance was a joint result of sleep and learning input. In the period from Day 2 to Day 7, there was only one learning session on Day 3. However, from Day 7 no further learning input was given to participants. Clearly it was the effect of consolidation which took place from Day 3 to Day 7. As a result of exposure of stimuli in three days and the consolidation of sleep in seven days, the link from orthographic representation to semantics became solid and robust, resulting in such behavioural improvement.

Conclusion

In this experiment, we further explored the facilitatory effect of semantically grouped learning on the formation of lexical representation of Chinese characters. For the first time we demonstrated that within a time span of seven days, robust phonological representations in L2 Chinese could be formed. Links from phonological and orthographic forms to meaning were also constructed well. Using pseudo-characters without shared semantic radicals, we eliminated the interference from orthography on the effect of semantic grouping. The crucial milestones for the formation of semantic representations was on Day 2 while for phonological representations on Day 3. We also demonstrated that long-term phonological representation was well established from the result of priming tasks.

CHAPTER 5

EFFECTS OF PHONOLOGICAL GROUPING BY RHYME

Introduction

This chapter aims to further examine the effect of phonological grouping on the initial learning of Chinese characters through answering the following questions: what is the effect of learning characters in partial phonological overlaps? What is the effect of phonological grouping on the learning of characters in the long term? Could we identify critical milestones during the initial stage of L2 Chinese character learning?

Results of the homophone grouping experiment (Experiment 2) showed a facilitatory effect in terms of reaction times but this effect only emerged on the second day of testing. However, literature in skilled L1 reading suggested that partial phonological overlap, such as rhymes (Lukatela & Turvey, 1996), and repeated syllables (Szmalec et al., 2009; Wilcox & Medina, 2013), and identical phonological overlap, such as homophones (e.g., Y. A. Lee et al., 1999; Lukatela & Turvey, 1994; Rastle & Brysbaert, 2006), had different effects on the reading process. Therefore, it should be expected that different results from overlapping than identical phonological representations in the context of L2 vocabulary learning.

As mentioned in Chapter 1, all Chinese characters are monosyllabic. Each consists of a rime with or without an onset at the beginning. The onset is short and difficult to identify while the rime is relatively easy to discover. Rhyme is a very popular phenomenon in Chinese literature as well as in everyday life. We therefore chose rhyming characters as learning materials instead of characters those share onsets.

Experiment 2 only contained one learning session. Although most participants were able to reach a certain level of accuracy, whether the learning was enough for them was unknown. The poor result of the learning of pronunciation suggested the identity learning might take more time. Therefore, we are going to extending the study in the same way as in Experiment 3. More learning sessions on the second, and the third day were introduced to make sure the phonological representations form properly. The learning result could be seen though the accuracy of testing sessions. Moreover, the proper formation of lexical representation would be tested after a longer time span, as suggested by previous studies, seven days after the learning.

The present study used a categorisation task on the second day as well as on the seventh day to test the result of category learning. The result of identity learning could be reflected by a priming task on the seventh day. It is hoped that the transition from category learning to identity learning could be observed in the result.

The experiment used a between subject design. Learning materials for the grouped and ungrouped condition were the same 20 Chinese characters. The independent variable was the nature of the learning method: learning in groups of phonological overlapping words and learning in randomised groups. The rest of the settings were all identical to Experiment 3 in Chapter 4.

Methods

Participants

As with previous experiments, we used the Research Participation Scheme (RPS) for participant recruitment. There were 39 volunteers to participate this study. Six of them were male while the others were female. They were all adult British English speakers with normal or corrected to normal vision and no language impairments such as aphasia or dyslexia. They had no prior learning experience of East Asian languages, such as Chinese, Japanese, and Korean etc.

Materials

Learning materials were 20 simplified Chinese characters selected from *the List of General Standard Chinese Characters* (2013). They all represented monosyllabic Chinese words. They could be categorised into 5 sets of rhymes. Characters within each set shared segmental structure and tone. Four tones of modern Mandarin Chinese could all be found among these 20 characters. All characters were daily used, simple, and concrete nouns to maintain consistency of difficulty across sets.

Four characters in each set of rhyming words were free from semantic relationships and had minimal orthographic overlapping. They were translated into English with everyday simple English words. There was no significant difference among sets within the character list in regard to number of strokes, $F(4, 15) = 1.26$, $p = .33$. Neither was there any difference in the frequency of English translations, $F(4, 15) = 2.378$, $p = .10$. We collected the frequency of English translations from the package N-Watch (Version 2007.10.1) (Davis, 2005). For a full list of materials, see Table 5.1 (P. 148).

Same methods and settings of generating visual stimuli in previous experiments were followed here. Visual stimuli, including Chinese characters and English translations, were created with Adobe Photoshop (Version 12.0) in bmp format. Detail settings were specified as follows: For characters, 425*425 pixels and 530 kB in Kai font; while for English words, 425*122 pixels and 150 kB in Times New Roman.

Table 5.1

The Phonological Overlapping List of Experiment 4.

Chinese character	Pronunciation		Stroke number	English translation
	Pinyin	IPA		
兵	bīng	[piŋ˥]	7	*soldier*
钉	dīng	[tiŋ˥]	7	*nail*
晶	jīng	[tɕiŋ˥]	12	*crystal*
樱	yīng	[jiŋ˥]	15	*cherry*
猴	hóu	[xou˧˥]	12	*monkey*
楼	lóu	[lou˧˥]	13	*building*
骰	tóu	[tʰou˧˥]	13	*dice*
轴	zhóu	[tʂou˧˥]	9	*axis*
唇	chún	[tʂʰun˧˥]	10	*lip*
魂	hún	[???un˧˥]	13	*ghost*
轮	lún	[lun˧˥]	8	*wheel*
豚	tún	[tʰun˧˥]	11	*pig*
海	hǎi	[xai˨˩˦]	10	*sea*
铠	kǎi	[kʰai˨˩˦]	11	*armour*
奶	nǎi	[nai˨˩˦]	5	*milk*
崽	zǎi	[tsai˨˩˦]	12	*infant*
厕	cè	[tsʰɤ˥˩]	8	*toilet*
褐	hè	[xɤ˥˩]	14	*brown*
课	kè	[kʰɤ˥˩]	10	*lesson*
热	rè	[ɹɤ˥˩]	10	*heat*

The auditory stimuli were created in the same way as in Experiment 3 in Chapter 4. We used the reading app TTSApp with the audio base VW Hui (Chinese). In this experiment, we also fixed the length of all stimuli, we either stretched or compressed the original sound generated by the app and smoothing the treated sound with MATLAB (Version 8.1). The uniform length of the sound files was1000 ms. Similar to Experiment 3 in Chapter 4 (p. 100), the current experiment also used a longer and repeated version of the sound, which was 9 s long with six iterations. There was an interval of 500 ms between iterations. The longer version was created with Cool Edit Pro (Version 2.1).

Apparatus

The present experiment used the same experimental apparatus as we used in Experiment 3 (see p. 100 for further details).

Design and procedure

Identical to Experiment 3 in Chapter 4, the present experiment consisted of three learning sessions, five testing sessions, two categorisation tasks and one priming task across seven days (see Figure 4.1, p. 101).

Overall in the present experiment we followed the same design and procedure of the learning phase, testing phase as well as categorisation and priming tasks as we did in Experiment 3, Chapter 4. Specifications of differences will be addressed in the following sections.

Learning phase.

There were three learning sessions across 3 days from Day 1 to Day 3. As an independent manipulation, the 20 characters were presented either in group of the same rhyme or in an ungrouped way during the learning phase. As a consequence of the between subject design of the experiment, participants were categorised into two groups according to the way they learnt those 20 characters, i.e., half of the participants learnt the grouped character sets while the other half learnt ungrouped sets.

The grouped list consisted of five sets of characters, which within the set shared a rhyme and tone. Thus, there were four characters in each set. The ungrouped list was created by distributing the same 20 characters into five sets. Items of each set were four out of the 20 characters, while each item within the set was distinct for its pronunciation and meaning. During the experiment, we created five versions of set order by Latin square method. Order of items within each set was randomised by E-prime when presenting to participants.

Procedure and the time-course of learning trials was identical to Experiment 3, Chapter 4 (see Figure 4.2, p. 103 and pp. 102-104 for details about procedure of the learning phase).

Testing phase.

We used the same logic in the design of testing tasks of Experiment 3, Chapter 4 (pp. 104-106). As a Chinese character could be paired with either a sound or an English word, there were two kinds of pairings consequently. Each pairing could be matching or mismatching, therefore, we got the following four conditions:

Chinese character + the correct English translation;

Chinese character + the correct Chinese pronunciation;

Chinese character + an incorrect English translation;

Chinese character + an incorrect Chinese pronunciation.

They are further illustrated in Figure 5.1.

The same procedures were applied to the present experiment. All the rest settings remained unchanged as in Experiment 3 (see pp. 106-108 for details about procedures of testing phase).

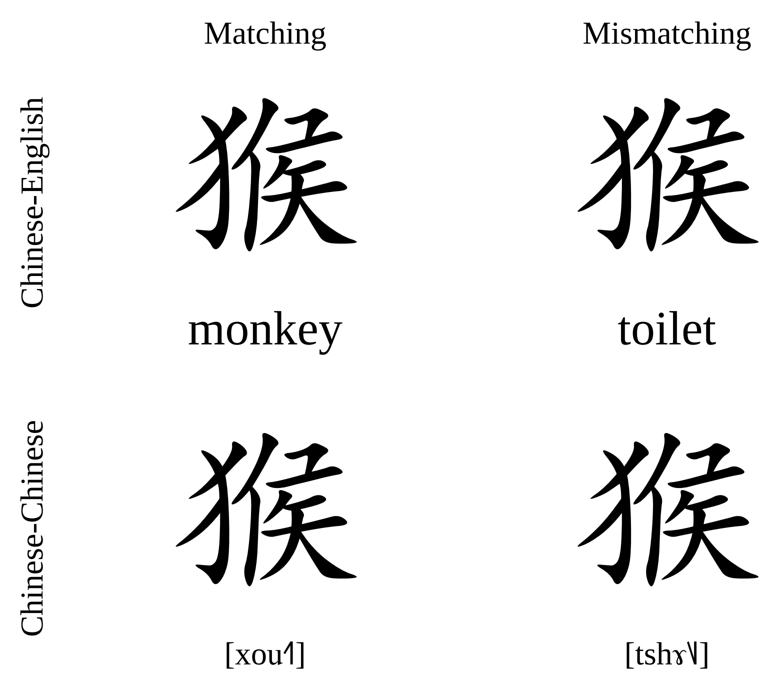


Figure 5.1. The design of testing phase stimuli of Experiment 4, where the character 猴 means *monkey*; while [ts^{h}ɤ˥˩] is the pronunciation of the character 厕, which means *toilet*.

Categorisation test.

We kept the basic design of categorisation test of Experiment 3 in Chapter 4 (pp. 108-109). However, in this experiment, any two characters from the character list were rhyming or not rhyming instead of semantically related or not as in Experiment 3. Each participant went through two categorisation tests, one on Day 2 and the other on Day 7. We used the same set of programmes for both learning groups. In this test, participants were asked to judge whether the two characters belonged to the same phonological category or not. Below are a set of examples (Figure 5.2).

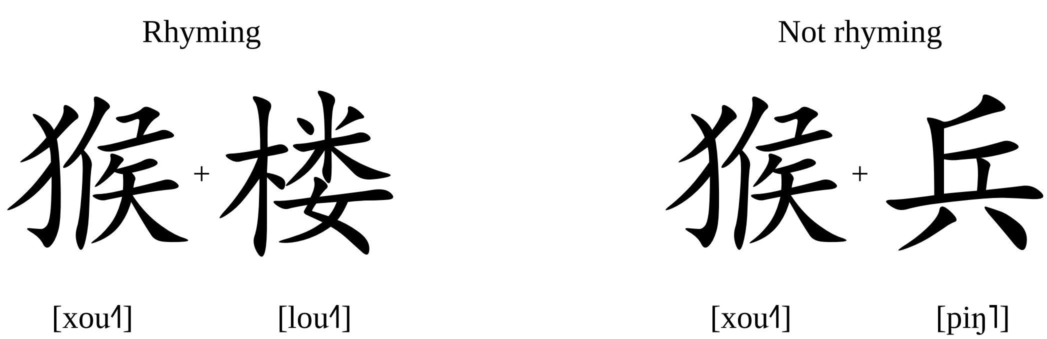


Figure 5.2. The design of categorisation test stimuli of Experiment 4, where the character means *monkey*; the character 楼 means *building*; and the character 兵 means *solider*.

Procedure.

We followed the same procedure of categorisation tasks as in Experiment 3 (pp. 109-111). In a typical trial is shown in Figure 5.3 (p. 153), the participant judged whether these two characters rhyme or not by pressing the button Yes or No. Other procedures and settings were also identical to Experiment 3.

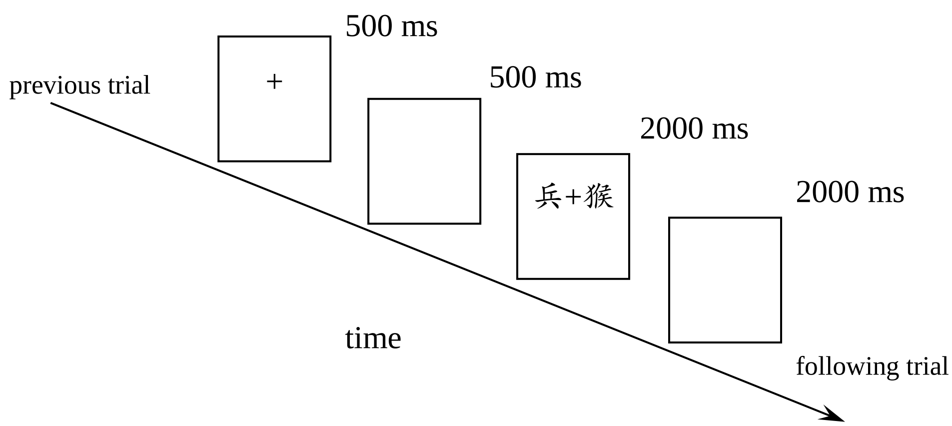


Figure 5.3. Timing of a categorisation trial of Experiment 4, where the character 兵 [piŋ˥] means *soldier*; while the character 猴 [xou˧˥], which means *monkey*.

Priming test.

The logic in designing the priming task was identical to Experiment 3 (pp. 111-113). In this experiment, the prime could either be the identical character to the target character, a phonologically related character, or a phonologically unrelated character. Participants were instructed to judge whether the target character match with the English words or Chinese sounds or not. Below we illustrated the design of the priming task in Figure 5.4 (p. 154). The same randomisation method and testing procedure applied to this experiment as well (see pp. 113-114 for further details).

