INAUGURAL-DISSERTATION
zur Erlangung der Doktorwürde der Philosophie (Dr. phil.)

des Fachbereichs Germanistik und Kunstwissenschaften der Philipps-Universität Marburg

The influence of orthography on L1 and L2 spoken word recognition: Evidence from Chinese learners of German

Vorgelegt von

Gan, Lu

aus P.R. China

Marburg/Lahn im August 2023

Vom Fachbereich Germanistik und Kunstwissenschaften der Philipps-Universität Marburg als Dissertation angenommen am

Tag der mündlichen Prüfung/Disputation am 08.11.2023

Erstgutachter: Prof. Dr. Ulrike Domahs

Zweitgutachter: Prof. Dr. Christina Kauschke

Acknowledgments

First of all, I would like to express my very great appreciation to my main supervisor Prof. Dr. Ulrike Domahs for her valuable suggestions throughout the planning and development of this research work, from which I really have profited greatly. I appreciated that she is always patient to answer any questions of mine. I am also very grateful to my second supervisor Prof. Dr. Christina Kauschke for her support, and it is an honor to receive guidance from her.

Next, I wish to thank all my AG members: Steffi, Vera, Francie, Jonas, and Michelle. Their advice, coding support, and assistance in participant recruitment have been essential to the completion of this research. I really appreciated the lovely time that we've had together in every group meeting, and it will definitely be an unforgettable memory in my life. I would also like to extend my thanks to all the staff in my department who assisted with my experiments, as I can't finish my research smoothly without them.

Additionally, I would also like to thank my friends who encourage me over these years and were willing to take part in my experiments or helped me find suitable participants, especially during the challenging pandemic period.

Finally, and most importantly, I am very grateful to my parents and my husband for their unwavering love and support, because of them I can do what I want to do. Their encouragement has been invaluable during my Ph.D. program.

Content

List of Figures

List of Tables

Abstract in English

Numerous studies have already shown that not only does phonology have an influence on visual word recognition, but conversely, orthographic influences on spoken word recognition are also clearly evident in different languages. However, its impact on L2 auditory word processing, particularly for learners with a non-alphabetic L1 background like Chinese, has not yet been investigated. Unlike German, which features a shallow alphabetic orthography with consistent grapheme-phoneme correspondences, Chinese is a logographic language with a deep orthography and arbitrary mappings between written characters and phonological form.

The main objective of my study is to investigate which role orthographic and phonological representations play during word recognition in both L1 and L2, and to examine how language proficiency in L2 affects the orthographic effect in auditory processing of words in Chinese learners of German. To explore this issue, two groups of Chinese participants with intermediate and high language proficiency in German and a German control group were recruited to perform semantic judgment tasks on implicit processing of orthographic and phonological information in spoken word retrieval. The results showed that orthographic information doesn’t impact auditory word processing in German native speakers, whereas such influence is found in Chinese participants with high proficiency. But the L2 subgroup with intermediate proficiency did not show such a significant orthographic effect. This finding suggested that whether the orthographic form influences L2 spoken word recognition is proficiency-dependent but not associated with the target language.

In addition, previous research has demonstrated that long stays abroad in an L2-dominant environment can influence language knowledge of L1. This implies that long-term residence in an L2 language environment can also have an impact on orthography processing in L1, especially when orthography plays a crucial role in L1, as is the case with Chinese. Hence, this study applied the idea of the residence binding factor to the Chinese auditory task to explore how orthographic similarities affect spoken word

recognition when participants live in the L1- or L2-spoken environment respectively. A comparison of the orthographic effect in Chinese spoken word recognition was conducted between Chinese-German bilinguals living in Germany and a group of Chinese monolinguals living in China. The results revealed an obvious influence of orthographic similarity only in Chinese speakers who have been living in China, but not in Germany. This finding suggested that the associations between exposure to printed words and the effect of orthography on spoken word recognition are tightly linked, which means that the acquisition of an alphabetic L2 and long-term immersion in an alphabetic L2-spoken environment may reduce the orthographic effect during L1 spoken word recognition for speakers with a logographic L1 background.

It is generally emphasized in my studies that the unique characteristics of each language should be taken into account when investigating orthographic effects during auditory processing. The role of orthographic information in L1 spoken word recognition is language-dependent and in L2 spoken word recognition is proficiency-dependent.

Abstract in German

In zahlreichen Studien wurde bereits gezeigt, dass nicht nur die Phonologie einen nachweislichen Einfluss auf die visuelle Worterkennung hat, sondern umgekehrt in verschiedenen Sprachen auch orthografische Einflüsse auf die gesprochene Worterkennung deutlich erkennbar sind. Aber ob die Orthographie eine Rolle in der auditiven Wortverarbeitung bei L2-Sprechern spielt, deren L1 (Chinesisch) ein logographisches Schriftsystem aufweist und die eine Alphabetschrift wie Deutsch lernen, ist bislang noch nicht untersucht worden. Im Gegensatz zum Deutschen, das eine transparente alphabetische Orthographie mit überwiegend konsistenten Graphem-Phonem-Korrespondenzen aufweist, ist Chinesisch eine logographische Sprache mit einer tiefen Orthographie und arbiträrer Korrespondenz zwischen Schrift und phonologischer Form.

In meiner Dissertation geht es darum, wie sich orthographische und phonologische Repräsentationen auf die auditive Worterkennung sowohl in L1 als auch in L2 auswirken, und ob sich dabei Unterschiede zwischen L2-Lernern mit unterschiedlichem Kompetenzniveau zeigen. Um diese Frage zu klären, wurden zwei Gruppen chinesischer Probanden mit mittlerem und hohem Sprachniveau im Deutschen sowie eine deutsche Kontrollgruppe rekrutiert, um Studien mit orthographischem und phonologischem Priming durchzuführen. Die Ergebnisse zeigten, dass orthographische Informationen die auditive Wortverarbeitung bei deutschsprachigen Muttersprachlern nicht beeinflussen, während ein solcher Einfluss bei chinesischen Probanden mit hohem Sprachniveau festgestellt wurde. Jedoch waren keine offensichtlichen orthographischen Effekte in der L2-Subgruppe mit mittlerem Sprachniveau zu erkennen. Aus den Ergebnissen geht hervor, dass der Einfluss der Orthographie bei der Erkennung gesprochener Wörter in L2 sprachkompetenzabhängig ist, aber nicht mit der Zielsprache zusammenhängt.

Außerdem haben frühere Forschungen nachgewiesen, dass lange Auslandsaufenthalte in einer L2-Umgebung zu einer Verringerung der L1 Kenntnisse führen können. Dies

könnte bedeuten, dass der langfristige Aufenthalt in einer L2-Sprachumgebung einen Einfluss auf die Verarbeitung der Orthographie in L1 haben kann, insbesondere wenn die Orthographie in der L1 eine kritische Rolle spielt, wie es bei der chinesischen Sprache der Fall ist. Daher wurde der Faktor der Umgebung in einer weiteren Studie bei chinesischen Teilnehmenden in Deutschland und China untersucht, um herauszufinden, wie orthographische Ähnlichkeiten die Erkennung gesprochener Wörter beeinflussen, wenn die Probanden in der L1- oder L2-Sprachumgebung leben. Ich habe den orthographischen Effekt bei der Erkennung gesprochener chinesischer Wörter zwischen in Deutschland lebenden chinesisch-deutschen Bilingualen und einer in China lebenden chinesische monolinguale Gruppe verglichen. Die Ergebnisse zeigen einen deutlichen Einfluss der orthografischen Ähnlichkeit auf auditive Wortverarbeitung nur bei chinesischen Sprechern, die in China, aber nicht in Deutschland gelebt haben. Diese Ergebnisse deuten darauf hin, dass die Beziehung zwischen dem Output zu geschriebenen Wortformen und dem Effekt der Orthographie auf die gesprochene Worterkennung eng miteinander verknüpft ist, was bedeuten kann, dass der L2-Erwerb einer Alphabetschrift und der langfristige Aufenthalt in einer L2-Sprachumgebung mit einer Alphabetsprache den orthographischen Effekt in einer logographischen L1 verringern kann.

Es wird in meinen Studien generell betont, dass bei der Untersuchung orthografischer Effekte während der auditiven Verarbeitung die linguistischen Merkmale jeder Sprache berücksichtigt werden sollten. Die Rolle der Orthographie bei der Verarbeitung von gesprochenen Wörtern in L1 ist sprachabhängig und in L2 hängt von Kompetenz in den jeweiligen Schriftsystemen ab.

Chapter 1. Introduction

1.1 Motivation and research questions

The written forms and sounds of language play crucial roles in both first language (L1) and second language (L2) learning. However, in earlier studies, more emphasis was placed on the processing of written language in reading, while the sounds of language received relatively less attention. In linguistic research, we name the spelling of written language “Orthography” and the sounds of spoken language “Phonology”. Over the last three decades, there is growing research on the interaction between phonology and orthography during lexical processing in the psycholinguistics field. Not only phonology has been shown to impact visual word recognition, but orthographic influences on spoken word recognition have been previously demonstrated in different languages among native speakers (W. F. Chen et al., 2016; Pattamadilok et al., 2009; Perre, Midgley, et al., 2009; Qu & Damian, 2017; Ventura et al., 2007; Ziegler et al., 2003; Ziegler & Ferrand, 1998; Zou et al., 2012). There is evidence suggesting that auditory speech can be effectively processed without being affected by orthographic information, while others argued that spoken language is not a system of pure sound-meaning connections, orthography can probably change the way the spoken language is processed (Ziegler, Ferrand & Montant, 2004).