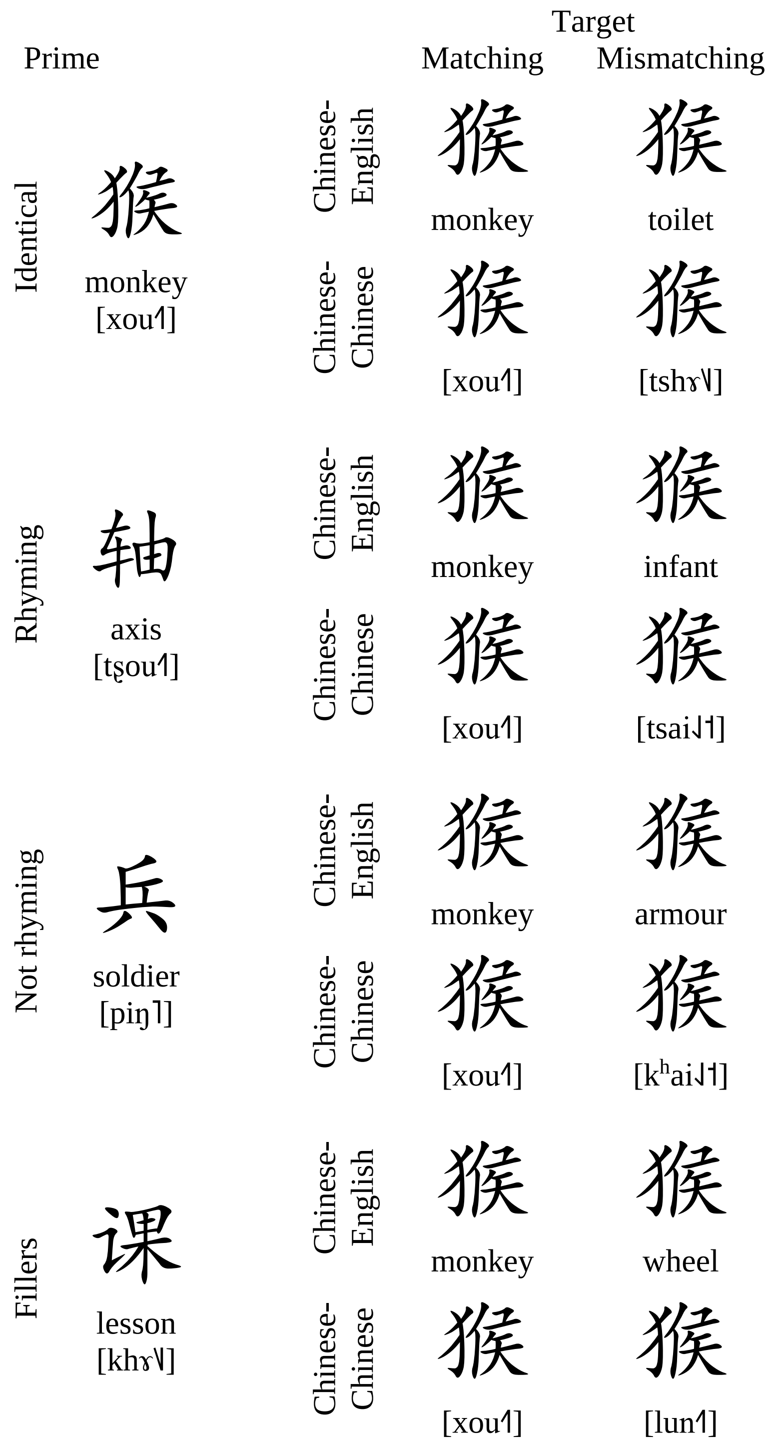


Figure 5.4. Examples of priming conditions in Experiment 4, where the character 猴 means *monkey*; [tsai˨˩˦]] is the pronunciation of the character 崽, which means *infant*; [k^hai˨˩˦] is the pronunciation of the character 铠, which means *armour* and [lun˧˥] is the pronunciation of the character 轮, which means *wheel*.

Results

Data analysis procedure

In this study, procedures of analysing the data were generally identical to Experiment 3 in Chapter 4 (see pp. 114-116 for further details). We excluded data from trials with incorrect responses during reaction time analyses, including 6,734 out of 27,200 trials (25%) in testing tasks, 3,626 out of 8,160 trials (44%) trials in categorisation tasks, and 1,484 out of 8,160 trials (18%) in priming task.

Nearly all the factors which we manipulated in the experiment were the same to those in Experiment 3, except the following ones:

For categorisation task, the factor *Category* was either *phonologically related* or *phonologically unrelated*.

In the priming task, factors *Matching* and *Pairing* were also included for the design of the trial. *Priming*, the type of the prime, could be *identical to the target*, *rhyming* and *not rhyming*.

Results of the Testing Tasks

Results of error rates and reaction time are presented in the following sections separately.

Error rate.

Table 5.2 and Figures 5.5-5.7 summarise the error rates data. All factors were deviation coded while Learning was dummy coded with the reference level set as Learning 1.

Participants became increasingly accurate with more learning inputs.

The significant two-way interaction Grouping by Learning 2 vs. 1 (p = .02) and Grouping by Learning 3 vs. 1 (p = .006) is shown in Figure 5.5. Based on results of post hoc Tukey's tests for ungrouped learners, the difference between any two learning conditions reached significance at $p < .001$ (i.e., Learning 1 vs. 2, Learning 1 vs. 3 and Learning 2 vs. 3). Similar trend was also found among grouped learners, except for the error rates difference between Learning 2 and 3 was at the level of <.05 significance, while for the other two comparisons, both at $p < .001$.

Furthermore, increasing learning inputs lead to the decrease in error rates for both groups overall ($p < .001$ for both Learning 2 vs. 1 and Learning 3 vs. 1).

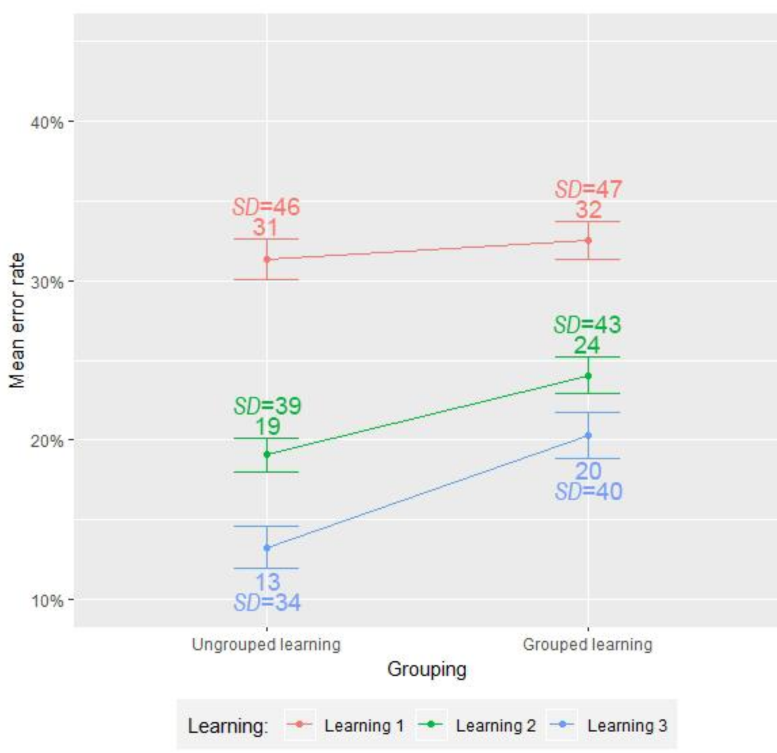
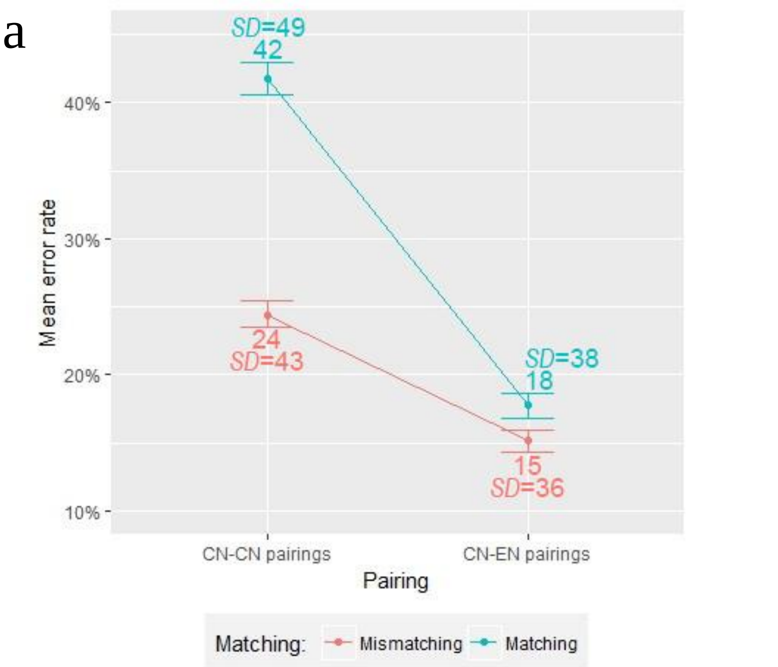
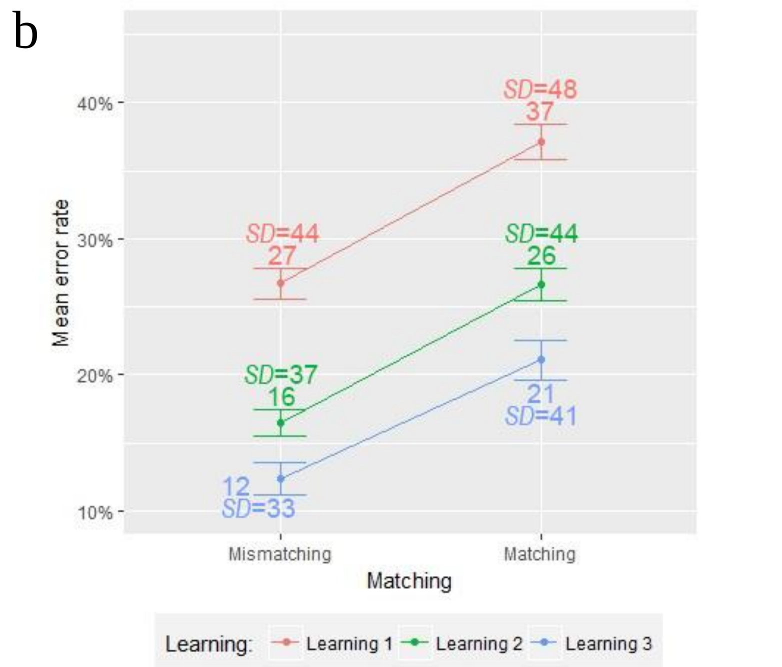


Figure 5.5. The results of error rate in terms of Grouping by Learning of Testing tasks, Experiment 4.





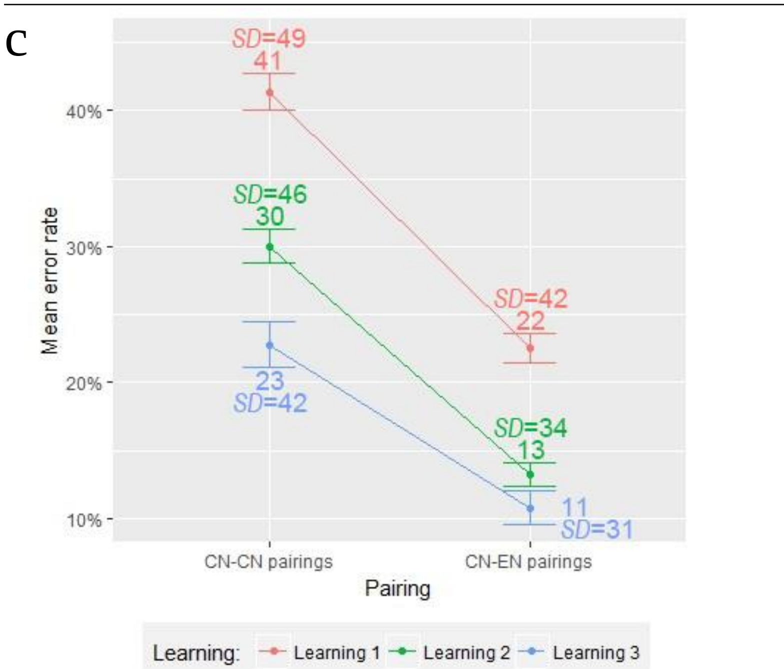


Figure 5.6. Results of error rate in terms of

(a) Pairing by Matching,

(b) Matching by Learning, and

(c) Pairing by Learning, Testing tasks, Experiment 4.

All participants were less error prone for mismatching trials compared to matching trials, and less error prone for C-E trials than C-C trials.

We found that lower error rates were always found among mismatching trials and among C-E pairing, compared to other Matching or Pairing conditions. As seen in Figure 5.6a (p. 157), the error rates for C-E matching and mismatching trials did not differ much. However, for C-C trials, error rates for matching trials were higher than error rates for mismatching trials at < .001 level of significance based on the result of post hoc Tukey HSD tests. That was confirmed by the significant two-way interaction Pairing by Matching ($p < .001$).

Moreover, Learning significantly interacted with Matching ($p = .02$ for Matching by Learning 2 vs. 1, and $p = .03$ for Matching by Learning 3 vs. 1, see Figure 5.6b, p. 157). As seen in Figure 5.6c (p. 157), the interaction of Learning by Pairing also reached significance ($p < .001$ for Learning 2 vs. 1) or approached significance ($p = .07$ for Learning 3 vs. 1). They all showed that from Learning 1 to Learning 3, the drops in error rates for different Matching and Pairing conditions varied from each other: for the factor of Matching, the decrease in error rates was larger for mismatching trials than for matching trials, while for the factor of Pairing, the decreased error rates were larger for C-C pairings than for C-E pairings.

Participants became more error prone after sleep, regardless of learner group.

We found a marginal significant main effect of Sleep ($p = .09$) in the reaction time data. As shown in Figure 5.7 (p. 159), error rates increased after sleep, which showed that in terms of error rates, sleep did not contribute to the consolidation of learning as far as the reduction of error rates.

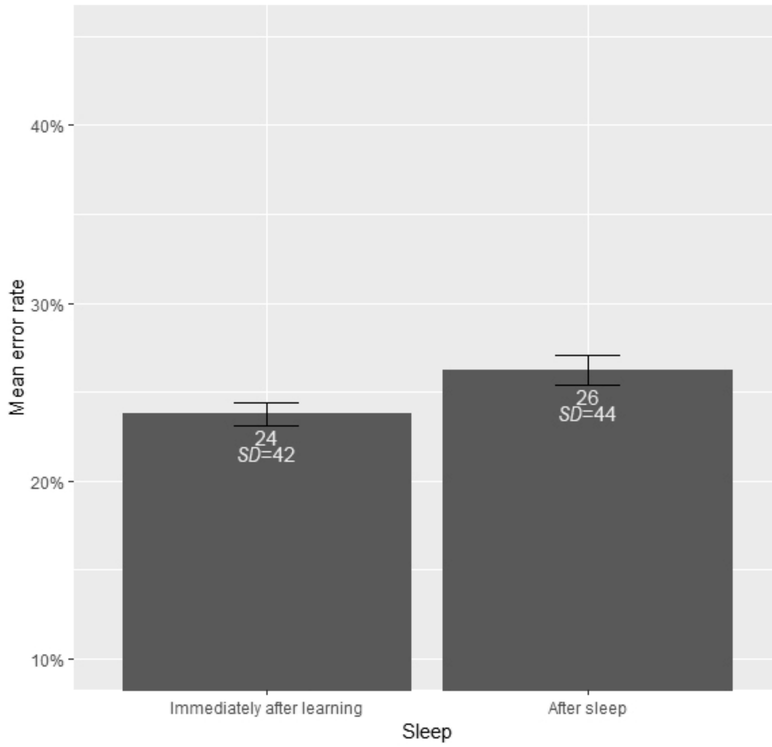


Figure 5.7. Error rates results of the main effect of Sleep, Testing tasks, Experiment 4.