At present, there is numerous evidence from research conducted on alphabetic languages, highlighting the important role of orthography in processing spoken words. However, the majority of these studies primarily concentrate on alphabetic languages, whose orthography and phonology cannot be separated completely, and these studies primarily investigate the impact of orthography on spoken word processing within the context of L1. Even in the domain of L2 studies, experiments have predominantly focused on alphabetic languages as either the first or second language. So far, limited research has been undertaken to investigate the orthographic effect on spoken word recognition when the listeners' L1 is a nonalphabetic language while their L2 is an alphabetic language. To address this gap, we selected Chinese learners of German, along with a control group of German native speakers, as our participants. Unlike German, which has a shallow alphabetic orthography with highly regular grapheme-phoneme correspondences, Chinese as a logographic writing system has a deep orthography with an arbitrary mapping between written characters and phonological form. Hence, the Chinese language presents numerous linguistic distinctions from German, such as visually similar characters can possess entirely different pronunciations and meanings, and similar pronunciations can correspond to distinct characters and meanings. But it is important to note that in German, visually similar written forms are typically associated with similar sounds. By taking these linguistic features into account, investigating the impact of orthography on L2 spoken word recognition in Chinese learners of German might offer a unique insight into this research field. Also, we expect interesting differences in orthographic roles played in auditory word processing by comparing the native group in an alphabetic language like German with that in a non-alphabetic language like Chinese.

In addition, based on my personal experiences and observations, I have noticed that long-term immersion in an L2-spoken environment and reduced use of L1 can lead to changes in the native language itself. For instance, some of my Chinese friends living abroad have encountered hesitations while attempting to use certain Chinese words which rarely arise in their daily lives or even faced challenges in recalling the correct written form of some characters. Prior research has also suggested a reduction in native language skills due to long periods of living abroad in an L2 environment (Cook, 2002; Grosjean, 1989). Therefore, we are interested in investigating whether Chinese native speakers, who have been living abroad for a long time with limited opportunities to use the written form of their native language, may exhibit distinct patterns in processing orthographic information during spoken word recognition, as compared to those without knowledge of German and any experience of living abroad, who solely use

Chinese in their daily lives.

To summarize, this thesis attempts to contribute further evidence regarding the role of orthography in both native and non-native spoken word recognition. The study addresses several key questions:

a) How does orthography impact spoken word recognition for German native speakers processing German words?

b) What is the influence of language proficiency in L2 on the processing of auditory words for Chinese learners of German participating in the same task conducted in German?

c) How does the prolonged residence of Chinese native speakers in Germany, where they use German in their daily life, affect the involvement of orthographic and phonological information in their native word recognition, in comparison to Chinese native speakers residing in China?

All of these questions will be thoroughly addressed in the subsequent chapters with empirical investigation.

1.2 Thesis structure

This thesis is composed of five chapters. In this introductory chapter, I have presented the motivation behind the study and outlined the research questions that will be explored.

In Chapter 2, the theoretical background of word recognition will be reviewed. It will begin by defining the process of word recognition and exploring its various aspects, containing both L1 and L2 processing. After that, the relevant models about how the visual and auditory words are processed in both L1 and L2 contexts will be introduced. Furthermore, this chapter will offer an overview of orthographic activation during spoken word processing in native and non-native languages, considering various writing systems. I will present important findings that have investigated the impact of orthography on word recognition in both L1 and L2. In addition, the writing system of the German language and the Chinese language will be briefly introduced to enhance the understanding of how orthography plays a role in different linguistic systems.

In the third chapter of this thesis, an overview of the methodological design for the experimental study will be provided first. After that, a series of four experiments will be described in this chapter. Two of these experiments will focus on comparing L1 and L2 participants, as well as two subgroups of L2 participants with distinct language proficiency. Additionally, the other two experiments will concentrate on performance differences between two Chinese native groups with different residences. Experiment 1 will investigate the orthographic role during German spoken word recognition in the German native group. Experiment 2 will follow a similar experimental design but will be conducted with Chinese participants possessing two different levels of language proficiency in German. Experiment 3 will explore the orthographic role during Chinese spoken word recognition in Chinese native speakers residing in Germany. To contrast this, Experiment 4 will involve Chinese participants who reside in China and have never had any experience of living abroad, providing valuable insights into the influence of orthography on Chinese auditory word processing. For each experiment, I will provide detailed information about the participants, the materials used, and the data collection procedure. Also, the data analysis and a thorough discussion of the results for each experiment will be presented.

Chapter 4 will present a comprehensive conclusion of all the observed findings, discussing explanations for the findings and their relationship to the models and literature introduced in Chapter 2. Furthermore, a new model in spoken word recognition for Chinese learners of German will be proposed, taking into consideration the unique aspects of their language background and proficiency. Additionally, the limitations of the conducted experiments will be described, and potential areas for improvement in future research in the psycholinguistic field will be suggested.

Chapter 2. Theoretical background and previous research

This chapter aims to provide a detailed description of how words are recognized when we hear or read them, especially offering an overview of word recognition in L2. The relevant models, the theoretical grounding of the empirical part, as well as studies about orthographic effects on spoken word recognition, will be described in the following sections.

2.1 Word recognition

Generally speaking, word recognition involves the comprehension of both spoken and written words. The process of word recognition contains the identification of a specific word from other candidates stored in the mental lexicon, which operates like a dictionary storing all the words we have acquired. Within the mental lexicon, various types of information are encoded, such as pronunciation, spelling, and meaning. When we encounter a word through reading or hearing, the visual or auditive input will search for the best-matched word in the "dictionary", thereby enabling us to comprehend the meaning of this word.

Also, the recognition of spoken words differs from that of printed words. Because we hear spoken words usually only once and the word lasts for a brief duration, whereas with printed words, we have the option to review what we just read as many times as necessary; second, we can tell the approximate word length when we read the printed words at the first sight, but for spoken language, the word length is not apparent when we hear only the initial part of it; moreover, word boundaries are not necessary for the spoken word, and coarticulation causes sounds to blend into one another in a continuous speech while visually presented words can be identified much more easily with word boundaries and single letters (Dornbusch, 2012).

In this thesis, I will focus on spoken word recognition. The term *word recognition* has been used to refer to the end-point of the selection phase when a listener successfully identifies the specific lexical entry that was actually heard, and the word recognition point denotes the exact moment at which this identification takes place (Frauenfelder & Tyler, 1987). There is already some evidence to indicate that words are usually recognized by listeners before they are fully heard, whether in isolation or context (Grosjean, 1980; Marslen-wilson, 1984; W. Marslen-Wilson & Tyler, 1980).

Spoken word recognition is not just a matter of understanding what the word means, the overall process of spoken word recognition could be divided into three fundamental functions, namely access, selection, and integration (Marslen-Wilson, 1987): the access function is responsible for establishing the relationship between the recognition process and the sensory input, so the speech signal could be mapped onto the representations of written forms stored in the mental lexicon; the selection function operates as a mediator between the access and integration functions, discriminating the best-match from the available input; the integration function addresses how the recognition process relates to the higher-level representation of the utterance, ensuring that syntactic and semantic information associated with the word is integrated to complete the recognition process.

Various factors influence the speed and accuracy of word recognition. These factors include the clarity of articulation, the frequency and degree of familiarity associated with the given word, the presence and frequency of competing neighbors during the recognition process, and point at which the word could be distinguished from other candidates, as well as the influence of top-down information (syntactic, semantic, and pragmatic sources); also extralinguistic factors such as the context, the person we are speaking to, and the speaker's knowledge of the world play a role in word recognition (Grosjean & Byers-Heinlein, 2018). All these elements interact in the complex process of recognizing spoken words. Knowing that so many factors are influential, we need to explore how and to what extent these factors affect the process of word recognition. To do this, we must first know the concrete process of word recognition, which is presented as models that have been developed over the past several decades. Later in the thesis, some prominent models of visual and auditory word recognition will be introduced. In order to look deeper into the processing of words within the mental lexicon, this study will proceed by introducing the foundational findings subsequent to the presentation of these theoretical models. Since it is widely accepted that both visual and auditive word processing differs between L1 and L2 groups, the models of word recognition in L1 and L2 will be described separately in the following sections.

2.1.2 Models of visual word recognition in L1

Although this thesis focuses on spoken word recognition, there are some models of visual word recognition that are relevant to the theoretical background of the empirical investigation of auditory word recognition.

The IA model

The earliest and one of the most influential models of visual word recognition is the interactive activation model (IA model, McClelland & Rumelhart, 1981) (see Figure 1), which is also very important to develop models of auditory word recognition. The lines in Figure 1 ended with arrows representing excitatory effects and those ended with dots mean inhibitory interactions. According to the interactive activation model, words are represented as nodes in a network that are connected by inhibitory links, and the orthographic and phonological codes are with automatic links between them (McClelland & Rumelhart, 1981). The proposed model comprises three distinct levels: the feature level, the letter level, and the word level, each comprised of a set of units or nodes (see Figure 1). When a visual stimulus containing a string of letters is presented, the feature level nodes are immediately activated, thereby triggering the activation of all letter nodes that correspond to these features while inhibiting other irrelevant nodes at the same time; subsequently, the best-matched letter nodes receive the greatest excitation and proceed to activate the corresponding word nodes; in turn, these word nodes engage in competitive processes with all other word nodes, and they also send feedback activation to the letter nodes that are consistent with them (McClelland & Rumelhart, 1981).

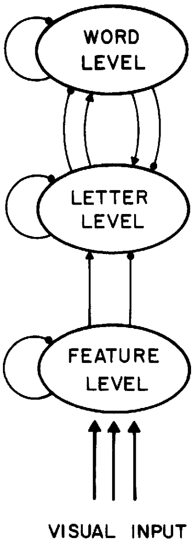


Figure 1. The various levels of processing considered in the interactive activation model and their interconnections. (McClelland & Rumelhart, 1981)

The interactive activation model also provided a basic interaction account for auditory word recognition.

DRC model

The dual route cascaded (DRC) model from Coltheart et al. (Coltheart et al., 1993, 2001; Coltheart & Rastle, 1994) is regarded as one of the most influential models of visual word recognition. It represents an extension of the interactive activation and competition model originally proposed by McClelland and Rumelhart (1981) and by Rumelhart and McClelland (1982). Figure 2 illustrated the overall architecture of the DRC model, in which arrows represent excitatory links between units, and circles indicated inhibitory links between units.

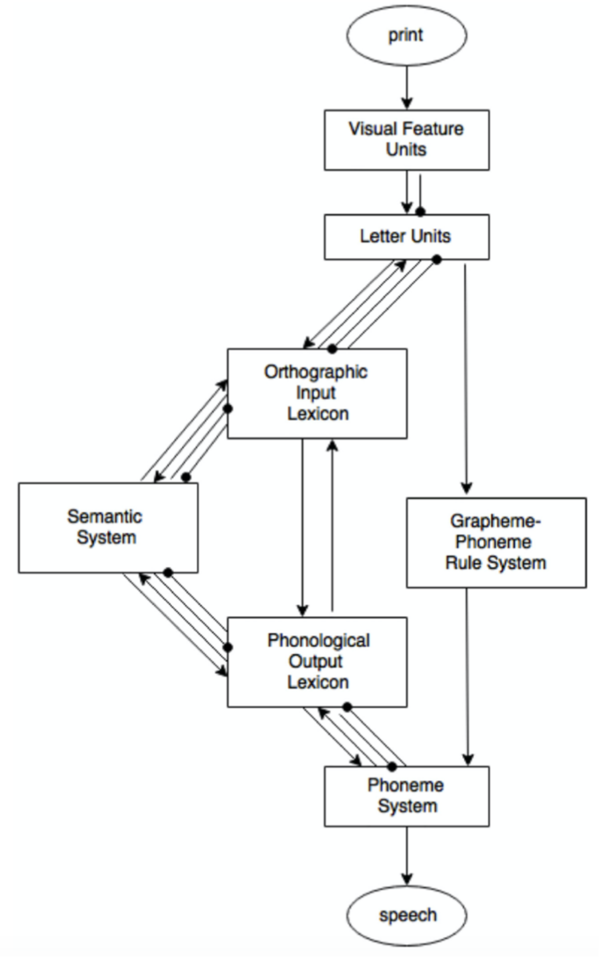


Figure 2. The dual-route cascaded model of visual word recognition and reading aloud. (Coltheart et al. 2001)

According to Coltheart et al. (2001), the model comprises three distinct routes: the lexical semantic route, the lexical nonsemantic route, and the GPC (grapheme–phoneme conversion) route; and within each route, multiple interacting layers are present, with each layer comprising sets of units that represent the smallest symbolic components of the model. For instance, these units may represent words within the orthographic input lexicon or individual letters at the letter unit level. The interaction between units from different layers takes place in two primary ways: inhibition and excitation (Coltheart et al. 2001): inhibition occurs when the activation of certain units hinders the rise of activation in other units, while excitation occurs when the activation of a unit facilitates the activation of other units. Moreover, inhibition operates within the same level, where units within a given level inhibit one another through inhibitory lateral connections. Furthermore, at the levels of orthographic and phonological lexicons, connections between units are exclusively excitatory, facilitating information flow between these levels.

The DRC model operates as follows (Coltheart et al., 2001): firstly, the visual feature units are interconnected with the features of the letter string, facilitating the transmission of activation from the feature level to the letter level; next, the orthographic lexicon activated and is also fed back to the letter level, leading to cascaded processing, this activation process results in a build-up of activation in the phonemic layer, with feedback from the phoneme layer to the letter layer; meanwhile, the GPC system contributes activation to the phoneme layer.

The BIAM

The BIAM (Bimodal Interactive Activation Model) is a bimodal adaptation of McClelland and Rumelhart's (1981) interactive activation model, and it posits bidirectional activation between phonological and orthographic units at both the sub-lexical and lexical levels (Grainger & Ferrand, 1994, 1996; Grainger et al., 2003; Ziegler et al., 2003).

It is worth noting that the BIAM predicts fast phonological priming, which is not present in the DRC model proposed by Coltheart and colleagues (2001), so phoneme representations from the input can transmit activation to lexical phonological representations through pathways that also facilitate auditory word recognition in the model (Diependaele et al., 2010). The sublexical phonological effects on visual word recognition are intermediated via phonological input, with the phonological representations being rapidly activated upon the presentation of a printed word (Grainger & Holcomb, 2009).

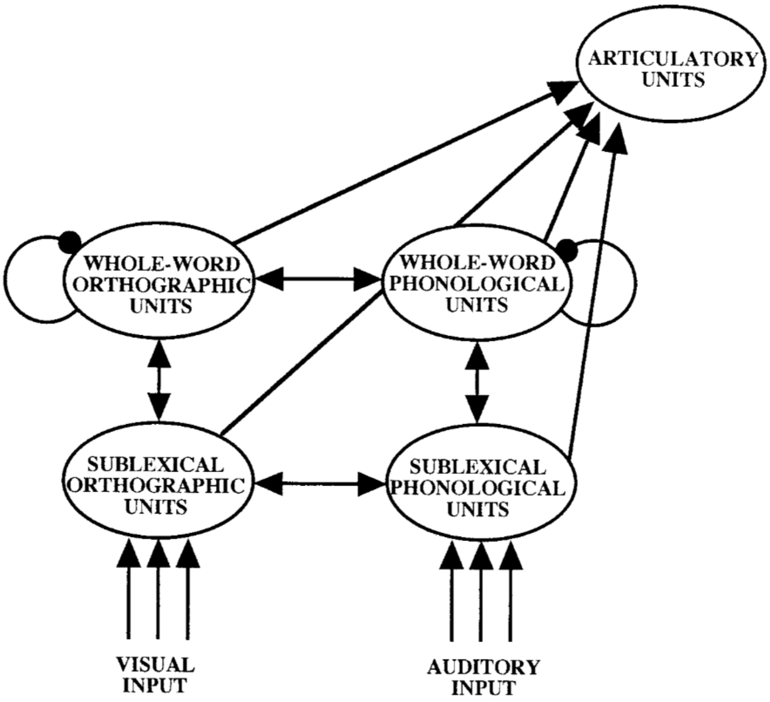


Figure 3. Grainger and Ferrand's (1996) bimodal interactive activation model.

We can see from Figure 3 that the inter-level connections at both sublexical and lexical processing levels operate in a facilitatory manner, whereby units at sublexical phonology level activate correspondent units at sublexical orthography level (Grainger & Ferrand, 1996). This operation enables the model to account for the impact of orthographic information on phonological processes. Additionally, within each processing level, similar units demonstrate inhibition through lateral inhibition mechanisms. As a result, the BIAM can simulate inhibitory neighborhood density effects, which were commonly observed in the studies of auditory word recognition (Ziegler et al., 2003).

The BIAM is not limited to visual word recognition but also encompasses the architecture necessary for bimodal processing of spoken words. We can see from Figure 3 that the articulatory features in the BIAM also activate sublexical phonological units, meanwhile, these sublexical units subsequently activate the whole-word orthographic and phonological representations, facilitating an integrated word recognition process.

2.1.3 Models of spoken word recognition in L1

Over the past several decades, there has been a growing emphasis on developing models for word recognition. While current models of spoken word recognition may exhibit variations in their implementations, there are at least three aspects in which they generally concur: 1) as a word is being heard, numerous word candidates are activated simultaneously; 2) The extent of correspondence between the incoming speech signal and the stored lexical representations exerts an influence on the activation levels of various word candidates; 3) the activated candidates compete with each other during the process of recognition (Weber & Scharenborg, 2012).

The current debate in the study of word recognition focuses on whether distinguishable levels of processing may or may not interact (Cutler et al., 1987). Some models of word recognition propose information flow between processing levels, and they are considered interactive accounts. On the other hand, some models don't allow interaction and adopt autonomous processing at each level, relying solely on serial and bottom-up information flow. Models belonging to the interactive models class include the cohort model of auditory word recognition (W. Marslen-Wilson & Tyler, 1980) and the TRACE model (Elman & McClelland, 1984; McClelland & Elman, 1986). Autonomous models differ from interactive models, and one example of an autonomous model is the Shortlist model of word recognition (Norris, 1994). In this section, I will introduce these three important models, which were basically from the perspective of alphabetical languages.

The Cohort Model

The Cohort model was regarded as the first psycholinguistic model of word recognition developed particularly for spoken language (Perre, Pattamadilok, et al., 2009). According to the Cohort Model, phonology is perceived sequentially and the word-initial sound activates a cohort of lexical candidates (W. Marslen-Wilson & Tyler, 1980; Tyler, 1984). Word recognition begins as soon as the first phoneme of a word is heard, rather than waiting for the whole word to finish. During this process, the acoustic input is mapped onto a word in the listener's lexicon, and competing candidates are inhibited, gradually narrowing down the cohort of potential words until only one candidate remains, leading to the recognition of the word (Marslen-Wilson & Tyler, 1980). After the recognition, the word is selected and integrated into the context.

Taking the spoken word "cat" as an example, "cat" shares the same initial phoneme /k/ with "cup", "cash" and "car". And the phoneme string /kæ/ could also be the beginning of "car" and "cash", but when /t/ comes out, only one candidate "cat" remains. Because once /kæt/ is perceived and no other English words are spelled in that way, the word "cat" can be recognized. We can also say the /t/ is the "recognition point" of /kæt/.