Table 5.2

Summary for Fixed Effects in the Mixed Logit Model for Error Rate of Testing Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-0.93	0.12	-7.53	< .001***
Grouping	-0.04	0.09	-0.46	.64
Pairing	0.53	0.07	7.46	< .001***
Matching	-0.27	0.08	-3.28	.001**
Learning 2 vs. 1	-0.81	0.10	-7.91	< .001***
Learning 3 vs. 1	-1.34	0.16	-8.22	< .001***
Sleep	0.04	0.02	1.69	.09[†]
Grouping by Pairing	0.03	0.04	0.65	.52

Pairing by Matching	-0.15	0.03	-4.31	< .001***
Matching by Learning 2 vs. 1	-0.08	0.04	-2.31	.02*
Matching by Learning 3 vs. 1	-0.10	0.05	-2.11	.03*
Pairing by Learning 2 vs. 1	0.14	0.04	3.85	< .001***
Pairing by Learning 3 vs. 1	0.09	0.05	1.78	.07[†]
Grouping by Learning 2 vs. 1	-0.17	0.07	-2.32	.02*
Grouping by Learning 3 vs. 1	-0.34	0.12	-2.73	.006**

Notes. [a]*N* = 27,200, log-likehood = -12,546.5. The random effect structure included an intercept and slopes of Matching, Grouping, Pairing, and Learning for items, and an intercept and slopes of Learning, Sleep, and Matching by Pairing for subjects. It was the maximal random effect structure that reached convergence.

[†]p. < .1. *p < .05. **p < .01. ***p < .001.

Reaction time.

The reaction time data are summarised in Figures 5.8 and 5.9 and Table 5.3.

As shown in Table 5.3, we found a clear pattern among fixed effects: nearly all interactions containing the factor Pairing were significant. That included the four-way interaction Grouping by Pairing by Learning by Matching (p = .02 for Learning 3 vs. 1), Grouping by Pairing by Sleep (p =.002), Pairing by Matching (p = .04), and Pairing by Learning (p = .008 for Learning 2 vs. 1). Therefore, we split the data over the factor Pairing and analysed the subsets respectively. For each subset, we started from the model that contained all fixed effects and interactions but only excluding the factor of Pairing. The results of the two subsets are also shown different sections in Table 5.3.

Inhibitory effect of Grouping: grouped learners reacted slower than ungrouped learners, especially among C-E pairings.

Here in the main model, we found a marginal significant main effect of Grouping ($p = .08$), which showed that grouped learners reacted more slowly than ungrouped learners for all trials. At the same time, we also discovered a significant main effect of Grouping ($p = .03$) in the split model for C-E pairings, which shows that grouped learners needed more time than ungrouped learners to react to C-E pairings (see Figure 5.8).

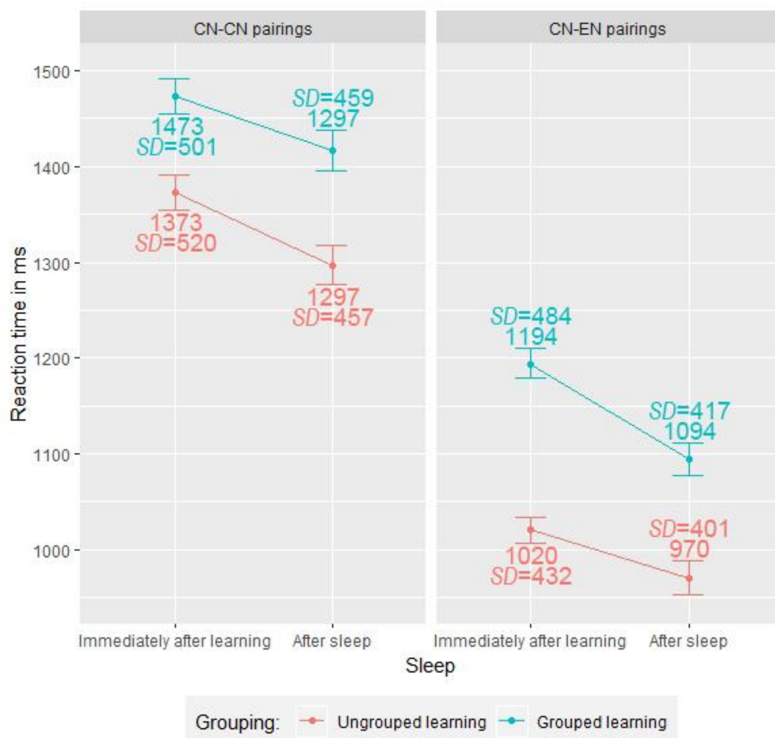


Figure 5.8. The result of three-way interaction Grouping by Pairing by Sleep for reaction time in Experiment 4.

The drop in latencies among C-E trials was larger for grouped learners compared to ungrouped learners.

For C-C pairings, the decrease in reaction time after sleep for grouped learners and ungrouped learners did not differ much. However, for C-E pairings, there was a significant two-way interaction Grouping by Sleep (p = .04), as shown on the right panel of Figure 5.8 (p. 161). Results of post hoc Tukey's tests showed that the decrease in reaction time for grouped learners and for ungrouped learners all reached significance at the level of < .001. Only before sleep showed the difference between grouped and ungrouped learners at marginal significance of .08.

Speed-accuracy trade-off over the factor of Sleep among all participants.

While the significant main effect of Sleep was found regarding the reduction of reaction time in the main and split models (p < .001 for each of the three models respectively), drops in reaction time varied among different pairings and groups of participants. Compared to the result of error rates, we can see an interesting picture: immediately after learning, participants were slower but maintain lower error rates, while after sleep, participants became faster and making more mistakes. That shows a speed-accuracy trade-off (SAT) before and after sleep.

Learning inputs contribute to reduced reaction times: A Drop in RT among C-E trials on Day 2 and on Day 3, a RT drop among C-C trials.

We found a significant two-way interaction Pairing by Learning in the main model (p = .008 only for Learning 2 vs. 1) in the reaction time data. In Figure 5.9 (p. 163), it was clear that the main drop in reaction times for C-C pairings was on Day 3 ($p < .001$). In contrast, participants were much faster for C-E trials after Learning 2. The drop of reaction time on Day 2 and 3 were both significant compared to Day 1 ($p < .001$ for both Learning 2 vs. 1 and Learning 3 vs. 1). This pattern was further reflected through the split data, as the main effects of Learning were both significant for the two subsets.

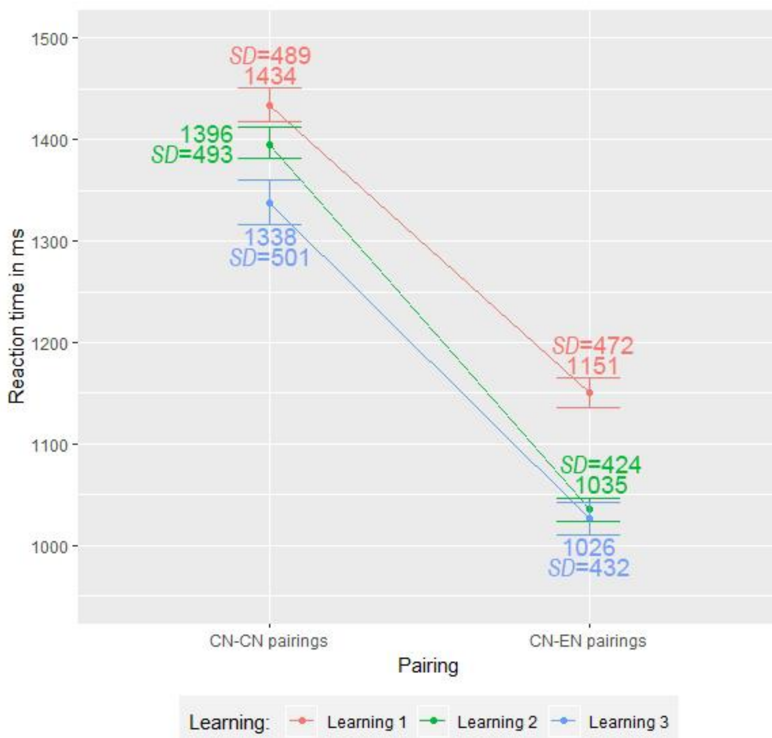


Figure 5.9. The two-way interaction of Pairing by Learning for reaction time in Experiment 4.

Participants were less error prone AND faster for C-E pairings than C-C pairings.

In addition, there was a significant main effect of Pairing ($p < .001$) and a significant main effect of Learning ($p < .001$) in the main model. Both show that participants were generally faster for C-E pairings compared to for C-C pairings, and they became gradually faster as they received more learning inputs.

Table 5.3

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time of Testing Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1305.05	41.11	36.42	31.75	< .001***
Grouping	-71.91	39.77	32.29	-1.80	.08[†]
Pairing	143.94	15.89	42.84	9.06	< .001***
Learning 2 vs. 1	-63.61	27.21	36.72	-2.33	.02*
Learning 3 vs. 1	-158.57	38.89	34.50	-4.07	< .001***
Matching	-4.73	11.34	31.20	-0.42	.68
Sleep	58.23	7.17	34.80	8.12	< .001***
Grouping by Pairing	15.30	14.14	33.53	1.08	.29
Grouping by Learning 2 vs. 1	7.23	26.10	32.04	0.28	.78
Grouping by Learning3 vs. 1	14.34	38.06	31.97	0.38	.71
Pairing by Learning 2 vs. 1	42.92	15.48	39.87	2.77	.008**
Pairing by Learning 3 vs. 1	14.35	16.23	38.37	0.88	.38
Grouping by Matching	-5.05	7.48	33.96	-0.68	.50
Pairing by Matching	-21.53	9.90	31.64	-2.18	.04*
Learning 2 vs. 1 by Matching	6.56	9.55	25.54	0.69	.50
Learning 3 vs. 1 by Matching	9.20	11.02	29.26	0.84	.41
Grouping by Sleep	-6.36	6.89	31.60	-0.92	.36
Pairing by Sleep	-0.25	2.93	20088.95	-0.08	.93
Grouping by Pairing by Learning 2 vs. 1	-5.80	13.65	31.95	-0.42	.67
Grouping by Pairing by Learning 3 vs. 1	-13.00	14.29	35.62	-0.91	.37
Grouping by Pairing by Matching	-7.82	5.49	97.22	-1.42	.16
Grouping by Learning 2 vs. 1 by Matching	9.17	6.98	62.51	1.31	.19
Grouping by Learning 3 by Matching	7.74	9.10	45.94	0.85	.40
Pairing by Learning 2 vs.1 by Matching	3.23	5.97	19824.60	0.54	.59
Pairing by Learning 3 vs.1 by Matching	3.74	7.05	19586.15	0.53	.60
Grouping by Pairing by Sleep	9.15	2.93	20095.02	3.12	.002**

Grouping by Pairing by Learning 2 vs. 1 by Matching	8.05	5.96	20034.75	1.35	.18
Grouping by Pairing by Learning 3 vs. 1 by Matching	16.97	7.04	19839.56	2.41	.02*
Model for Chinese-Chinese trials[b]					
Intercept	1446.25	44.83	33.56	32.26	< .001***
Learning 2 vs. 1	-19.31	36.94	28.89	-0.52	.60
Learning 3 vs. 1	-141.35	47.61	36.96	-2.97	< .001***
Matching	-17.95	12.76	38.13	-1.41	.16
Grouping	-57.96	41.99	32.71	-1.38	.18
Sleep	56.96	6.75	37.81	8.44	< .001***
Grouping by Sleep	3.81	4.70	175.86	0.81	.42
Model for Chinese-English trials[c]					
Intercept	1161.63	42.95	42.39	27.05	< .001***
Learning 2 vs. 1	-107.44	22.59	37.29	-4.76	< .001***
Learning 3 vs. 1	-174.22	35.08	36.20	-4.97	< .001***
Matching	20.05	13.36	21.00	1.50	.15
Grouping	-87.79	38.20	32.12	-2.24	.03*
Sleep	59.27	8.38	35.37	7.07	< .001***
Grouping by Sleep	-11.97	5.45	30.66	-2.19	.04*

Note. [a]*N* = 20,466. The maximal random effect structure for which the convergence was realised included a random intercept and slopes of Matching, Grouping, Pairing, Learning and Sleep for items and a random intercept and slope of Learning, and Matching by Pairing for subjects.

[b]*N* = 9,102. The maximal random effect structure that converged contained a random intercept and slopes of Grouping, Learning, Matching, and Sleep, and random slopes and intercept for Learning, Matching, and Sleep.

[c]*N* = 11,364. The random effect structure consisted random slopes and intercept for Grouping, Learning, Matching and Sleep for items, and random intercept and slopes of Learning, Matching and Sleep. That was the maximal random effect structure for which convergence is reached.

[†]*p*. < .1. **p* < .05. ***p* < .01. ****p* < .001.

Results of Categorisation Tasks

According to results of categorisation tasks for Experiment 4, there were no significant interactions for either reaction time data or error rates data. However, a significant main effect was observed for both data sets.

Error rate.

Table 5.4 and Figure 5.10 summarise the error rate data of the categorisation task. The reason why the insignificant interaction Grouping by Category existed in the model was the result of backward model selection. A model excluding the interaction Grouping by Category did not fit better than a model including this interaction ($\chi^2(1) = 105.68$, $p < .001$). Note that the significant *p* value suggested there was a significant contribution of the interaction. Therefore, the two-way interaction was kept in the model.

Participants were more error prone for rhyming pairings than for not rhyming pairings.

As shown in Figure 5.10 (p. 167), participants made significantly more mistakes for phonologically related pairings than for phonologically unrelated pairings ($p < .001$). That showed participants tended to judge two characters from the list as not rhyming, which further indicated they were better at distinguishing different rhymes than identifying rhyming pairings.

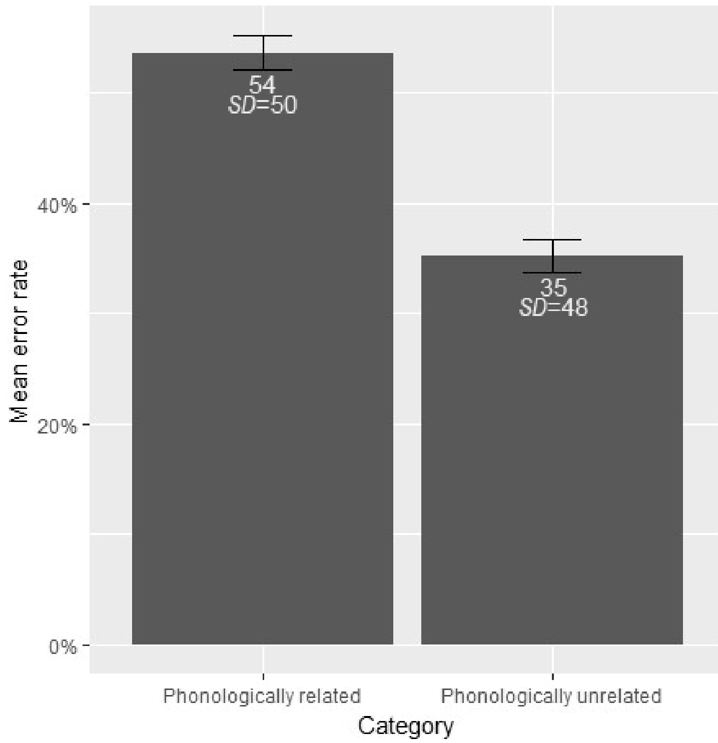


Figure 5.10. The result of error rates in terms of the factor Category of the Categorisation tasks in Experiment 4.

Table 5.4

Summary for Fixed Effects in the Mixed Logit Model for Error Rate of Categorisation Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-0.26	0.07	-3.58	< .001***
Grouping	0.11	0.07	1.58	.11
Time	0.04	0.03	1.42	.16
Category	0.48	0.12	4.02	< .001***
Grouping by Time	-0.04	0.03	-1.49	.14
Grouping by Category	-0.06	0.12	-0.52	.60

Note. [a]*N* = 8,160, log-likehood = -5030.6. The maximal random effect structure which reached convergence included a random intercept for subjects and items and random slopes of Time, Category, Grouping by Session, and Grouping by Condition for items, and random slopes of Time by Category for subjects.

***p < .001.

Reaction time.

The reaction time data of categorisation tasks are summarised in Figure 5.11 and Table 5.5.

Drop in latencies on Day 7: Evidence of established phonological categorical representation.

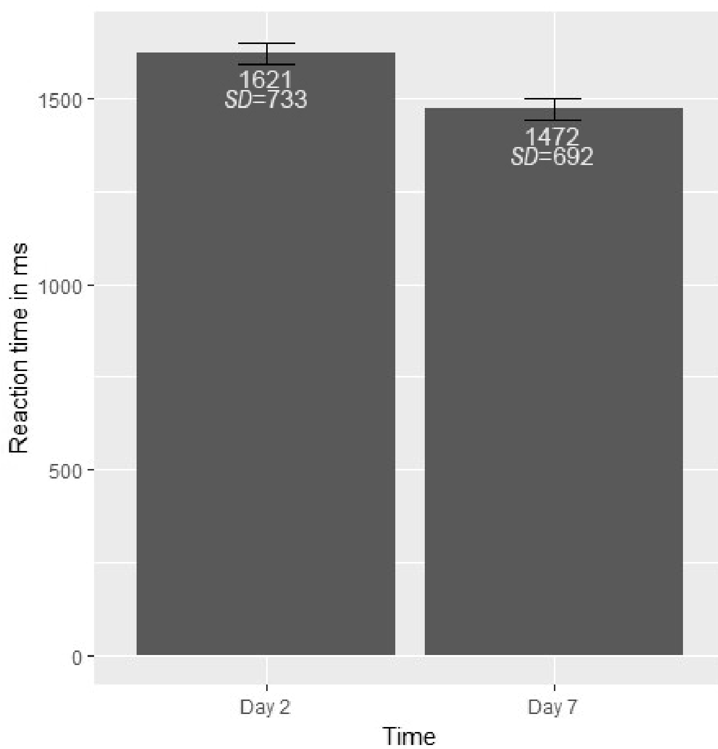


Figure 5.11. The result of error rates in terms of the factor Category of the Categorisation tasks in Experiment 4.

As seen in Figure 5.11, reaction times on Day 7 were significantly lower than on Day 2 ($p = .004$). Based on the information in Table 5.4 (p. 167) about error rates results, there was no significant main effect of Time ($p = .16$). That meant the reaction time dropped while they did not get less error prone as time went. Therefore, accuracy was a better measure of learning of representations than speed.

Table 5.5

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time of Categorisation Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1516.82	79.74	32.26	19.02	< .001***
Grouping	-79.47	77.38	31.77	-1.03	.31
Time	81.37	26.27	33.31	3.10	.004**
Category	0.21	15.16	32.06	0.01	.99

Note. [a]*N* = 4,534. The maximal random effect structure which reached convergence contained a random intercept for subjects and items respectively, and random slopes of Grouping, Time, and Category for items, and random slopes of Category and Time for subjects. That is the maximal random effect structure which could converge.

**$p < .01$. ***$p < .001$.

Results of the Priming Tasks

Error rate.

Table 5.6 and Figures 5.12-5.15 presented the error data for the priming task.

Accuracy was higher for mismatching trials than matching trials.

Figure 5.12 clearly presents that Pairing significantly interacted with Matching ($p < .001$). Based on results of post hoc Tukey's tests, the difference in reaction time between C-C and C-E trials for mismatching trials (at p < .05) was significantly smaller than that for matching trials ($p < .001$). Overall, participants made more mistakes for matching trials than for mismatching trials ($p = .01$), which showed they were better at rejecting incorrect pairings than at recognising correct pairings.

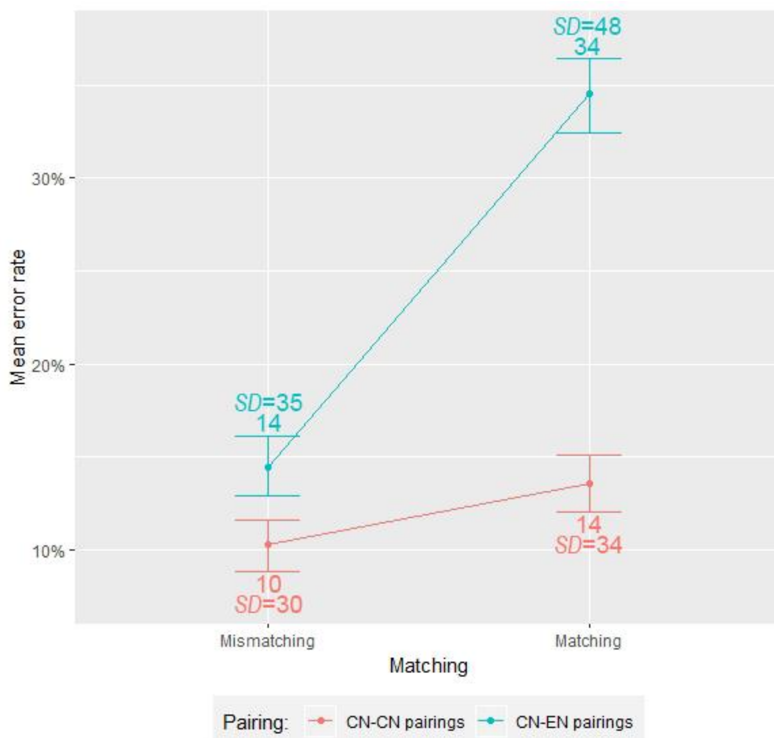


Figure 5.12. The result of error rates of the interactions Pairing by Matching of the priming task in Experiment 4.

Accuracy was higher for C-C pairings than for C-E pairings.

Interestingly, the error rate for C-C pairings was lower than that for C-E pairing ($p < .001$). That shows participants were able to recall the phonological information and the link between character and the pronunciation. This was consistent with better formation of the phonological representations of Chinese characters than the formation of semantic representations.

Negative priming effect was found among matching trials primed by phonologically related and unrelated characters.

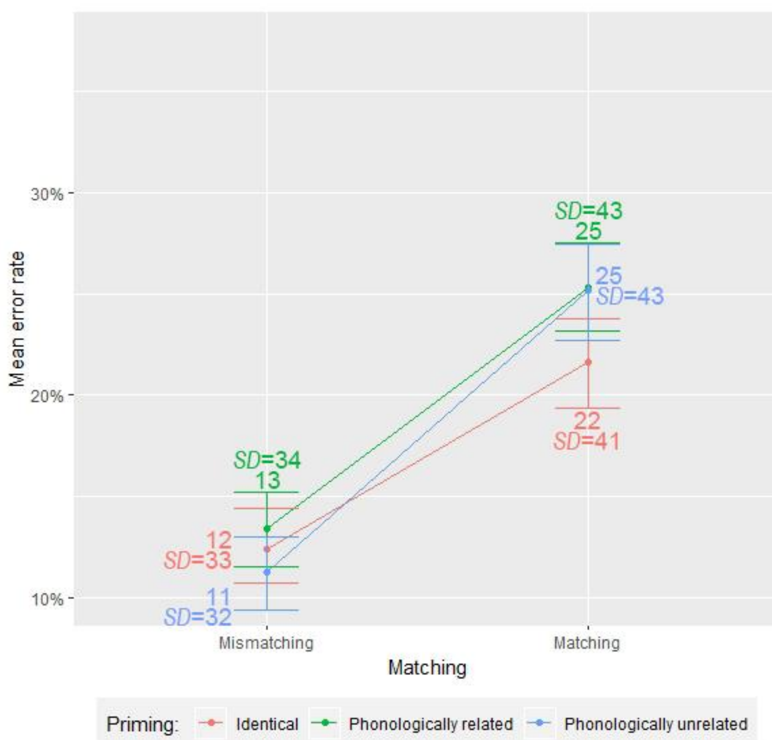


Figure 5.13. Error rate results for the interaction of Priming by Matching in the priming task in Experiment 4.

We found that the two-way interaction between Priming 3 vs. 1 with Matching reached significance (p = .03) (Figure 5.13, p. 171). According to post hoc Tukey's tests, there was no significant difference among the three priming conditions for mismatching trials as well as for matching trials. The difference between matching and mismatching trials primed by phonologically related characters reached significance at $p < .05$. The difference between matching and mismatching phonologically unrelated trials also reached significance at $p < .001$.

Table 5.6

Summary for Fixed Effects in the Mixed Logit Model for Error Rate of Priming Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-2.12	0.20	-10.90	< .001***
Priming 2 vs. 1	0.15	0.10	1.50	.13
Priming 3 vs. 1	0.09	0.10	0.96	.34
Matching	-0.27	0.11	-2.52	.01*
Grouping	0.21	0.18	1.22	.22
Pairing	-0.53	0.08	-7.01	< .001***
Priming 2 vs.1 by Matching	-0.09	0.08	-1.04	.30
Priming 3 vs.1 by Matching	-0.18	0.08	-2.18	.03*
Matching by Grouping	0.01	0.08	0.08	.93
Grouping by Pairing	0.01	0.06	0.22	.83
Matching by Pairing	0.24	0.06	3.74	< .001***

Note. [a]N = 8,160, log-likehood = -3097.4. The maximal convergent random effect structure contained a random intercept for subjects and a random intercept for items, and random slopes of Priming, Grouping by Matching, Grouping by Pairing, and Matching by Pairing for items, and random slopes of Priming, and Matching and Pairing for subjects.

*$p < .05$. ***$p < .001$.

Reaction time.

The reaction time data are summarised in Table 5.7 and Figures 5.14 and 5.15.

Faster AND more accurate for C-C pairings than for C-E pairings: clear evidence of the establishment of Chinese phonological representations.

We found that Pairing significantly interacted with Priming 2 vs. 1 ($p < .001$) (Figure 5.14a, p. 174). As shown in post hoc Tukey's results, for the reaction time C-E pairings, participants were faster for trials primed by phonologically related characters compared to trials primed by phonologically unrelated characters at .09 level of marginal significance. While for C-C pairings, RTs were slow but no significant difference between trials of difference priming conditions.

Overall, latencies were longer for C-E pairings than for C-C pairings ($p < .001$). Together with the result of error rates, it presented a very interesting picture: they were faster and more accurate for C-C pairings compared to C-E pairings. This indicated the establishment of new phonological representation linked to characters. In contrast the link between the English semantics and the Chinese characters was less robust.

Longest latencies among trials with phonologically unrelated (not rhyming) primes.

In addition, as shown in Figure 5.14b (p. 174), longer RTs were observed for trials primed by characters which did not rhyming with the target character than the other two priming conditions. This was confirmed by the marginally significant main effect of Priming 3 vs. 1($p = .09$).

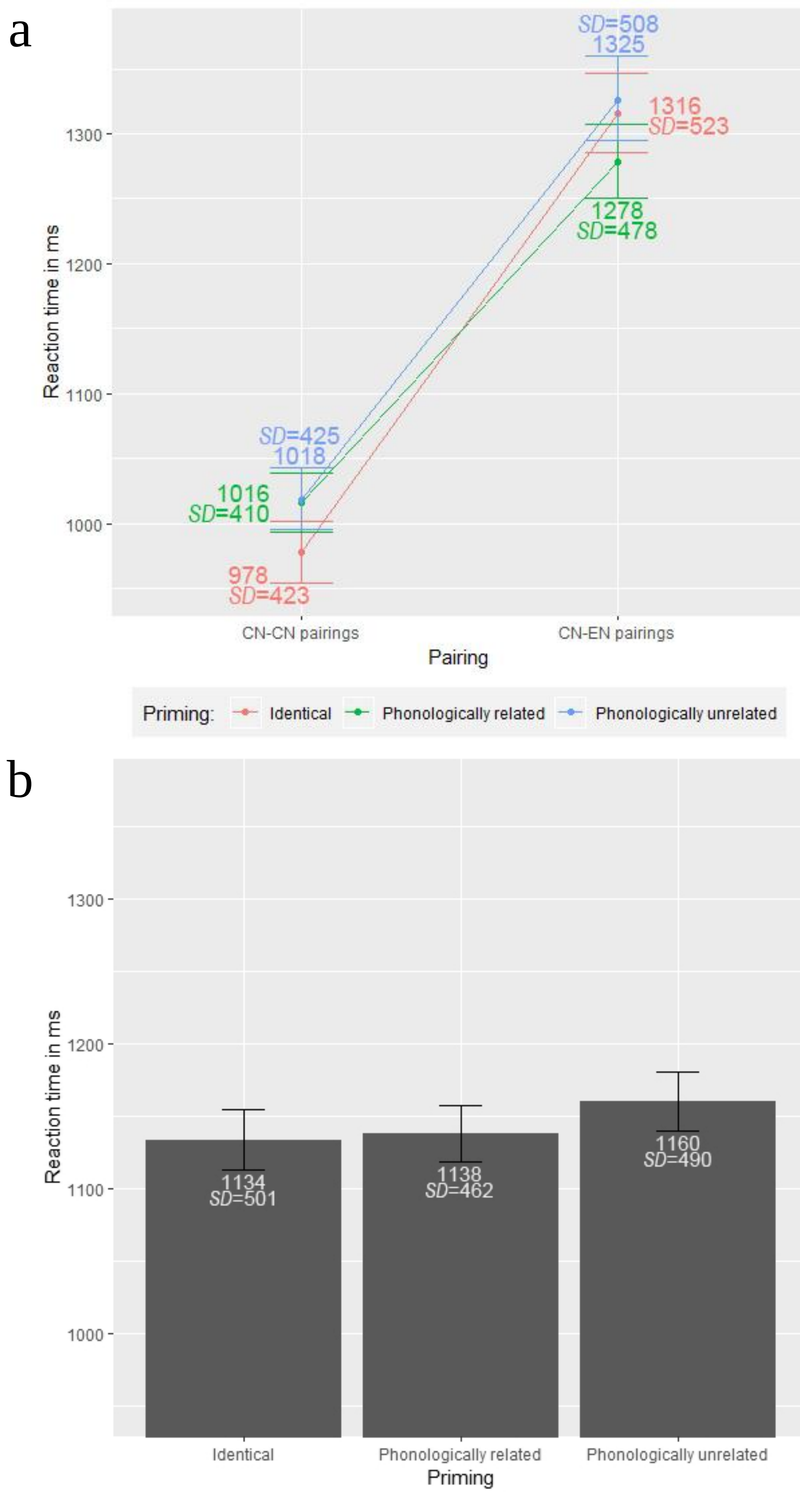


Figure 5.14. Results of reaction time regarding (a) Priming by Pairing, and (b) the main effect of Priming in the priming task of Experiment 4.