There are some observations from experiments that supported the Cohort Model. One of the main results is obtained in an auditory lexical decision task (see Marslen-wilson, 1984). The participants were presented with a set of stimuli containing both real words and nonwords, and they were instructed to indicate whether each item was identified as a nonword or not by pressing specific buttons. The nonwords were created by manipulating real words at different positions within the phoneme string, such as by altering the initial phoneme of a real word (e.g., changing "zawritude" to a nonword after /z/), or were formed by modifying the middle position of the phoneme string in a real word (e.g., transforming "trenker" into a nonword after /k/). The results showed that decision time from critical phoneme offset remained constant, regardless of the presence point of critical phoneme in the sequence and how the sequence is. Following Marslen-Wilson's (1984) observations, it was proposed that the specific position of the deviation point, also known as the uniqueness point, did not significantly influence response times, and the critical factor for word recognition should be the point at which only one candidate word remained within the cohort of potential matches.

But the finding from Taft and Hambly (1986) challenged the Cohort model. It demonstrated that higher-frequency words are recognized more effortlessly by listeners than lower-frequency words, even when both types of words reach the same recognition point in the model. Consequently, the Cohort model's ability to explain the impact of word frequency on on-time recognition remains limited. After that, Marslen-Wilson (1987) put forward a new Cohort model which still uses the activation concept, but with an account for frequency effects. The results suggest a significant advantage in recognition time for high-frequency words. Additionally, the new Cohort model operates as a bottom-up system, with no influence from top-down contextual information on the actual lexical recognition units (Marslen-Wilson, 1987).

However, the Cohort model still faces a question about the efficient processing strategy in auditory word recognition assumed from Marslen-Wilson (1984), that is, the role it plays during the real-time processing of continuous speech (Luce, 1986). Luce (1986) suggested that an optimally efficient recognition strategy might be limited to longer, low-frequency words. He also raised the possibility that some words could be challenging to recognize in isolation. Thus, a more efficient model for shorter and high-frequency words in continuous speech is necessary.

The TRACE Model

McClelland and Elman (1986) introduced a model called the TRACE model, which is based on the principles of interactive activation and is regarded as the first computationally implemented model of spoken word recognition (Weber & Scharenborg, 2012). According to this model, although bottom-up input is crucial, other factors such as top-down context also affect speech recognition (McClelland & Elman, 1986). For instance, an experiment conducted by Grosjean (1980) revealed that participants took significantly longer time to correctly identify a word in isolation than when the same word was placed in a sentential context.

How bottom-up and top-down processes interact, is the focus of the TRACE model. In other words, word activation is initiated by phonemes (bottom-up), and subsequently, the activated words provide feedback to activate their corresponding phonemes (top-down). This bidirectional flow of activation forms a key aspect of the TRACE model. For example, when word items share some phonological similarities, like “cat” and “hat” which share rhymes, the similarity competes for recognition as well.

The TRACE model is characterized as a dynamic processing structure comprising a large number of units organized into three distinct levels: the feature level, the phoneme level, and the word level. We can see from Figure 4 that the words in TRACE model are represented as phonemic strings, and these strings are converted to multi-dimensional features that present acoustic-phonemic patterns; on the phoneme-level, the individual units which received the bottom-up information inhibit each other; then the activated phonemes encode into candidate words on word-level, and finally, the best-matching word will be recognized (e.g., the word *bat* in Figure 4) (McClelland & Elman, 1986). Not like the Cohort Model which showed the sequential direction of processing, according to the TRACE model, auditory input can go through each level, from features to phonemes and then to words, or in opposite direction, or get off at each level and explore within the level. But there is no inhibition between levels in TRACE, feedback connections from the word-level to the phoneme-level make TRACE an interactive model and therefore affect perception (Weber & Scharenborg, 2012).

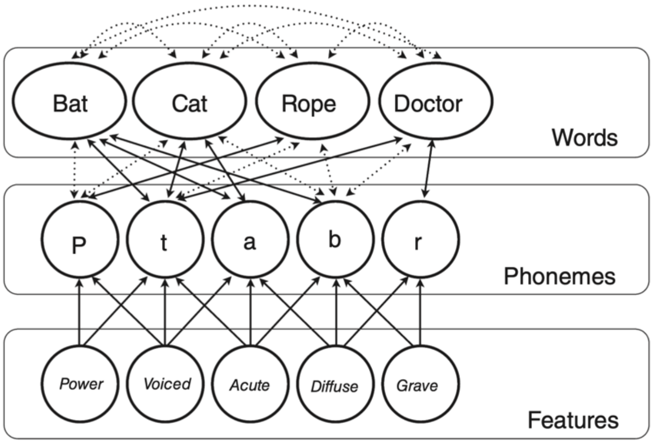


Figure 4. The TRACE model of auditory word recognition. (Joanisse & McClelland, 2015)

Despite both the TRACE and Cohort Models allowing for top-down influences on spoken word recognition, they exhibit some notable distinctions. The Cohort Model primarily emphasizes word-level processing, whereas the TRACE model places greater emphasis on the identification of features and phonemes. Furthermore, the Cohort Model heavily relies on clear initial phonological input to activate cohorts, while the TRACE model permits the activation of shared features, such as rhymes.

With TRACE model we can explain some facts derived from the experiments. For example, Cutler et al. (1987) observed that a consonant was detected faster in a word than in a non-word in a phoneme monitoring task. According to the "word superiority effect", lexical activation influences phoneme recognition even when the auditory signal is clear, which is consistent with TRACE model that the word-level feeds back and activates phonemes. But when phonemes form a non-word, such facilitation from the word-level exists not anymore. Besides, Allopenna and colleagues (1998) found that participants looked more at the phonologically related picture (words that share the same initial phonemes or share the same rhymes) than unrelated ones in an eye-tracking experiment. These results strongly support the validity of the TRACE model in explaining the overall pattern of spoken word recognition, as it suggests that the process is unlikely to be strictly sequential.

But there are some restrictions of the TRACE model, due to the vocabulary being limited to one-syllable words, and word frequency effects were not considered in the TRACE model.

Shortlist

In contrast to TRACE, which supports a highly interactive view of auditory word processing, autonomous models (Forster, 1976, 1979; Seidenberg, 1985; Tanenhaus et al., 1985) posit that lexical access and selection are modular processes. According to these models, the processing from signal analysis to word selection is autonomous, and each module operates independently, motivated primarily by bottom-up sensory information.

Shortlist is one of the autonomous models which in the absence of top-down effects and is entirely bottom-up; meanwhile, it can easily perform simulations with vocabularies of tens of thousands of words (Norris, 1994). The Shortlist model encompasses two processing stages. In the first stage, a shortlist of word candidates is generated through an exhaustive lexical search. These word candidates can then be compared to the input but are limited to a set of words (maximum 30) at each segment. If there are an excessive number of candidates, those with the lowest bottom-up activation are eliminated from the shortlist, while the candidates with higher bottom-up activation continue to undergo processing. In the second stage, the best-matching lexical candidates are connected into a small interactive activation network which is equivalent to the word level of TRACE, so that the candidates which shared some features inhibit each other in proportion to how many phonemes by which they overlap. As Figure 5 presented, an initial candidate set is generated on the basis of bottom-up input alone, which consists of acoustic-phonetic features; and some of the multiple lexical candidates will be selected through competition processing within the generated candidate set, without influencing the generation stage. Besides, the recognition of phonemes is not influenced by lexical processing and is also not required for lexical processing (Houkema, 2001).

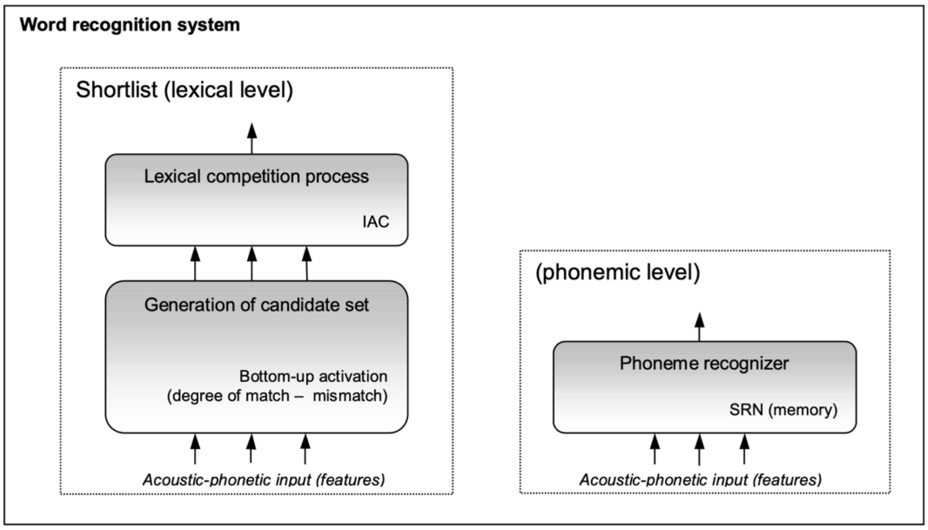


Figure 5. Schematic representation of the autonomous Shortlist model of spoken word recognition (Norris, 1994b, adapted from Houkema, 2001)

The Shortlist model possesses two distinctive features: first, it incorporates the influence of lexical stress in constraining word activation; second, the model implements the possible-word constraint, which reduces the activation of candidate words when they are surrounded by the input that cannot form a viable word (for instance, in English, the activation of “apple” in the string “fapple” is decreased, as a single consonant cannot constitute a word (Weber & Scharenborg, 2012).