A speed-accuracy trade-off found over the factor of Matching.

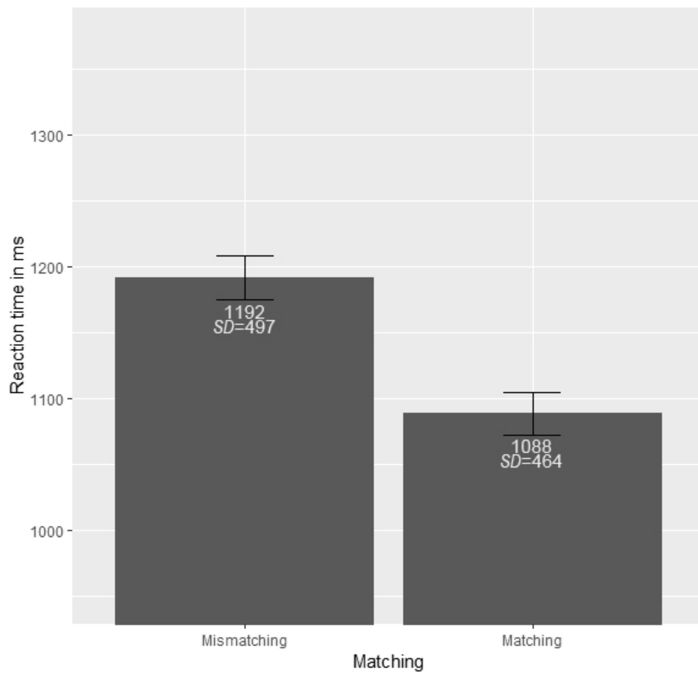


Figure 5.15. The main effect of Matching for reaction time, the priming task in Experiment 4.

Here in the reaction time data, we observed longer RTs for mismatching trials than for matching trials (p = .004, see Figure 5.15). The error rates of the priming task were higher for matching trials than for mismatching trials. Therefore, it is clear that there was a speed-accuracy trade-off due the factor of Matching.

Table 5.7

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time of Priming Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1161.16	48.55	34.97	23.92	< .001***
Grouping	51.34	41.18	33.07	1.25	.22
Priming 2 vs.1	1.22	12.69	23.32	0.10	.92
Priming 3 vs.1	26.70	13.30	22.80	2.01	.09†
Pairing	-173.26	12.59	64.03	-13.76	< .001***
Matching	29.46	9.42	31.21	3.13	.004**
Priming 2 vs.1 by Pairing	35.84	10.88	6946.67	3.30	< .001***
Priming 3 vs.1 by Pairing	13.20	10.84	6470.13	1.22	.22

Note. [a]*N* = 6,676. In the maximal random effect structure which converged, there were a random intercept of item and subject respectively, while it contained random slopes of Priming, Pairing, Grouping and Matching for item, and random slopes of Priming, Pairing and Matching for subject.

†*p*. < .1. ***p* < .01. ****p* < .001.

Discussion

In this study, we studied the effect of rhyme grouping on the initial stage of L2 Chinese learning in a span of time of seven days. The effect of rhyme grouping was examined across the three tasks. Through different tasks, we also identified some critical time which could reflect the building of phonological representations and links from forms to meaning. The effect of sleep consolidation was also tested in this study. I will address them respectively in the following sections.

Effects of Rhyme Grouping across Seven Days

As discussed in the introduction, the discrepancy between the rhyme priming effects and homophone priming effects could be found across several paradigms. Therefore, rhyme grouping was expected to work differently compared to homophone grouping. And that prediction was generally correct.

Overall, the effect of rhyme grouping was inhibitory. Firstly, grouped learners reacted significantly slower than ungrouped learners, overall. They spent significantly more time than ungrouped learners especially for C-E trials. For error rates of testing tasks, grouped learners showed trend to be more, although not significantly, error prone than ungrouped learners. These results above suggested that constant, explicit exposure to rhymes hinders grouped learners from building strong phonological representations for individual characters. This indicated ungrouped learning lead to more robust link from phonological and orthographic representations to the existing semantics than grouped learning. The underlying reason of such inhibitions could come from the phonological overlap effect in L1 skilled reading. Previous studies showed in L1 skilled reading, there was evidence that rhyming words could cause inhibitory priming effect in a primed word naming tasks, regardless of whether the rhyme shared orthographic feature (e.g., *hose-nose*) or not (*nose-rows*) (Lukatela & Turvey, 1996). Moreover, at a sentence level, it was shown that rhyming words in a sentence could cause confusion in meaning comprehension (Acheson & MacDonald, 2011).

The inhibitory effect of rhyming grouping we found in this experiment contradicted findings in previous literature in L1 and in L2 learning. For L1 vocabulary learning, repeated phonological features contributed to better recall afterwards (Szmalec et al., 2009). While for the learning of L2 words, presenting in phonological groups during learning with shared segments (syllables or beginnings) contributed to improved performance in the short term (Wilcox & Medina, 2013). It is worth pointing out that investigating the effect of rhyme grouping in most alphabetic language has a major inherent defect the transparency in the mapping between phonology and orthography in most alphabetic languages. This transparency could be potential confounds orthographic and phonological grouping effects. In Szmalec et al.'s study about L1 word learning (2009), they used legal consonant-vowel structures in Dutch (L1) and created nonwords with those syllables. While in Wilcox and Medina's study (2013), Spanish words were used as L2 learning material. Both languages are featured with high phonological transparency. In both experiments, stimuli were presented on screen and so were these tasks. Therefore, it is quite likely that phonology and orthography both contribute to the facilitation effects of phonological grouping found in both experiments.

By contrast, such consistency between orthography and phonology is (but not always) absent in Chinese characters. In this study, we deliberately eliminated rhymes with shared phonological radicals during the stimuli selection. In this way, no confound of orthography could possibly be found when investigating the effect of rhyme grouping. We therefore claim that effects of rhyme grouping on the L2 vocabulary learning are generally inhibitory. The facilitation from rhyme grouping found in some other languages could be a result of confounding with shared orthographic segments.

It shall be pointed out that within each set, characters shared segmental structure as well as tone. There were several reasons to keep consistency of tone within each rhyme set. Firstly, sharing rhyme as well as tone could reduce the workload of participants. The learning tasks could be too demanding for novice learners to concentrate on rhyme and tone at the same time. Secondly, tone was not a factor of manipulation in this thesis. No experiment was designed to investigate the effect of tones during the initial stage of character learning. However, it could be worthy investigating the effect of grouping in rhyme with different tones as a follow-up study in the future.

What Happened in Seven Days: Reconstructing the Timeline

In the testing tasks, all participants were less error prone and faster for C-E trials than C-C trials. This means the orthographic representations were better formed than phonological representations. It also suggested that the link from orthographic representations to semantics was more robust than the link from phonological representations to semantics.

In the categorisation task, we found further evidence that the links between phonological representations and orthographic representations were well formed. During the categorisation tasks, latencies dropped on Day 7 when compared to Day 2 while the accuracy stayed the same.

In the priming task on Day 7, all participants were faster and less error prone for C-C pairings compared to C-E pairings. This suggested phonological representations for individual characters was also formed rather well. Compared to results of the testing session, we found that the pattern of accuracies for the C-E/C-C pairings reversed. That is, in the testing session, participants were more error prone for C-C trials than C-E trials. However, during the priming task, regardless of priming condition, participants were more error prone for C-E pairings than for C-C pairings.

There were several factors contributed to the reversal of differences in accuracy between C-E and C-C trials. In testing tasks, participants were much more familiar with these English words than these Chinese sounds. During the first few days of the experiment, participants could easily link the orthographic representation with the English word. At the same time, they struggled with Chinese sounds while gradually formed the link between the sound and the character. However, after four days of consolidation, these Chinese phonological representations were well formed and showed strong robustness through the result of the priming task. A possible explanation is that the link between phonological and orthographic forms needs more time to form than the link between semantic representations and orthographic representations. According to the lexical constituency model, the former process needed well specified phonological and orthographic representations; while the latter, only specified orthographic representation was needed. The reversal of trend was the joint result of the consolidation and the formation of phonological representations.

In the priming task, we found a speed-accuracy trade-off over the factor of matching, namely, overall, they were faster for matching tasks yet making more mistakes, while they were slower but they were more accurate for mismatching trials. That tendency showed that for matching trials, participants were likely to give a response quickly although that would sacrifice the accuracy, whereas for mismatching trials, they tended to respond with some thinking.

We found that both groups were less error prone for mismatching trials compared to matching trials, while no significant difference in latencies between matching and mismatching trials in the testing task. In other words, participates were more accurate to identify the mismatch between meaning/pronunciation and the character. This pattern was also found in the priming task, where accuracy was higher for mismatching trials than matching trails. We believe that was due to the underlying nature of the two conditions. As for matching trials, participants need to recruit the exact lexical representations in order to make a judgement; while for mismatching trials, the correct rejection of incorrect pairings do not involve information at such details. We believe that explanation also applied to the following phenomenon: in the categorisation tasks, participants were better at rejecting non-rhyming pairings than correctly recognising rhyming pairings.

We also found inhibitory priming effect among trials primed by phonologically related (rhyming) and unrated (not rhyming) characters in terms of lower accuracy and longer latencies, compared to trials primed by rhyming characters at 1500 ms SOA.

We observed the time-course of the building of different level of lexical representations during initial stage of L2 Chinese character learning. For all participants, as more learning inputs they received, on Day 2, the reaction time dropped significantly for C-E pairings, while on Day 3, participants were significantly faster for C-C pairings than they did previously. A possible reason could be that for a complete novice learner, building and linking the phonological representations in another language surely required longer time than connecting the Chinese orthography to the English semantics. Nevertheless, empirical results for the fast formation of phonological representations and strong link to semantics still suggested a very swift process during the initial stage of L2 character learning.

The Effect of Sleep: Similar Findings to Experiment 3

Similar to our findings in semantic grouping experiment (Experiment 3) in the previous chapter, sleep did not contribute to the lexicalisation of the newly learnt characters in the testing tasks. Again, we found that after sleep, significant difference between grouped learners and ungrouped learners emerged. That was the decrease in latencies for grouped learners was significantly larger than that for ungrouped learners, although grouped learners spent longer time than ungrouped learners overall. That suggested the grouped learners benefited more from the sleep consolidation effects than ungrouped learners. However, we found a speed-accuracy trade-off for sleep in testing tasks. That was, before sleep, participants were less accurate but spent shorter time than they did after sleep. In the categorisation task, however, we found a decrease in reaction time on Day 7 compared to their performance on Day 2. Again, this finding was consisted with our findings for the semantic grouping experiment (Experiment 3). Detailed discussion will be addressed in the following chapter.

Conclusion

In this experiment, we investigated the effect of phonological grouping in terms of learning Chinese characters in rhyming sets. In contrast to findings from previous literature, we found inhibitory effects of rhyme grouping on L2 vocabulary learning. The absence of phonological transparency in our stimuli could be accounted for that difference. We successfully observed the time-course of the building of different levels of lexical representations during the initial learning of L2 Chinese characters. The critical time for the establishment of the link between the Chinese orthography and English semantics happened on Day 2 while the formation of Chinese phonological representations on Day 3. That was reflected through the reduction in error rates in the testing tasks. We also discovered that the phonological representations of Chinese characters were able to last at least 96 hr after the last learning input, suggesting the robustness of the newly built representations in long-term memory.

In conclusion, we testified that the effect of phonological grouping by rhyme was generally inhibitory. We also demonstrated that at such an initial stage of learning, the formation of different levels of lexical representations could happen and be observed and reflected through behavioural results.

CHAPTER 6

SUMMARY AND GENERAL DISCUSSION

In this thesis, we demonstrated for the first time that the learning of Chinese characters can take place in a short time span of seven days. After only three input sessions, representations of learnt characters entered the long-term memory and could be retrieved as long as 96 hr without further inputs. These results also showed that learning characters in semantic or phonological groups affects learning results. In this chapter, I will start with a summary of key findings of these four experiments. Followed by that, I will present findings from the analysis of the first two testing sessions of Experiments 3 and 4 to show the robustness of effects of grouping found in these experiments. Then I will discuss implications of these findings on L2 Chinese learning as well as for future studies. I will finish this chapter with a conclusion of studies in this thesis together with future directions of my research.

Semantic Grouping: Summaries and Implications

Effects of semantic grouping were investigated through Experiments 1 and 3. In Experiment 1, we used real characters with shared semantic radicals. In Experiment 3, we created pseudo-characters that do not share semantic radicals in the same semantic category. Relations among items in each semantic category could be represented by symbols like [S±O]. In such symbols, "±S" represents whether shared semantic information is involved or not, while "O" stands for whether shard orthographic characteristics is included or not. Therefore, in Experiment 1, the shared semantic and orthographic radical among characters could be represented as [S+O+] for Experiment 1, and for Experiment 3, [S+O-].

In this section, I will mainly emphasise on these findings from matching trials. As shown in results of testing and priming tasks in Experiments 1 and 3, tendencies presented from matching and mismatching trials were dramatically different. Therefore, it was reasonable to believe that these two conditions had their distinct underlying natures. According to Perfetti's lexical constituency model (2005), successful word reading relies on well specified lexical representations. For matching trials, lexical representations needed to highly specified and actively recruited in order to make a correct judgement. However, for mismatching trials, a correct rejection of incorrect pairings does not involve information at such details. Consequently, results from matching trials could be more conclusive than results from mismatching trials.

Table 6.1 summarises key findings of Semantic grouping, including Experiments 1 and 3.

Grouping words according to their semantic category in vocabulary learning is as a tool in pedagogical practice (Finkbeiner & Nicol, 2003). However, some researchers like Tinkham (1993, 1997) and Waring (1997) pointed out that semantic grouping caused difficulties during the initial stage of L2 vocabulary learning. Our findings of S+O+ grouping generally showed tendencies to confirm their view as compared to ungrouped learners, grouped learners were more error prone when responding to Chinese characters which appeared with Chinese sounds. They were also slower to respond to Chinese characters which appeared with English words than ungrouped learners after sleep. At the same time, grouped learners made less errors for Chinese characters pairing with English words overall than ungrouped learners. However, in the experiment about the S+O- grouping (Experiment 3), when the effect of orthography was eliminated, our findings challenged the view of an inhibitory effect of semantic grouping. Our finding backed up the effectiveness of semantic grouping in the learning of L2 vocabulary.

Previously researchers like Hashemi and Gowdasiaei (2005) claimed that semantic grouping generally benefits learners of intermediate or higher levels. Our findings showed that grouped learning could result in better performance compared to ungrouped learning at the very initial stage of L2 character learning. Grouped learners were either less error prone or faster than ungrouped learners for the lexical judgement task and the priming task. However, it shall be pointed out that only last piece of finding in Table 6.1 is conclusive, while the other findings were only tendencies with marginal or no significant difference from post hoc analysis.

Table 6.1

Summary of Key Findings of the Effect of Semantic Grouping among Matching Trials.

	Learning Materials	Task	Main findings
			Effects of [S+O+] grouping
Experiment 1: Semantic radical grouping	Five sets of four characters with shared radicals	Lexical judgement task	Compared to ungrouped learners, a. grouped learners tended to made more errors when responding to Chinese characters appeared with Chinese sounds. b. grouped learners tended less errors when responding to Chinese characters appeared with English words. c. grouped learners showed tendency of being slower when responding to Chinese characters appeared English words after sleep.
			Effects of [S+O-] grouping
Experiment 3: Pseudo-characters grouping	Five sets of four pseudo-characters without shared radicals	Lexical judgement task	Grouped learners tended made fewer errors than ungrouped learners after sleep, while all participants made more error after sleep.
		Priming task	Grouped learners were significantly faster than ungrouped learners overall.

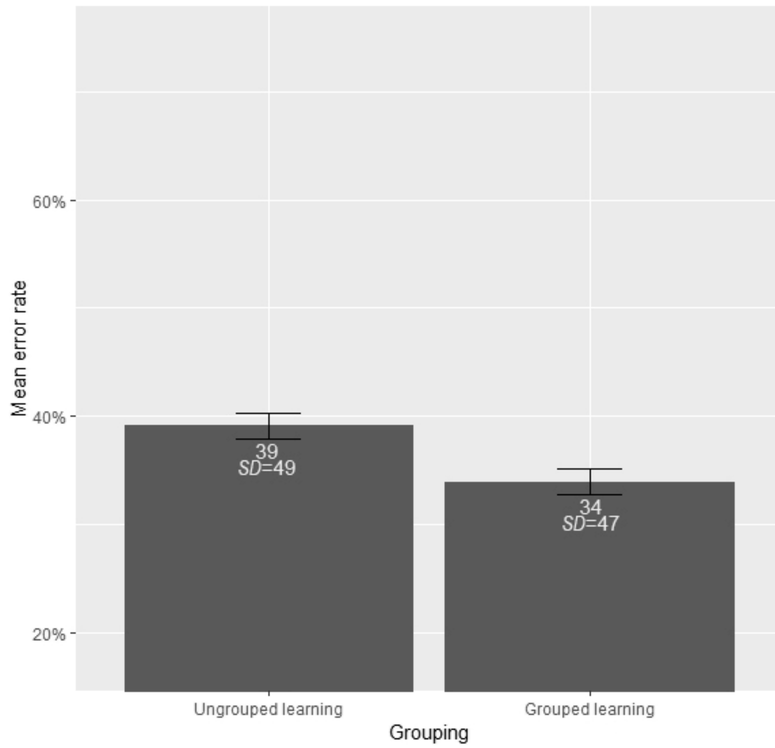


Figure 6.1. Error rates result of main effect of Grouping, Testing sessions 1 and 2, S+O-grouping experiment (Experiment 4).

However, the question arises: why did results of the two-way of grouping differ dramatically? Was it due to the removal of semantic radicals in the S+O- grouping? Was it due to the increasing input of learning? Or was there a transition point when the effect turned from inhibitory to facilitatory? In order to answer these questions, we carried out a further analysis of the first two testing sessions of Experiment 3. Using the same models of data analysis for the semantic radical grouping experiment, we only analysed results of the first two testing sessions of the S+O- grouping experiment (for further details, see Tables A.1 and A.2 in Appendix F, pp. 218-220).

We found a marginal significant main effect of Grouping (p = .08) in the main model for error rates. As seen in Figure 6.1, grouped learners were less error prone than ungrouped learners. The effect of semantic grouping without semantic radicals was proven to be facilitatory even only with the first two testing session analysed. While the other parameters stayed the same, the difference between results of semantic radical grouping experiments and results of the first two sessions of the pseudo-character grouping experiment is therefore attributable to the removal of the shared semantic radical. The inhibitory effect of semantic grouping by shared semantic radicals was due to the confound of the semantic radical which is shared among characters in each semantic category. The interference was also reflected from the interview. Many participants claimed that they could identify the radical clearly and linked the orthography (the radical) to the semantics, while neglected the meaning of individual characters.

Therefore, we conclude that the effect of semantic grouping on the initial stage of learning L2 Chinese characters is facilitatory.

Phonological Grouping: Summaries and Implications

Table 6.2 (p. 189) summarises the significant effects of phonological grouping, including Experiment 2 (phonological grouping by homophone) and Experiment 4 (phonological grouping by rhyme).

The two experiments presented different effects of phonological grouping on the learning of first-shot learning of Chinese characters. Previous studies (Rastle & Brysbaert, 2006) showed that primed by (pseudo-)homophones led to confusion and therefore had inhibitory priming effect during lexical decision. However, results of homophone grouping provide evidence for the beneficial nature of phonological grouping for learning.

Table 6.2

Summary of Key Findings of the Effect of Phonological Grouping.

	Learning Materials	Task	Main findings
			Effects of homophone grouping
Experiment 2: Grouping by homophone	Five sets of four homophones	Lexical judgement task	After sleep, grouped learners were faster than ungrouped learners overall. After sleep, ungrouped learners made more errors in responding to Chinese characters and English words that were not correctly matched than grouped learners.
			Effects of rhyme grouping
Experiment 4: Grouping by rhyme	Five sets of four rhyming characters	Lexical judgement task	Grouped learners made more mistakes when more learning sessions were given compared to ungrouped learners. Grouped learners were slower in responding to C-E pairings than ungrouped learners.
		Priming task	No clear evidence of rhyme grouping.

In the homophone grouping experiment, when grouped participants learnt homophones, they were less error faster than ungrouped learners overall. After sleep, grouped learners also made much fewer errors than ungrouped learners in responding to mismatched Chinese characters with English words compared to ungrouped learners. These findings all provided new evidence for the view that homophone grouping benefits learners. A possible explanation of such facilitatory effect could be the reduced number of phonological representations. The limited number of phonological representations helped to reduce the workload of learners and therefore made it easier for participants to contrast these sounds. It was quite likely that presenting homophones in groups helped grouped learners focus more on the difference among characters of the same set. As suggested by Perfetti's lexical constituency model (2005), phonology is a constituent when identifying words. Fully specified phonology helps to identify the word in general. In our experiment of homophone grouping, the number of phonological representations was limited. Participants were more likely to develop a fully specified phonological representations for these characters and link it with existing semantics. As a consequence, a robust link could be developed through that process.

Results of some studies like Szmalec et al (2009) show that repeated syllables among words contributes to better recall after learning. Similarly, Wilcox and Medina (2013) showed that shared onsets led to better performance in the short and long term. However, findings in rhyme grouping challenged the facilitatory effect of phonological grouping by a shared segment such as a phoneme or a rhyme. That could be due to the following reasons. Firstly, representations of rhyming neighbours could be easily be coactivated for grouped learners. During learning, rhyming neighbourhoods were learnt together for grouped learners. The confusion of similar sounds during learning could hinder them from building specified phonological representations. On the contrary, ungrouped learners had the opportunity to build a well specified phonological representations during learning, as those sounds learnt in a set differed. Secondly, participants were completely unfamiliar with Chinese sounds. Because of this reason, learning similar sounds together could lead to further confusion among characters in the same set. As ungrouped learners learnt five sounds without overlaps, they were more likely to get a better learning results compared to grouped learners.

It is worthy pointing out that the number of sounds differed between the homophone grouping experiment (Experiment 1) and the rhyme grouping experiment (Experiment 1). The former only had five sounds while the latter had 20. The growing number of sounds could also result in inferior learning results in Experiment 4 when being compared to Experiment 1.

However, results of Experiment 4 showed participants maintained low average error rates overall, while in Experiment 1, error rates were at a higher level.

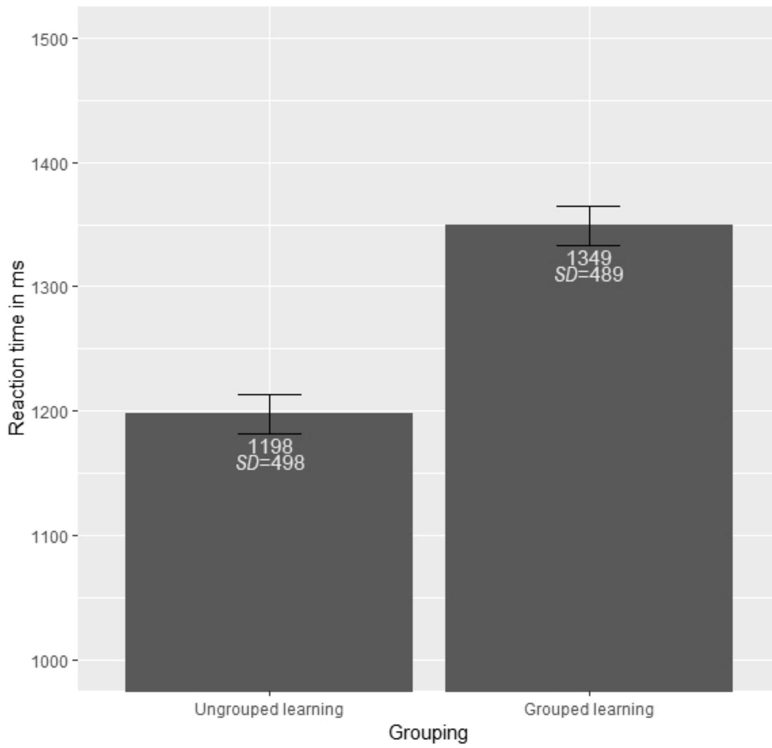


Figure 6.2. Inhibitory effect of rhyme grouping in the first two testing sessions, rhyme grouping experiment.

There were more learning input sessions for rhyme grouping experiment (three sessions) when compared to homophone grouping (one session only). It could be possible therefore that there is a point during learning when the effect of phonological grouping switched to inhibitory from facilitatory. Therefore, it is necessary to carry out an analysis of results of the first two testing sessions in the rhyme grouping experiment (Experiment 4). We used the same model of the analysis for the homophone grouping data for the analysis of the first two testing sessions of the rhyme grouping experiment (for further details, see Tables A.3 and A.4 in Appendix F, pp. 220-222).

We only find a significant main effect of Grouping (p = .02) for the reaction time data overall and a significant main effect of Grouping (p = .02) for the Chinese-English pairings. Grouped learners were significantly slower than ungrouped learners especially for Chinese-English pairings. The first two sessions clearly showed that the effect of phonological grouping by rhyming was inhibitory, which was identical to pattern of the overall data. Increase in learning session did not lead to a switch point of the effect of phonological grouping whatsoever. It is true that the increasing number of phonological representations could affect the quality of learning. Nevertheless, based on our findings, when compared with ungrouped learning, learning Chinese characters in homophones helps participants form well specified lexical representations while learning in rhyming sets, does not.

Towards the Timeline of Initial L2 Chinese Character Learning

From the pseudo-character grouping experiment (Experiment 3) and the rhyme grouping experiment (Experiment 4), we collected solid evidence for the well formation of lexical representations as well as links between forms and meaning after only three learning input sessions. Namely, in pseudo-character grouping experiment, we found that on Day 7, links from orthographic representations to semantic representations were well formed as the accuracy of the categorisation task was increased. The phonological representations of newly learnt characters were also well specified based on the high accuracy of trials of Chinese characters and Chinese sounds pairing as well as the Chinese characters and English words pairings.

In the homophone grouping experiment, orthographic representations were better formed than phonological representations as all learners made fewer errors in responding to Chinese characters and English words, rather than to Chinese characters to Chinese sounds. This also means that the link from orthographic representations to semantics were more robust than the link from phonological representations to semantics. Moreover, on Day 7, phonological representations for individual characters were well formed as regardless of group, all participants were faster and made fewer errors for pairings of Chinese characters and Chinese sounds than pairings of Chinese characters with English words.

In both experiments, we found that on Day 2, there was either a drop in error rates or in latencies for trials of Chinese characters with English words and another one on Day 3 for the Chinese characters-Chinese sound pairings.

On Day 2, for the pseudo-character experiment, all participants made fewer errors for Chinese-English pairings compared to Day 1, while there was a drop in reaction time among Chinese-English pairings in the rhyme grouping experiment.

On Day 3, all participants made fewer errors for the Chinese-Chinese pairings in the pseudo-character grouping experiment. Meanwhile in the rhyme grouping experiment, participants became faster in Chinese-Chinese trials. That suggested the formation of phonological representation took longer than the establishment of the link between the character and the English meaning. As participants were not familiar with Chinese phonology at all, they probably need more time to form the phonological representation and link it to the existed semantics. There was a significant decrease in reaction time on Day 2 for matching trials for both groups.

The Consolidation Effect of Sleep

In this thesis, we took it into account and modelled the effect of sleep consolidation. In most cases, effects of grouping became significant only after sleep. Major findings of the sleep consolidation effect in the four experiments are summarised in Table 6.3.

Table 6.3

Summary of Key Findings of Experiments 1 to 4 in Terms of the Consolidation Effect of Sleep.

Task	Grouping	Main findings
Lexical judgement task	Semantic radical	After sleep, all participants made fewer errors.
	Homophone	All participants became faster in responding to correctly matched Chinese characters and English words after sleep.
	Pseudo-character	Sleep did not contribute to a higher accuracy or a shorter latency. After sleep, difference in error rates between grouped and ungrouped learners was larger after sleep.
	Rhyme	Sleep did not contribute to a higher accuracy or a shorter latency. After sleep, decrease in latencies for grouped learners was significantly larger than that for ungrouped learners, although grouped learners spent longer time than ungrouped learners overall.

Our findings in these experiments for effects of semantic radical and homophone were consistent with the view that the sleep consolidation effect for lexicalisation could be applied into L2 vocabulary learning (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Mirkovic & Gaskell, 2016). Mostly, the effect of Grouping became significant after sleep, while the performance of both groups was improved to some extent. However, the effects of sleep on pseudo-character grouping and rhyme grouping showed that the sleep did not result in a closer integration into lexical competition. Therefore, we may ask that what sleep “consolidates” is not limited to the vocabulary and the integration into lexicon. Previous literature suggested that sleep also consolidate procedural information and motor skills (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Maquet, 2001). What we observed in our experiment might provide a broader scope of the sleep consolidation effect, which is worth more in-depth research.

Limitations and Suggestions for Future Studies

The major limitation of the present study was the size of sample. As shown from results of analysis, the difference between grouped learners and ungrouped learners were too week to detect. Sometimes, it only showed tendencies of reaching significance. The small size of sample could result in weakness of power to detect difference between two groups. There were around 40 participants for each experiment. For a moderate effect (effect size of .30), this size of sample could only result in the power of .50. For a power of .80, the sample size should be more than 46 (for the effect size of .40) or .85 (for the effect size of .30). Sizes of sample of three of all the four experiments were smaller than this limit. It means the present study was under powered to some extent.