Among the various models discussed so far, a central theme in the research of spoken word recognition is whether the processing is predominantly bottom-up or influenced by top-down information. When interactive models permit bidirectional information flow from lower to higher levels and vice versa, autonomous models posit that information flow occurs primarily from the bottom up. The list of models is not complete enough, as it does not encompass L2 lexical processing. The following section will provide a brief overview of some spoken word recognition models, with a specific focus on L2.

2.1.4 Models of visual and spoken word recognition in L2

As mentioned earlier, word recognition in L1 and L2 differs significantly due to the distinct language experience between these two groups. L1 speakers are solely immersed in their native language, whereas L2 learners must deal with at least two languages when they read or hear words in a second language. As a result, the existing model of word recognition in L1 must be adapted to suit the L2 group.

The BIA model

The Bilingual Interactive Activation (BIA) model (Dijkstra & van Heuven, 1998) is an extended version of the Interactive Activation (IA) model of monolingual visual word recognition developed by McCelland and Rumelhart (1981), so it actually shares some basic architecture with the IA model, like visual features, letters, and the word-level (see Figure 6). But the BIA model has one more level of representation units called language-level. The level of language contains two nodes, one for Dutch, and one for English.

When a string of letters is presented to the BIA model, the feature nodes at each position are activated, subsequently activating the letter nodes that correspond to these features, and the inhibition occurs when letter nodes do not match. After that, the activated letters transmit their activation to word nodes in both languages, then the activated word nodes send their activation back to the corresponding language nodes and also provide inhibitory feedback to the letter level, effectively inhibiting all other words (Dijkstra & van Heuven, 1998). Step by step, the best matching word candidates become most active.

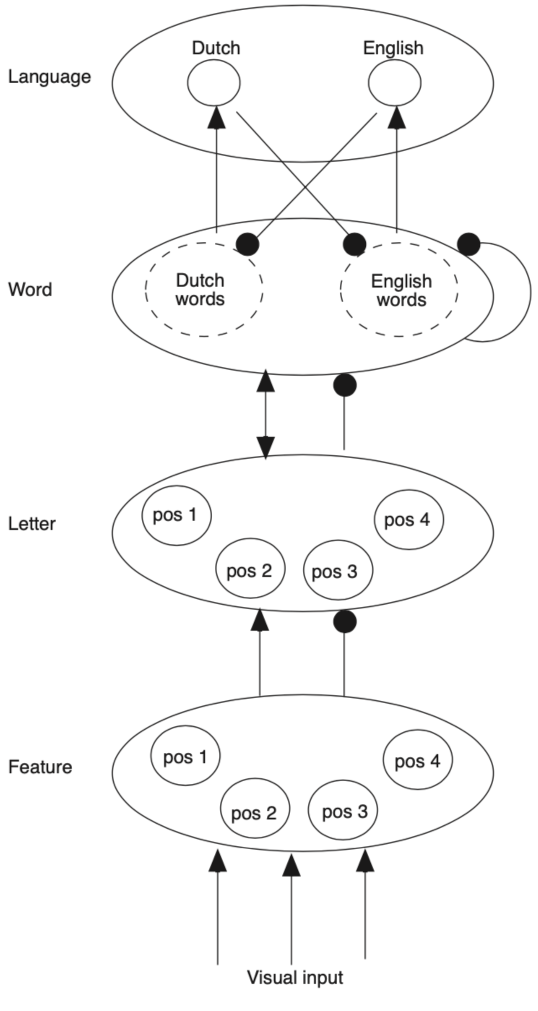


Figure 6. The Bilingual Interactive Activation (BIA) model for bilingual word recognition. Arrowheads indicate excitatory connections; ball-headed lines indicate inhibitory connections. (Dijkstra & van Heuven, 1998)

In the IA model, differences in word frequency influence the recognition process: higher-frequency words have a higher resting level activation when the recognition process begins, and therefore the recognition moment from them will be reached earlier than from other less frequent words (McClelland & Elman, 1986). Similarly, the BIA model also adopts this assumption, suggesting that distinction in first and second-language proficiency leads to differences in the frequency of word usage between the two languages, so words from the more frequently used language have a higher resting level of activation compared to the less practiced language (Dijkstra & van Heuven, 1998). This BIA model also provides a view that the top-down inhibition effects from two languages to word level are asymmetric, L1 words might send more inhibition to L2 words than in opposite direction.

The BIA+ model

The BIA model has its limitations. On the one hand, there are no semantic or phonological representations in the model, but word recognition is also affected by phonological and semantic information. On the other hand, the BIA model can only recognize words of the maximal length of four letters, which makes it hard to investigate some effects when words with more than four letters are used as stimuli in experiments. Due to these limitations, the BIA model was updated to the BIA+ model by Dijkstra and Van Heuven (2002). The BIA+ model consists of two systems: an identification system and a task/decision system (see Figure 7). The identification system assumed that the language lexicon is non-selective, so orthographic, phonological, and semantic representations are integrated stored. Consequently, when visual input matches the orthographic representations, it triggers activation of both orthographic and phonological lexical as well as sublexical representations; after that, semantic representations are activated, and word candidates from any language are selected in the end. Therefore, in cross-linguistic context, bilingual word recognition is influenced not only by orthographic overlap but also by phonological and semantic similarities (Dijkstra & van Heuven, 2002). Besides, a task/decision system receives continuous input from the identification system and is called task schema. Task schema determines which task-specific response procedures must be applied to the task at hand and reads out the activation in the identifying system continually; moreover, the decision often depends on lexical selection (see Figure 7). Hence, it implies that except for linguistic effects, we should also consider task-dependent and language-dependent impacts.

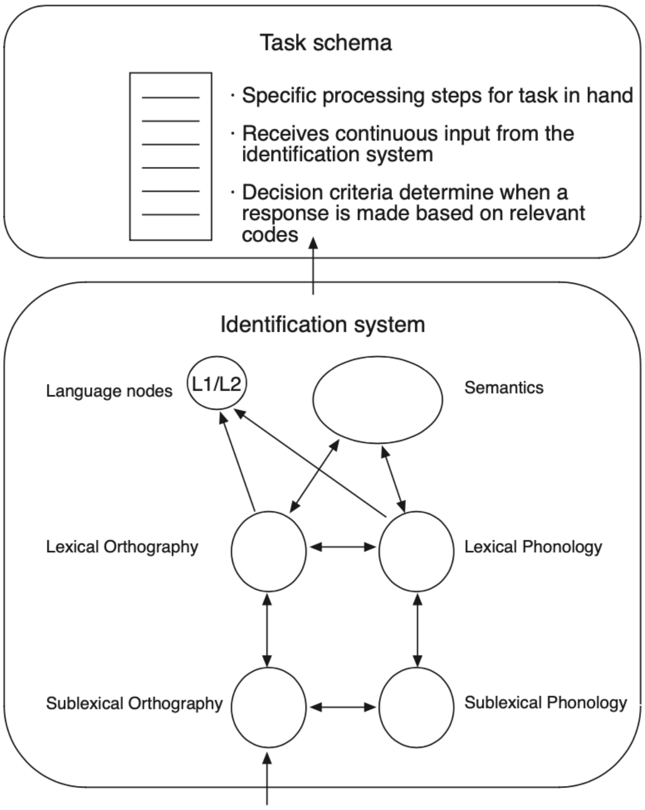


Figure 7. The BIA+ model. Activation flows between representational pools are shown by arrows. Within pools, inhibitory connections are omitted. Lemma representations between word form and meaning representations might be connected to language nodes. Only the task schema is affected by the non-linguistic environment (Dijkstra & van Heuven, 2002).

It should be noted that the initial phase of word recognition in BIA+ model works similarly to how they do in the BIA model and the internal representation will be more activated when there is a greater overlap between the input string and a representation in the mental lexicon (Dijkstra & van Heuven, 2002).

Additionally, the BIA+ model predicted lower resting level activations in L2 than in L1, which means that compared to L2 word recognition, L1 word recognition requires less extra activation to reach the recognition threshold. The results of studies about a larger frequency effect in L2 than in L1 supported this idea (Cop et al., 2015; Duyck et al., 2008). However, there is a precondition that both languages have similar orthographies, and the number of activated orthographic or phonological codes is determined by neighborhood density and word frequency. Hence, word candidates cannot be activated across alphabetic and logographic writing systems.

The BIMOLA model

Based on BIA model, Léwy and Grosjean developed a computational model of spoken word recognition (BIMOLA) (in Grosjean, 2008), which was also inspired by McClelland and Elman's TRACE model. BIA and BIMOLA models are both based on interactive activation models (see McClelland & Rumelhart, 1981), but BIA used stimuli from Dutch and English words whereas BIMOLA used French and English words. Figure 8 presents a simplified visual representation of BIMOLA, which contains three levels of nodes: features, phonemes, and words. The phoneme- and word-level nodes are organized independently, whereas the feature-level nodes are shared by the two languages, such as the allophonic variants in English and French. As a result, each language is represented by a small subset of units and a larger system containing these subsets. This representation is evident at the phoneme and word levels, where units can have both close and distant neighbors, as depicted in Figure 8, with the extent of darkness illustrating the closeness of these relationships (darker color represents the closer neighbors). At the word level, frequency is represented by the size of units. The bidirectional arrows between the phoneme level and word level illustrate the bidirectional activation connections, whereas the activation connections between the features level and the phonemes level are solely bottom-up (Grosjean, 2008). We can also see from Figure 8 that features activate phonemes first and then activate words. Furthermore, a top-down pre-activation of words is implemented, which relies on external information regarding the listener's language mode and higher linguistic expertise (Grosjean, 2008). Subset activation and lateral inhibition operate at both the phoneme and word levels but are limited to a single language. Phonotactic activation, on the other hand, is exclusively present at the phoneme level. Compared to the BIA model, there is no cross-language inhibition in the BIMOLA model, which means, units within a level inhibit one another but only within a language (Grosjean, 2008).