Due to the limited time, the present study did not include *Pinyin,* the official romanisation of Chinese language, in the design. *Pinyin* as a tool to present the pronunciation to learners of Chinese language is fairly common in the classroom context. In the interview, some participants claimed that the prounciation sounded “the same” during the learning phase. Therefore, would it lead to a better result if *Pinyin* was included in the learning phase? The experiment has been completely designed and wait to be carried out now.

Of course, to teach characters based on same/similar sounds and/or semantic category to complete novoice learners was only a laborartory imitation of the learning process. However, such grouping methods have distinctive features: prosody could be found in homophone/rhyme grouping while the semantic radical could help learners guess the meaning of unknown characters. In the real world, many computer-based Chinese learning course are available online or offline. Many of courses designed for beginning L2 Chinese learners are in traditional practical approprach with dialogues according to scenes. For example, global online course provider Coursera provides a 11-hr recorded course in cooperation with Peking University[9] starting from simple dialogues and pinyin. The course focuses on everyday use while the reading and learning of Chinese vocabulary were not emphasised. It is well acknowledged that boosting vocabulary is one of keys to the suceess in learning a new language. While trying to remember new vocabulary could be a tedious job, learning them in semantic/phonological group could be a useful supplement to the traditional approach. Findings of the present study could also be applied into the field of teaching Chinese for special purposes. For example, learning programme could be designed for those people with no or little knowledge of the language but need to read Chinese literature in a particular field. Characters used in a field have semantic relations and therefore could be grouped in semantic sets. Learning those characters could help them locate useful information and avoid unecessary cost in time and money for translation.

Conclusions

This thesis for the first time fully examined the effect of grouping on the initial learning of Chinese characters with empirical evidence. Learning Chinese characters in semantic group

[9] For further details, please visit https://www.coursera.org/learn/learn-chinese.

without semantic radical and in phonological group by homophones contributes to better learning results compared to ungrouped learning. Grouping in semantic category with shared radicals or in rhyming sets inhibits the specification of representations. Future studies should focus on the time-course of the building of representations and the role of *Pinyin* on the learning of Chinese characters.

LIST OF REFERENCES

Acheson, D. J., & MacDonald, M. C. (2011). The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language*, *65*(2), 193–207. https://doi.org/10.1016/j.jml.2011.04.006

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, (15), 17–29.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240–247.

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.)*, The science of reading: A handbook* (pp. 6–23). Oxford: Blackwell. https://doi.org/10.1002/9780470757642.ch1

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*. https://doi.org/10.1037/0033-295X.100.4.589

Damian, M. F., & Martin, R. C. (1998). Is visual lexical access based on phonological codes? Evidence from a picture-word interference task. *Psychonomic Bulletin & Review*, *5*(1), 91–95. https://doi.org/10.3758/BF03209461

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition, 81*(3), 77–86. https://doi.org/10.1016/S0010-0277(01)00135-4

Davis, C. J. (2005). N-watch: a program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70. https://doi.org/10.3758/BF03206399

de Groot, A. M., & Keijzer, R. (2000). What is hard to learn is easy to forget: the roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110

DeFrancis, J. (1989). *Visible Speech: The Diverse Oneness of Writing Systems*. Honolulu: University of Hawaii.

Diependaele, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the bimodal interactive activation model. *European Journal of Cognitive Psychology*, *22*(5), 764–778. https://doi.org/10.1080/09541440902834782

Dumay, N., & Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition*, *123*(1), 119–132. https://doi.org/10.1016/j.cognition.2011.12.009

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words: Research report. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Duyck, W., & Brysbaert, M. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, *13*(03), 359–371. https://doi.org/10.1017/S1366728909990344

Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, *36*(3), 407–422. https://doi.org/10.1016/j.system.2008.02.005

Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society*, *31*(2), 119–122.

Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *47*(2), 365–382. https://doi.org/10.1080/14640749408401116

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, *24*(03), 369–383. https://doi.org/10.1017/S0142716403000195

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698. https://doi.org/10.1037/0278-7393.10.4.680

Frisson, S., Bélanger, N. N., & Rayner, K. (2014). Phonological and orthographic overlap effects in fast and masked priming. *The Quarterly Journal of Experimental Psychology*, *67*(9), 1742–1767. https://doi.org/10.1080/17470218.2013.869614

Fu, I. P. (2005). Student Approaches to Learning Chinese Vocabulary By Student Approaches to Learning Chinese Vocabulary. *Faculty of the Virginia Polytechnic Institute and State University*.

Gairns, R., & Redman, S. (1986). *Working with words: A guide to teaching and learning vocabulary*. New York: Cambridge University Press.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1994.1011

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, *3*(1), 128–156.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*(3), 518–565.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighbourhood frequency effect. *Perception and Psychophysics*, *45*(3), 189–195. https://doi.org/10.3758/BF03210696

Han, J. (2012). *Chinese Characters* (G. Wang & L. Zhou, Trans.). Cambridge: Cambridge University Press.

Hashemi, M. R., & Gowdasiaei, F. (2005). An attribute-treatment interaction study: Lexical-set versus semantically-unrelated vocabulary instruction. *RELC Journal*, *36*(3), 341–361. https://doi.org/10.1177/0033688205060054

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: John Wiley & Sons.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Oxford: Blackwell.

Henning, G. H. (1973). Remembering foreign language vocabulary: Acoustic and semantic parameters. *Language Learning*, *23*(2), 185–196. https://doi.org/10.1111/j.1467-1770.1973.tb00654.x

Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, *2*(2), 170–175. https://doi.org/10.1016/S0022-5371(63)80082-1

Hoshino, Y. (2010). The categorical facilitation effects on L2 vocabulary learning in a classroom setting. *RELC Journal*, *41*(3), 301–312. https://doi.org/10.1177/0033688210380558

Jacobs, A., & Grainger, J. (1994). Models of visual word recognition - Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1311–1334.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jin, Z., Lee, J., & Lee, Y. (2013). The phonological process with two patterns of simplified chinese characters. *Research in Language*, *11*(4), 389–403. https://doi.org/10.2478/rela-2013-0005

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, (265), 679–682.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: the role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory, Volume 2* (pp. 389–395). London: John Wiley & Sons.

Kroll, J. F., & Hermans, D. (2011). Psycholinguistic perspectives on language processing in bilinguals. *Modeling Bilingualism: From Structure to Chaos : In Honor of Kees de Bot*, *43*, 15–36. https://doi.org/10.1075/bjl.25.04nov

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches.* (pp. 531–553). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199540914.003.0011

Kroll, J. F., Van Heel, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognitionn*, *13*(3), 373–381. https://doi.org/10.1017/S136672891000009X

Kwok, R. K. W., & Ellis, A. W. (2015). Visual word learning in skilled readers of English. *Quarterly Journal of Experimental Psychology*, *68*(2), 326–349. https://doi.org/10.1080/17470218.2014.944549

Lee, H. W., Rayner, K., & Pollatsek, A. (1999). The time course of phonological, semantic, and orthographic coding in reading: Evidence from the fast-priming technique. *Psychonomic Bulletin and Review*, *6*(4), 624–634. https://doi.org/10.3758/BF03212971

Lee, Y. A., Binder, K. S., Kim, J. O., Pollatsek, A., & Rayner, K. (1999). Activation of phonological codes during eye fixations in reading. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(4), 948–964. https://doi.org/10.1037/0096-1523.25.4.948

Lesch, M. F., & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 285–294. https://doi.org/10.1037/0278-7393.19.2.285

Liu, Y., Dunlap, S., Fiez, J., & Perfetti, C. (2007). Evidence for neural accommodation to a writing system following learning. *Human Brain Mapping*, *28*(11), 1223–1234. https://doi.org/10.1002/hbm.20356

Liu, Y., Perfetti, C. A., & Wang, M. (2006). Visual analysis and lexical access of Chinese characters by Chinese as second language readers. *Language and Linguistics*. https://doi.org/2006-0-007-003-000171-1

Liu, Y., Wang, M., & Perfetti, C. A. (2007). Threshold-style processing of Chinese characters for adult second-language learners. *Memory & Cognition*, *35*(3), 471–480. https://doi.org/10.3758/BF03193287

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is intially phonological: I. Evidence from associativepriming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, *123*(2), 107–128.

Lukatela, G., & Turvey, M. T. (1996). Inhibition of naming by rhyming primes. *Perception & Psychophysics*, *58*(6), 823–835. https://doi.org/10.3758/BF03205486

Maquet, P. (2001). The role of sleep in learning and memory. *Science*, (294), 1048–1052.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature*, *375*(6619), 813–815.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, (7), 703–704.

Mirkovic, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PLoS ONE*, *11*(4). https://doi.org/10.1371/journal.pone.0152489

Nation, I. S. P. (1982). Beginning to learn foreign vocabulary: A review of the research. *RELC Journal*, *13*(1), 14–36.

Nation, P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, *9*(2), 6–10. https://doi.org/10.1002/j.1949-3533.2000.tb00239.x

Nelson, J., Liu, Y., Fiez, J., & Perfetti, C. A. (2005). Learning to read Chinese as a second language recruits Chinese-specific visual word form areas. In *Paper presented at the Society for the Scientific Study of Reading Conference*. Toronto.

Ojima, S., Nakamura, N., Matsuba-Kurita, H., Hoshino, T., & Hagiwara, H. (2011). Neural correlates of foreign-language learning in childhood: a 3-year longitudinal ERP study. *Journal of Cognitive Neuroscience*, *23*, 183–199. https://doi.org/10.1162/jocn.2010.21425

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173

Perfetti, C. A., & Bell, L. (1991). Phonemic Activation during the First 40 ms of Word Identification : Evidence from Backward Masking and Priming. *Journal of Memory and Language*, *30*, 473–485.

Perfetti, C. A., & Liu, Y. (2005). Orthography to phonology and meaning: Comparisons across and within writing systems. *Reading and Writing*, *18*(3), 193–210. https://doi.org/10.1007/s11145-004-2344-y

Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L.-H. (2007). Reading in two writing systems: Accommodation and assimilation of the brain's reading network.

Bilingualism: Language and Cognition, *10*(02), 131. https://doi.org/10.1017/S1366728907002891

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., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 101–118. https://doi.org/10.1037/0278-7393.24.1.101

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

Qu, Q., & Damian, M. F. (2017). Orthographic effects in spoken word recognition: Evidence from Chinese. *Psychonomic Bulletin & Review*, *24*(3), 901–906. https://doi.org/10.3758/s13423-016-1164-9

R Development Core Team (2011). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, Retrieved from http://www.R-project.org/

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology*, *53*(2), 97–145. https://doi.org/10.1016/j.cogpsych.2006.01.002

Rayner, K., Pollatsek, A., & Schotter, E. R. (2013). Reading: Word identification and eye movements. In A. F. Healy, R. W. Proctor, & I. B. Weiner (Eds.), *Handbook of psychology: Experimental psychology* (pp. 548–577). Hoboken, NJ, US: John Wiley & Sons Inc.

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 787–799. https://doi.org/10.1002/wcs.68

Seal, B. D. (1991). Vocabulary learning and teaching. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (2nd ed., pp. 296–311). Boston: Heinle & Heinle.

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

Sereno, S. C., & Rayner, K. (1992). Fast priming during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 173–184. https://doi.org/10.1037/0096-1523.18.1.173

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *NeuroImage*, *33*(2), 774–783. https://doi.org/10.1016/j.neuroimage.2006.07.008

Suárez-Coalla, P., & Cuetos, F. (2017). Semantic and phonological influences on visual word learning in a transparent language. *Quarterly Journal of Experimental Psychology*, *70*(4), 772–781. https://doi.org/10.1080/17470218.2016.1164733

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, *54*(2), 99–132. https://doi.org/10.1016/j.cogpsych.2006.05.001

Szmalec, A., Duyck, W., Vandierendonck, A. A., Mata, A. B. B., & Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Quarterly Journal of Experimental Psychology*, *62*(918399474), 435–443. https://doi.org/10.1080/17470210802386375

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second language vocabulary. *Bilingualism: Language and Cognition*, *2*(1), 45–58.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience*, *30*(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Tan, L. H., Hoosain, R., & Siok, W. W. T. (1996). Activation of phonological codes before access to character meaning in written Chinese. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(4), 865–882. https://doi.org/10.1037/0278-7393.22.4.865

Tan, L., & Perfetti, C. (1997). Visual Chinese character recognition: Does phonological information mediate access to meaning? *Journal of Memory and Language*, *37*(37), 41–57. https://doi.org/10.1006/jmla.1997.2508

Tang, L. 唐兰 (1981). *Guwenzixue daolun* [古文字学导论]. Jinan: Qilu Press. (Original work published 1934)

Tang, L. 唐兰 (2005). *Zhongguo wenzixue* [中国文字学]. Shanghai: Shanghai Ancient Works Publishing House, Shanghai Century Publishing Group. (Original work published 1949)

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, *21*(3), 371–380. https://doi.org/10.1016/0346-251X(93)90027-E

Tinkham, T. (1997). The effect of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*(2), 138–163. https://doi.org/10.1016/0346-251X(93)90027-E

Tongyong guifan hanzi biao [通用规范汉字表]. (2013). Beijing: Language & Culture Press.

Wang, H.-C., Savage, G., Gaskell, M. G., Paulin, T., Robidoux, S., & Castles, A. (2017). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review*, *24*(4), 1186–1193. https://doi.org/10.3758/s13423-016-1182-7

Wang, M., Perfetti, C. A., & Liu, Y. (2003). Alphabetic Readers Quickly Acquire Orthographic Structure in Learning to Read Chinese. *Scientific Studies of Reading*. https://doi.org/10.1207/S1532799XSSR0702

Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, *25*(2), 261–274. https://doi.org/10.1016/S0346-251X(97)00013-4

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*(2), 439–456. https://doi.org/10.1080/713755710

Wilcox, A., & Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System*, *41*(4), 1056–1069. https://doi.org/10.1016/j.system.2013.10.012

Wilkinson, K., & Mazzitelli, K. (2003). The effect of "missing" information on children's retention of fast-mapped labels. *Journal of Child Language*, *30*(1), 47–73. https://doi.org/10.1017/S0305000902005469

Xu, S. 许慎 (Comp.), Xu, X. 徐铉 Annotated (1963). *Shuowen jiezi* [说文解字]. Beijing: Zhonghua Book Company. (Original work published 121)

Yin, J. J. (2006). *Fundamentals of Chinese Characters*. New Haven, Connecticut: Yale University Press. https://doi.org/10.1126/science.79.2057.503

Yum, Y. N., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2014). An ERP study on initial second language vocabulary learning. *Psychophysiology*, *51*(4), 364–373. https://doi.org/10.1111/psyp.12183

Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A., & Grainger, J. (2000). Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *The Quarterly Journal of Experimental Psychology*, *53Q*(3), 671–692.