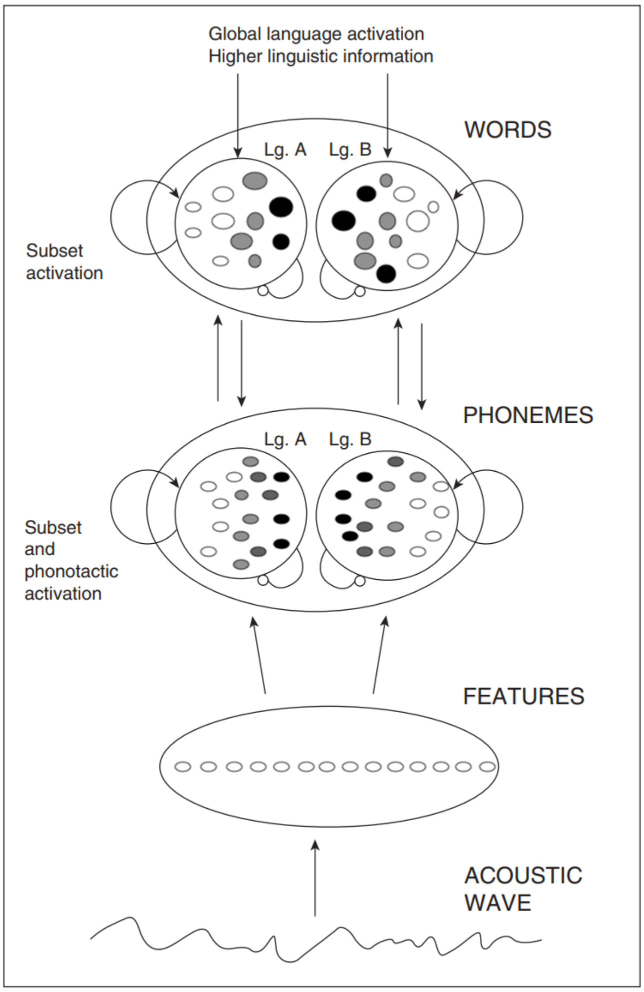


Figure 8. The Léwy and Grosjean bilingual model of lexical access (BIMOLA). Adapted from Fig. 11.1, p. 204, in *Studying Bilinguals* by François Grosjean (2008).

Grosjean's BIMOLA model presented distinct parallel phonological and lexical layers for two languages. Listeners use acoustic inputs to activate layers of both languages, ultimately selecting the most matching word based on the layer that produces the strongest activation between the two languages. However, we should note that the two languages described in the BIMOLA model are restricted to the case where L2 shares its alphabet with L1.

2.2 The difference between late L2 learners and L1 speakers

In this section, it is important to provide a clear definition of "L2 learners" and "L1 speakers" following the introduction of distinct models in L1 and L2 word recognition. According to Paradis and colleagues (2011), L2 learners denote children who have already made significant progress in acquiring L1 before they begin the acquisition of an L2. Whether the acquisition of a second language starts before or after the age of three differentiates between distinct conditions of second language learning: children who start to learn their second language later than by age three and use distinct languages at home, in the educational environment, and within the community are commonly referred to as sequential bilingual children or successive bilinguals, while simultaneous bilingual children typically acquire two languages at home before three years old (Paradis et al., 2011). Late L2 learners can be defined as individuals who are actively engaged in the process of learning an L2 to achieve functional and communicative goals, not solely for educational requirements satisfaction (Best & Tyler, 2007). Late L2 learners may not process their L2 in a native-like manner, while early learners have a higher likelihood of achieving native-like language processing abilities (Sabourin & Stowe, 2008).

Secondly, it is crucial to differentiate between SLA (Second Language Acquisition) learners and FLA (Foreign Language Acquisition) listeners in terms of their language-using environments. SLA involves the process of acquiring an L2 in natural communicative contexts, where the target language is commonly used and encountered in real-life interactions; and FLA takes place in more controlled and constrained settings, such as formal foreign language classrooms, where the target language is not as extensively used in daily life (Best & Tyler, 2007). As a result, FLA listeners have limited or no conversational experience with native speakers and primarily encounter the L2 through formal instruction in a restricted environment, and SLA listeners have more opportunities to interact with native speakers and experience a higher level of L2 exposure due to the dominant L2-speaking environment. Consequently, the performance of FLA listeners may differ from that of SLA listeners in various linguistic tasks, and we need to restrict participants to one of the two categories.

Furthermore, it is important to note that L1 speakers are typically child learners, whereas L2 learners are adult learners. This diversity in learning experiences leads to differences in language processing between L1 and L2 speakers (here specifically targeted "late L2 learners"). We can summarize the differences between them as follows: On the one hand, in L1 acquisition, native speakers typically learn the phonological system of their language before the orthographic system, and their exposure to the language is primarily through oral input. However, in the early stages of L2 language learning, learners often encounter both phonology and orthography simultaneously. Moreover, L2 learners often acquire the language predominantly in classroom settings, where they receive substantial exposure to orthographic information from the early stages of learning. Hence, extensive experience with written language becomes one of the most important features for L2 speakers, which might result in a different way of word processing when it is related to written forms.

On the other hand, L2 learners typically acquire the orthographic forms of their second language after becoming literate in their first language, so they may decode the orthographic forms of the L2 using the phoneme-grapheme correspondence established in their native language (Bassetti, 2017). Also, since L2 learners speak two languages and acquire the second language later in life, they generally have less language experience compared to native L1 speakers. This might lead to relatively weaker and less accurate mapping from phonological to orthographic representations in their memory, resulting in greater retrieval effort when attempting to comprehend spoken words in the L2.

In addition, in line with the arguments proposed by Cook (2002, 2003), learning a second language can influence the way our first language is represented in our minds, and the effects of L2 on L1 could be positive, negative, or neutral. For example, Laufer (2003) carried out studies with Russian-Hebrew bilinguals and found that the ability to recognize incorrect collocations in L1 was influenced by L2 word knowledge, also the diversity of vocabulary in L1 decreased with longer exposure to L2. However, some studies also have shown that exposure to an L2 can lead to acceleration in certain aspects of the first language. There was research that showed that English children who receive one hour of Italian language course per week demonstrate enhanced reading proficiency in English compared to their peers who do not have exposure to a second language (Yelland et al., 1993). One possible reason for this improvement could be the linguistic closeness between the two languages. It would be intriguing to investigate how prolonged immersion in an L2-speaking environment impacts the word processing of the first language, especially when the two languages differ significantly in their writing systems.

Not only the language experiences of L1 and L2 speakers are different, but previous studies also told us that some linguistic factors influence word recognition to a dissimilar extent among L1 speakers and L2 learners. In the L2 visual word recognition literature, there has been plenty of discussions about frequency effects. The study from Duyck and colleagues (2008) examined the size of the frequency effect in L1 and L2 during word recognition and found that Dutch–English bilinguals showed a significantly higher frequency effect in their L2 than in their L1, even though the frequency of stimuli was matched across languages. The same result is also found in a study by Cop, Keuleers, Drieghe, and Duyck (2015), which used a natural reading task with eye-tracking paradigm. Besides, they demonstrated that the frequency effect in both L1 and L2 reading became weaker when L1 proficiency increased, but it was not affected by L2 proficiency. As exposure appears to be the primary factor influencing lexical entrenchment within an integrated mental lexicon, it's important to note that not all groups of bilinguals necessarily exhibit lower L1 exposure than monolinguals, which suggests that a qualitative distinction in language processing between monolinguals, bilinguals L1, or bilingual L2 may not be necessary to explain reading performance (Cop et al., 2015).

Not only in visual word recognition but also in auditory word recognition there are findings that support that the impact of word frequency in L1 and L2 is different. For instance, Schmidtke (2014) conducted a study that involved the recording of pupil size while participants listened to English words and matched them to one of four pictures. The results revealed that bilingual speakers demonstrated an overall delayed pupil response compared to monolingual speakers. Furthermore, while the frequency effect remained consistent between early bilinguals and monolinguals, it was more pronounced in late bilinguals. Importantly, the authors did not merely attribute these results to the categorical difference between L1 and L2, but instead emphasized the early or late acquisition of L2 among bilinguals.

Within the L2 groups, there are also differences among individuals that we can discuss, such as language history, language use, and language proficiency. Language history and language use are difficult to separate because they are correlated with each other in most studies (Grosjean & Byers-Heinlein, 2018). These two factors normally refer to the age of L2 acquisition, the way in which the language was acquired (e.g., through classroom instruction or immersion with native speakers), length of residence in the target country, relative usage of L1 and L2, as well as the quantity and quality of input from native speakers. Shi and Morozova (2012) conducted a study exploring how the age of L2 acquisition (English), length of residence in the country, and daily exposure to L2 impact word recognition. The results demonstrated that the second-language learning history did indeed influence word recognition. These factors also play a role in the development of L2 language learning, ultimately determining language proficiency. Language proficiency, and vice versa, serves as a crucial factor during word recognition, as less fluent L2 groups may experience slower word recognition.

Apart from age of acquisition and language proficiency, other factors also influence word processing in L1 and L2 to varying degrees. Schröter and Schroeder (2018) reported greater effects of word length in L2 compared to L1 speakers, although the two groups they tested did not differ significantly in their overall performance on the lexical decision task. As previously mentioned, all these factors will be carefully considered when recruiting L2 participants for the studies in this dissertation.

Considering the presented findings, it is hypothesized that L1 speakers and L2 learners use distinct ways of processing lexical representations. Additionally, it is crucial to take into account the unique writing systems associated with L1 and L2. The aim is to investigate the processing of L2 learners whose L1 orthographic system does not overlap with the L2 orthography. Since this is central to my studies, an introduction to the language writing systems pertinent to this dissertation will be reviewed to enhance comprehension of the relevant empirical findings.

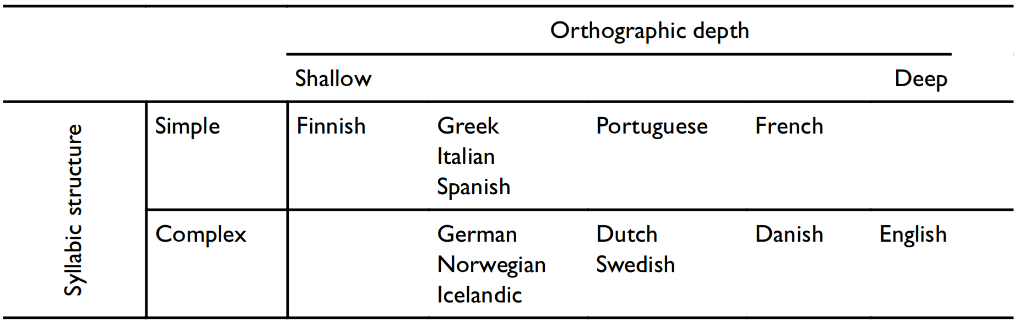
2.3 Orthographic systems

Different writing systems can refer to different characteristics of spoken languages and can be distinguished by the way in which phonological units are mapped onto orthographic representations: by phonemes, as in English; by syllables, as in Japanese Kana; or by morphosyllables, as in Chinese characters (Frost, 2008). Among these writing systems, it is important to distinguish their characteristics. In an alphabetic writing system, the elementary graphic units correspond to individual phonemes of the spoken language; a syllabic writing system employs elementary units that correspond to spoken syllables; in a logographic writing system, the elementary unit represents a spoken syllable, which coincides with a morpheme or a complete word (Perfetti & Liu, 2005). For example, in English, the letters in the writing system represent individual phonemes, allowing for different pronunciations of English words through various combinations of letters; in Japanese Kana, the graphemes represent syllables in the Japanese language, and its syllabic structure is based on vowel or consonant-vowel combinations, determining how the graphemes are pronounced, such as こ pronounced as /ko/ and ん pronounced as /n/, then こん can be pronounced as /kon/; in Chinese, the graphemic structure represents the meaning of morphemes and sometimes a spoken syllable can represent many different characters, so we can’t pronounce it barely through the grapheme, such as /qing1/ can be written as 青, 轻, 倾 and many other characters.

In my thesis, two orthographic systems will be focused on: alphabetic orthographies such as a focus on German, and logographic orthographies such as those used in Chinese. To understand the influence of orthographic information on spoken language recognition, the concrete characteristics of these two writing systems will be explained in the following sections.

2.3.1 The German writing system

Before I provide a general overview of the German writing system, two important characteristics that determine alphabetic orthographic systems will be first introduced. Seymour, Aro & Erskine (2003) classified the orthographies along the two dimensions of (1) syllabic complexity and (2) orthographic depth (see Figure 9).



		Orthographic depth				
		Shallow				Deep
Syllabic structure	Simple	Finnish	Greek Italian Spanish	Portuguese	French	
	Complex		German Norwegian Icelandic	Dutch Swedish	Danish	English

Figure 9. Hypothetical classification of participating languages relative to the dimensions of syllabic complexity (simple, complex) and orthographic depth (shallow to deep) (Seymour et al., 2003).

Syllabic complexity

Syllabic complexity denotes the distinction between Romance languages, such as Spanish, characterized predominantly by open CV syllables with minimal initial or final consonant clusters, and Germanic languages like German and English, which exhibit a higher frequency of closed CVC syllables and more complex consonant clusters in onset and coda positions (Seymour et al., 2003). Owing to the common roots of German and English, they share similarities in terms of phonology, and both utilize the same 26 letters of the Latin alphabet. However, German possesses a few additional characters, such as umlauted letters (ä, ö, and ü) and Eszett (ß).

Orthographic depth

Orthographic depth can be defined by the consistency of letter–phoneme correspondences in a language (Liberman, Liberman, Mattingly, & Shankweiler, 1980). In accordance with Liberman et al. (1980), orthographic depth is contingent upon two variables: the depth of the morphophonological representation itself and how well the orthography approximates this representation. It means whether the morphophonological representation is consistent with the phonemic representation, or more simply, how regularly the letters represent the sound. Due to the varying consistency of letter–phoneme correspondences, different languages exhibit varying degrees of orthographic depth. Schmalz and colleagues (2015) suggested that orthographic depth comprises two distinct concepts: the degree of complexity and unpredictability of print-to-speech correspondences in a given orthography. Therefore, we can define a language’s transparency based on its orthographic depth, ranging from very shallow to very deep. The orthographic depth can be described as shallow when with more reliable correspondences, which is that one letter represents only one phoneme and sounds can be predicted from the spelling; if letters often represent more than one phoneme (which is also called feedforward inconsistency) or there is often more than one way to spell a phoneme (which also called feedback inconsistency), we can say this language has a deep orthographic depth (Pytlyk, 2017).

Although German and English have partly common roots, the German writing system is considered to have highly regular and consistent orthography-to-phonology correspondences, whereas English has a more irregular and inconsistent orthography-to-phonology relation. For example, in English, the letter “a” exhibits distinct pronunciations in words like “bank”, “ball”, and “park”, while in German, the letter is consistently pronounced the same in words such as “Ball”, “Park”, and “Bank” (Goswami et al., 2005). Due to its regularity, the pronunciation of nearly every German word can be deduced from its spelling once the speakers have mastered the spelling rules. The irregularity in German words comes mainly from loan words, proper names, and geographical terms (Ziegler, Perry, & Coltheart, 2000).

Moreover, Katz and Frost (1992) proposed the orthographic depth hypothesis, positing that the route of reading is dependent from the nature of orthography. They claimed that shallow orthographies are better able to facilitate a phonologically based word recognition process, whereas in deep orthography, readers need to use their knowledge of word morphology, which is extracted from the visual characteristics of written words, in order to comprehend printed text. Ellis et al. (2004) explored the influence of orthographic depth on reading acquisition across alphabetic scripts (Albanian, Greek, and English), syllabic scripts (Japanese *hiragana*), and logographic scripts (Japanese *kanji*). The study revealed that in scripts with deeper orthography, there was reduced latency associated with word length, an increased proportion of errors categorized as no-responses, and a higher tendency for substantive errors to involve whole-word substitutions rather than mispronunciations of nonwords. Consequently, they demonstrated that orthographic depth had an impact on both the speed and reading strategy employed.

Also, according to the psycholinguistic grain size theory, orthographic transparency will have an impact on the ability to learn to read (Goswami, 2010). There is already evidence showing that German children outperform English children in reading, and German adult readers also exhibit faster response times and higher accuracy during non-word reading compared to the English group. This difference in performance is attributed to English readers relying more on large orthographic units, such as rimes and words, so they need to rely on higher levels of orthographic consistency to recode the phonological units in smaller grain sizes, whereas German readers rely on smaller grain sizes, such as phonemes and letters, to reduce the ambiguity of grapheme-to-phoneme mappings within their orthographic system (Goswami, 2010; Seymour et al., 2003; Ziegler & Goswami, 2005).

But the above-mentioned hypothetical classification is only considering European languages. So, I will then briefly present the characteristics of the Chinese writing system.

2.3.2 The Chinese writing system

Chinese is a logographic language, and the character is regarded as the foundational unit of its writing system. In this thesis, when we mention “Chinese”, we are referring to “Mandarin Chinese”, which is the official language of China. Not like Indo-European languages such as German, inflectional or derivational morphology rarely exists in Chinese word formation (C. . Li & Thompson, 1981). Moreover, Chinese characters are mostly compound characters, so we can see that each character is composed of one or more radicals and these radicals contain basic strokes. The position of radicals in these compound characters is usually left–right or top–bottom. Interestingly, around 85% of Chinese characters include both a semantic radical, providing information about the character’s meaning, and a phonetic radical, which represents its pronunciation (Perfetti & Tan, 1998). For example, 晴 (/qing2/, *sunny*) has a semantic radical 日 which delicates the meaning of sun, and the phonetic radical is 青 /qing1/, which shares the same *Pinyin* with 晴 but with a different tone.

Pinyin is an official phonetic system for representing the pronunciation of Chinese characters, and it is based on the Latin alphabet which is the same as English. In the education system of mainland China, children are required to learn *Pinyin* by the end of the first and second grades of primary school. Each *Pinyin* representation is associated with its own syllabic tone, which is crucial for distinguishing meaning of syllables that are phonetically identical. Tones in Mandarin Chinese can be classified into four types, including 1. level contour, 2. rising, 3. falling - rising, and 4. falling. We should note that *Pinyin* is also a highly transparent alphabetic system, so even if someone hasn't seen the character, he/she could pronounce the character with help of *Pinyin.* However, it should be noted that *Pinyin* may not be phonemically encoded at the same level as traditional alphabetic orthographies. Instead, it primarily operates at the syllable level and, to some extent, at the subsyllabic level of onset and rhyme (Gottardo, Yan, Siegel & Wade-Woolley, 2001). For example, every Chinese character has only one syllabic unit, the *Pinyin* of 带 (/dai4/, *bring*) can be divided into onset *d* and rhyme *ai*, but we can't add any phoneme after the rhyme or present the syllable structure at body-coda level (e.g. *trust* as tru+st in English).