APPENDIX A: PARTICIPANT CONSENT FORM

Psycholinguistics Lab: Participant Consent Form

Experiment Name: ___________

Name: ____

First Language:____
(Please note: this refers to the language you learned first as a child, not to the language you currently use most frequently)

Date:___________________________

I have read the experimental instructions. I am here out of my own free will, and I understand that I can leave at any moment during the experiment without explaining myself to the experimenter. I also understand that my data will be collected, stored and analysed anonymously, in accordance with the Data Protection Act. The computer-generated result files will bear ID numbers only. It will not be possible to link my ID number to the individual participants. Consent forms will be stored separately from the experimental results and will NOT bear the ID number. My data may be made available to other researchers in appropriate archives. If you wish to withdraw your data from this study this will be possible if the withdrawal request is made within 8 weeks of your participation. To withdraw your data please contact the researcher using the e-mail address above.

Signature: ______________________________

APPENDIX B: LANGUAGE BACKGROUND QUESTIONNAIRE

QUESTIONNAIRE

Please provide the following information about yourself (tick the appropriate box):

All information that you provide is confidential in accordance with the Data Protection Act (1998) and The British Psychological Society Code of Ethics and Conduct (2006).

Gender: Female [ ] Male [ ]

Dominant hand: Left [ ] Right [ ]
(that you usually write with)

Age (please specify): ……… yrs

Native British English speaker: Yes [ ] No [ ]

Are you fluent in any other Languages?

Yes [ ] No [ ]

If yes – please state which languages:

……………………………………………………………………………

Have you ever been diagnosed with a speech or language disorder?

Yes [ ] No [ ]

Have you been diagnosed with dyslexia?

Yes [ ] No [ ]

Thank You

Experiment: Date:

APPENDIX C: INSTRUCTIONS FOR EXPERIMENTS 1 AND 2

Last updated on 16 September, 2014
Edited by Kristian Suen

Researcher: Kristian Suen
e-mail: [redacted]

Study: **Learning Chinese characters**

To take part in this study you should be a native speaker of English with good reading and speaking skills and no known reading disorders. Your data will be stored anonymously and archived on CD-ROM. The computer-generated result files will bear ID numbers only. It will not be possible to link the ID numbers to the individual participants. Consent forms will be stored separately from the experimental results and will NOT bear the ID number. Your data may be made available to other researchers in appropriate archives. If you wish to withdraw your data from this study this will be possible if the withdrawal request is made within 8 weeks of your participation. To withdraw your data please contact the researcher using the e-mail address above. We are happy to answer any questions you have about the study both before and at any time after your participation.

Instructions

The aim of this experiment is to test your learning of Chinese characters. The experiment will have a learning phase and a testing phase. During the learning phase, you will see a Chinese character and see the written English translations on the screen. You will also hear the Chinese word through the headphones. Please try to learn the relationship between the Chinese characters and their sounds and meaning as well as you can.

The stimuli will be presented in sets of 4. You can look at them one at a time for as long as you like and you can repeat each set as often as you like in order to learn them. Then you can move on to the next set of four. You cannot go back to a previous set. There are 5 sets of four in total.

After a short break, you will then be tested on your learning by being asked to judge whether a character-Chinese word or character-English word combination is correct or incorrect. You will make your judgment by pressing buttons.

- Press the YES button for a correct pairing
- Press the NO button for an incorrect pairing.

Please use your dominant hand for the yes button.
Please press only once – don't try to correct mistakes.

continued overleaf

Each trial will start with a fixation cross (+) in the centre of the screen, shortly before the presentation of each stimulus. You will be given examples of trials before we begin the testing session.

There will be 8 blocks of 20 trials in the testing phase divided and there will be a pause between the blocks. The experimenter will check with you that you are ready to proceed to the next block. You may, of course, leave the experiment at any point. We will start with a practice block of 4 trials to get you used to the task.

Please respond as quickly and as accurately as you can.

APPENDIX D: INTERVIEW QUESTIONS FOR EXPERIMENTS 1 AND 2

Day 1

1. Which task during the testing session do you think is more difficult?

   ☐ the “Character + Pronunciation” matching task

   ☐ the “Character + Translation” matching task

2. Which learning strategies did you use in order to learn those characters?
3. Do you have any other comments?

Day 2

1. Which task during the testing session do you think is more difficult?

   ☐ the “Character + Pronunciation” matching task

   ☐ the “Character + Translation” matching task

2. Which learning strategies did you use in order to learn those characters?
3. Do you have any other comments?

APPENDIX E: INSTRUCTIONS FOR EXPERIMENTS 3 AND 4

Last updated on 05 October, 2015
Edited by Kristian Suen

Researcher: Kristian Suen
e-mail: [redacted]

Study: **Learning Chinese characters**

To take part in this study you should be a native speaker of English with good reading and speaking skills and no known reading disorders. Your data will be stored anonymously and archived on CD-ROM. The computer-generated result files will bear ID numbers only. It will not be possible to link the ID numbers to the individual participants. Consent forms will be stored separately from the experimental results and will NOT bear the ID number. Your data may be made available to other researchers in appropriate archives. If you wish to withdraw your data from this study this will be possible if the withdrawal request is made within 8 weeks of your participation. To withdraw your data please contact the researcher using the e-mail address above. We are happy to answer any questions you have about the study both before and at any time after your participation.

Instructions

The aim of this experiment is to test your learning of Chinese characters. The experiment will have a learning phase and a testing phase. During the learning phase, you will see a Chinese character and see the written English translations on the screen. You will also hear the Chinese word through the headphones. Please try to learn the relationship between the **Chinese characters** and their **sounds** and **meaning** as well as you can. You do NOT need to pronounce the sound.

The stimuli will be presented in sets of 4. You can look at them one at a time and hear the pronunciation 6 times. You can learn each set twice before you move on to the next set. There are 5 sets of four in total.

After a short break, you will then be tested on your learning by being asked to judge whether a character-Chinese word or character-English word combination is correct or incorrect. You will make your judgment by pressing buttons.

- Press the YES button for a correct pairing
- Press the NO button for an incorrect pairing.

Please use your dominant hand for the yes button.

Please press only once – don’t try to correct mistakes.

continued overleaf

Each trial will start with a fixation (—) in the centre of the screen, shortly before the presentation of each stimulus. You will be given examples of trials before we begin the testing sessions.

There will be 8 blocks of 20 trials in the testing phase divided and there will be a pause between the blocks. The experimenter will check with you that you are ready to proceed to the next block. You may, of course, leave the experiment at any point. We will start with a practice block of 10 trials to get you used to the task.

Please respond as quickly and as accurately as you can.

For Day 2/7

Today your learning will also be tested by judging whether two Chinese characters belongs to the same semantic category or not. You will also make your judgment by pressing buttons. The two characters will be presented together. There will be 6 blocks of 20 trials in this test. The rest requirements are similar to previous experiments.

- Press the YES button if they are related in meaning
- Press the NO button if they are not related in meaning

Please use your dominant hand for the yes button.

Please press only once – don’t try to correct mistakes

Please respond as quickly and as accurately as you can.

For Day 7

The new task today is generally the same with the earlier combination judgement task, expect for that you are asked to response to the character in the frame only.

Each trial will start with a fixation (—) in the centre of the screen. Then you will see an unframed character before the presentation of the framed stimulus.

You should response to the framed character only.

Please press only once – don’t try to correct mistakes

Please respond as quickly and as accurately as you can.

[The End]

APPENDIX F: ANALYSIS RESULTS OF THE FIRST TWO SESSIONS OF EXPERIMENTS 3 AND 4

Tables A.1 and A.2 summarise the data of first two testing sessions of the pseudo-character grouping.

Table A.1

Summary for Fixed Effects in the Mixed Logit Model for Error Rate of the First Two Sessions of Testing Tasks in Experiment 3.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-0.6	0.10	-6.52	< .001***
Grouping	0.14	0.08	1.72	.08†
Pairing	0.39	0.07	5.48	< .001***
Matching	-0.36	0.08	-4.58	< .001***
Sleep	0.02	0.02	1.03	.30
Grouping by Pairing	-0.03	0.05	-0.58	.56
Grouping by Matching	-0.04	0.05	-0.65	.52
Pairing by Matching	-0.13	0.02	-6.13	< .001***
Grouping by Sleep	-0.03	0.02	-1.11	.26
Matching by Sleep	0.05	0.02	1.89	.06†
Grouping by Pairing by Sleep	0.04	0.02	1.72	.08†
Model for matching trials[b]				
Intercept	-0.30	0.14	-2.14	.03*
Grouping	0.13	0.08	1.57	.12
Sleep	-0.02	0.04	-0.40	.69
Pairing	0.52	0.08	6.42	< .001***
Grouping by Sleep	-0.04	0.03	-1.22	.22
Sleep by Pairing	-0.03	0.03	-1.23	.22
Model for mismatching trials[c]				
Intercept	-1.08	0.12	-8.75	< .001***
Grouping	0.10	0.09	1.10	.27
Sleep	0.07	0.03	2.21	.03*
Pairing	0.30	0.10	2.85	.004**
Grouping by Sleep	-0.01	0.03	-0.24	.81
Sleep by Pairing	0.02	0.03	0.68	.50

Notes. [a]*N* = 12,480, log-likehood = -7,342.1. The random effect structure included an intercept and slopes of Grouping, Pairing, and Sleep by Matching for items, and an intercept and slopes of Pairing, Matching, and Sleep for subjects. It was the maximal random effect structure that reached convergence. A model excluding the interaction Grouping by Category did not fit better than a model including this interaction ($\chi 2(1) = 3.08$, $p = .08$).

[b]*N* = 6,240, log-likehood = -3,831.7. In random effect structure, we included a random intercept for items and subjects, and random slopes of Grouping, Sleep, and Pairing for items, and slopes of Pairing and Sleep by subjects. That was the maximal structure in which convergence was reached.

[c]*N* = 6,240, log-likehood = -3,459.3. The random effect structure which reached convergence consisted of a random intercept for subjects and items, as well as random slopes of Pairing and Sleep for subjects, and slopes of Grouping, Sleep, and Pairing for items.

[†]*p*. < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table A.2

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time in the First Two Sessions of Testing Tasks in Experiment 3.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1339.52	38.20	43.24	35.07	< .001***
Grouping	19.39	33.08	37.28	0.59	.56
Pairing	164.38	14.58	48.14	11.28	< .001***
Matching	-15.90	9.90	26.75	-1.61	.12
Sleep	109.43	12.28	37.69	8.92	< .001***
Pairing by Matching	-22.32	8.15	27.53	-2.74	.01*
Pairing by Sleep	7.06	4.28	7703.40	1.65	.10[†]

Note. [a]*N* = 7,912. For the random effect structure of the model of all trials, we included a random intercept for items and a random intercept for subjects, together with random slopes of Grouping, Sleep, and Pairing by Matching for items, and slopes of Sleep and Pairing by Matching for subjects.

†p. < .1. *p < .05. ***p < .001.

The data of the first two sessions of the thyme grouping experiment are summarised in Tables A.3 and A.4.

Table A.3

Summary for Fixed Effects in the Mixed Logit Model for Error Rate of the First Two Sessions of Testing Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*Wald Z*	*p*
Model for all trials[a]				
Intercept	-0.85	0.13	-7.42	< .001***
Grouping	-0.03	0.10	-0.31	.76
Pairing	0.55	0.08	7.15	< .001***
Matching	-0.28	0.08	-3.68	< .001***
Sleep	0.05	0.03	1.89	.06†
Grouping by Pairing	-0.00	0.06	-0.10	.92
Grouping by Matching	-0.03	0.06	-0.54	.58
Pairing by Matching	-0.13	0.02	-5.61	< .001***
Matching by Sleep	0.10	0.02	4.49	< .001***
Grouping by Pairing by Matching	-0.06	0.02	-2.55	.01*
Model for matching trials[b]				
Intercept	-0.65	0.13	-4.95	< .001***
Grouping	0.02	0.09	0.24	.81
Pairing	0.66	0.09	7.57	< .001***
Sleep	-0.05	0.04	-1.08	.28
Model for mismatching trials[c]				
Intercept	-1.29	0.17	-7.68	< .001***
Grouping	-0.12	0.11	-1.08	.28
Pairing	0.47	0.13	3.70	< .001***
Sleep	0.16	0.05	3.37	< .001***

Notes. [a]*N* = 10,880, log-likehood = -6,055.3. The random effect structure included an intercept and slopes of Grouping, Matching, Pairing, and Sleep for items, and an intercept and slopes of Pairing, Matching, and Sleep for subjects. It was the maximal random effect structure that reached convergence.

[b]*N* = 5,440, log-likehood = -3,190.1. The random effect structure which reached convergence consisted of a random intercept for subjects and items, as well as random slopes of Pairing and Sleep for subjects, and slopes of Grouping, Pairing, and Sleep for items.

[c]*N* = 5,440, log-likehood = -2800.6. In random effect structure, we included a random intercept for items and subjects, and random slopes of Grouping, Pairing, and Sleep for items, and slopes of Pairing and Sleep by subjects. That was the maximal structure in which convergence was reached.

[†]*p*. < .1. **p* < .05. ****p* < .001.

Table A.4

Summary for Fixed Effects in the Mixed Linear Model for Reaction Time in the First Two Sessions of Testing Tasks in Experiment 4.

Predictor	Coefficient	*SE*	*df*	*t*-value	*Pr*(>|*t*|)
Model for all trials[a]					
Intercept	1303.74	41.04	36.08	31.77	< .001***
Grouping	-87.13	37.74	32.22	-2.31	.02*
Pairing	144.48	15.83	41.71	9.13	< .001***
Sleep	95.94	10.02	33.17	9.57	< .001***
Matching	-3.26	11.19	29.34	-0.29	.77
Grouping by Pairing	15.89	14.22	33.19	1.11	.28
Grouping by Sleep	-13.87	9.43	31.51	-1.47	.15
Pairing by Sleep	3.19	4.59	7237.32	0.07	.49
Pairing by Matching	-21.11	11.46	24.31	-1.84	.08[†]
Grouping by Pairing by Sleep	13.76	4.59	7236.93	3.00	.003**
Model for Chinese-Chinese pairings[b]					
Intercept	1447.10	44.95	33.20	44.96	< .001***
Grouping	-66.40	43.85	32.56	43.85	.13
Sleep	97.93	12.47	32.17	12.47	< .001***

Matching	-24.86	17.44	32.56	17.44	.15
Grouping by Sleep	0.82	11.21	28.92	11.21	0.94
Model for Chinese-English pairings[c]					
Intercept	1159.40	42.81	41.84	27.08	< .001***
Grouping	-89.08	38.81	32.22	-2.90	.03*
Sleep	95.13	10.61	29.37	8.97	< .001***
Matching	18.66	14.66	19.13	1.27	.22
Grouping by Sleep	-29.80	9.77	31.01	-3.05	.005**

Note. [a]*N* = 7,406. For the random effect structure of the model of all trials, we included a random intercept for items and a random intercept for subjects, together with random slopes of Sleep, Pairing by Matching, and Grouping by Pairing for items, and slopes of Sleep and Pairing by Matching for subjects. It was the maximal random effect structure that converged.

[b]*N* = 3,192. The random effect structure which reached convergence consisted of a random intercept for subjects and items, as well as random slopes of Pairing and Sleep for subjects, and slopes of Grouping, Pairing, and Sleep for items.

[c]*N* = 4,214. In random effect structure, we included a random intercept for items and subjects, and random slopes of Grouping, Pairing, and Sleep for items, and slopes of Pairing and Sleep by subjects. That was the maximal structure in which convergence was reached.

[†]*p*. < .1. **p* < .05. ***p* < .01. ****p* < .001.