In addition, Chinese orthography differs from alphabetic orthography in terms of orthographic-phonological correspondence. In alphabetic languages like German, there exists a systematic correspondence between spelling and sound. But in Chinese, which is with logographic scripts, the mapping between pronunciation and spelling is basically arbitrary, which means that visually similar characters can have totally different pronunciations and meanings, while totally similar pronunciations can have different characters and meanings. For example, 柜(/gui4/, *cabinet*) and 拒 (/ju4/, *reject*) have very similar written forms but they are pronounced differently and with different meanings; 惊(/jing1/, *surprise*) and 晶(/jing1/, *crystal*) have the complete same pronunciation, but their orthographic form and meaning are fully different, which is also called homophony. This homophony of characters is another important feature of Chinese. Disregarding the tone of characters, about 5000 commonly used words in Chinese can be mapped onto about 400 distinct monosyllables (Qu & Damian, 2017). According to Zhou (1978), the phonetic components give clues to the pronunciation in only 38 % of Chinese characters. Therefore, from Chinese orthographic representation, the pronunciation cannot be deduced directly, but we can retrieve it from our memory (Patterson & Coltheart, 1987).

All these indications suggest an extremely weak link between orthography and phonology and frequent ambiguities in spoken words or syllables, particularly for characters with numerous homophones. Hence, orthography in written text is constructed to resolve homophony and to identify the meaning of a character (Qu & Damian, 2017). In other words, the orthographic form can distinguish homophonic morphemes.

In summary, the orthographic systems of languages are very different, which implies that the ways how orthographic representations are encoded during spoken word processing could vary from one language to another. While these orthographic systems represent sublexical units of the spoken language (such as phonemes, syllables, and morphophonemes) to varying extents, they all contain cues to the phonological information of written words. Even in the case of Chinese, where characters are predominantly compounds and homophones, the presence of the *Pinyin* system allows for pronunciation when reading characters becomes challenging.

2.4 The activation of orthography during spoken word processing

The overview of word recognition models in L1 as well as L2 has been provided in the last section, and the distinction between L1 and L2 word recognition as well as the different orthographic systems is introduced in sections 2.2 and 2.3. However, it is still not clear how orthography affects spoken word processing in L1 and L2. Therefore, in this section, I will present empirical studies that investigated the role of orthography during L1 and L2 spoken word recognition in alphabetic and logographic languages.

2.4.1 The role of orthography during spoken word recognition in L1 (alphabetic language)

In the present section, a closer look at the evidence from orthographic effects on L1 spoken word recognition in alphabetic language, which is relevant to the empirical part of my present study, will be provided.

Many studies supported that listening automatically activated the orthographic information of words online. Until the late 1990s, the investigation of orthographic effects on spoken word recognition primarily relied on metaphonological tasks, specifically rhyme judgment and phoneme manipulation tasks, which were commonly employed in linguistic studies. For example, Seidenberg and Tanenhaus (1979) explored how orthography influences auditory word recognition in three rhyme detection tasks, and their findings demonstrated that participants exhibited faster rhyme judgments for word pairs with orthographic similarity (e.g., *pie* - *tie*) compared to those that were orthographically different (e.g., *rye* - *tie*).

However, in Damian and Bowers' experiments (2010), when critical pairs were mixed with numerous fillers, the finding of Seidenberg and Tanenhaus (1979) was not replicable. In their experiment 1, they used identical materials and procedures as in the original study conducted by Seidenberg and Tanenhaus (1979), and the results showed that orthographic similarity had a significant impact on response times for word pairs that rhymed, while it hindered responses for pairs that did not rhyme. In experiments 2 and 3, when they manipulated the nature of the non-rhymes or added a large number of filler items, the orthographic effect was eliminated. This finding suggested that some strategic factors could be triggered during the rhyme judgment and the orthographic effect was not really generated.

Indeed, the presence of orthographic effects has been demonstrated to exist in various other tasks as well. Ziegler & Ferrand (1998) investigated the influence of orthographic consistency on auditory word processing with the lexical decision task, in which they manipulated the orthographic consistency of English spoken words in two categories: phonological rhymes of words could be spelled in many ways (inconsistent) or could be spelled in only one way (consistent). The results of the study revealed a notable different result, which is that consistent words with rhymes that can be spelled in only one way contribute to faster responses, in contrast to inconsistent words with rhymes that can be spelled in multiple ways. Recent studies have predominantly employed lexical decision tasks to investigate the orthographic role in spoken language processing. Such as the study from Perre et al. (2009), lexical decision task was also used to test whether orthographic information affects phonological priming in spoken word recognition, and the result showed that native English speakers respond more quickly to word pairs with both orthographic and phonological overlap, compared to pairs with only phonological similarity. However, this method has a potential weakness, as participants would strategically generate an orthographic image of the spoken word, which could influence their decisions on the word's lexical status (Pattamadilok et al., 2009; Qu & Damian, 2017). Hence, Pattamadilok et al. (2009) chose a semantic task, wherein participants listened to words with a focus on meaning, without thinking about the orthographic form, thus minimizing strategic factors. They used a go/no-go paradigm, where participants were required to press a button if they recognized the name of a human body part (go) and did not respond if they believed the word did not belong to a part of the human body (no-go). The results revealed that orthography indeed exerts a nonstrategic influence on spoken language processing (Pattamadilok et al., 2009). Besides, in semantic and gender categorization tasks, the orthographic consistency effect could also be found (Peereman, Dufour & Burt, 2009). As a result, it becomes evident that orthographic effects persist not only in lexical decisions or metaphonological tasks but also in semantic tasks. The semantic task affords the opportunity to explore not only the influence of phonology-orthography inconsistency on lexical processing without strategic influence but also the implicit role of orthography in retrieving semantic information. Thus, this approach will be adopted as the main experimental method in my study.

In contrast to these studies which supported an effect of orthography on lexical processing, some other researchers didn't find any effects of orthography. Ventura and collegeaus (2004) found that in Portuguese auditory lexical decision tasks, inconsistent words produced longer latencies and more errors than consistent words, but no orthographic consistency effect exists in standard shadowing tasks. Subsequently, the researchers conducted a comparison between two conditions, wherein a shadowing response depends on either a lexical or a phonemic criterion. The results revealed that only in the lexically contingent shadowing condition, orthographic consistency exerted a significant influence. Hence, they suggested that there is no effect of orthographic consistency on pre-lexical processes, but instead on lexical processes. This finding was replicated by Pattamadilok, Morais, Ventura, and Kolinsky (2007), but this time using French, a much more orthographically inconsistent language than Portuguese. The results were in line with those from Ventura et al. (2004): In the auditory lexical judgment task, inconsistent words caused longer responses than consistent words, whereas the word consistency effect was observed in the shadowing task when words were presented either combined with pseudowords or alone. Rastle et al. (2011) also confirmed the lack of spelling-sound consistency effects in a shadowing task, while robust orthographic effects were observed in a picture naming task. The authors provided an explanation that in the shadowing task, phonological activation might take over before orthographic input has an opportunity to influence speech production, while the picture naming task involves additional processing stages and extended processing time, and inconsistent orthographic feedback activates conflicting phonological representations, leading to delayed responses. Moreover, Türk and Domahs (2022) investigated the role of orthography during spoken word recognition with German native speakers and employed a similar paradigm from Perre et al. (2009), but didn't find a significant influence of orthography, which existed in the experiment conducted in English from Perre et al. (2009). So they demonstrated that the effect of orthography might be language-specific and depends on the orthographic depth of the target language. Some researchers have claimed that the effect of orthography may indicate strategic adaptations to a specific task environment, rather than an obligatory effect on the prelexical processing of speech (Cutler, Treiman, & van Ooijen, 2010; Cutler & Davis 2012). In this fact, orthography is seen to have only little role to play in conversational speech (Mitterer & Reinisch, 2015).

These findings indicate that the influence of orthography during L1 spoken word recognition is not always present when it is tested with different experimental methods, in different target languages, and different environments. Therefore, this dissertation will test the existence of an orthographic effect in spoken word recognition with a more transparent language, German, using a different priming paradigm, namely a semantic judgment task.

2.4.2 The role of orthography during spoken word recognition in L1 (Chinese)

2.4.2.1 The lexical processing of Chinese

Given the unique phonological and orthographic structure of Chinese, the lexical processing of Chinese varies from that of an alphabetic language. Thus, the previously mentioned models, which are suitable for explaining alphabetic word recognition, may not be as applicable to Chinese word recognition as to English due to the fundamental differences in the writing systems between alphabetic and logographic languages.
According to the multilevel-interactive model from Taft, Zhu & Peng (1999), there are three units in the Chinese lexical processing system: orthographic units, phonological units, and semantic units (see Figure 10). When the written form of a word is presented, the processing system is initiated through the activation of orthographic units, focusing on the lowest-level features, such as individual strokes and stroke combinations. Subsequently, the radical units are engaged and transmit their activation to the character units, eventually reaching the multicharacter units. Also at both the character and multicharacter levels, relevant phonological and semantic units can be activated (Taft & Zhu, 1995, 1997), implying that radical units are directly associated with both semantic and phonological representations. In this model, it is evident that radicals play a crucial role as input units for Chinese character recognition. Radicals are comparable with letters in an alphabetic script since they represent the smallest units associated with specific features (Taft et al., 1999). However, a key distinction is that radicals are sensitive to positional information, such as left-right and top-bottom. Consequently, during the recognition process of a compound character, an inhibitory or facilitatory effect might arise when characters share the same radicals.

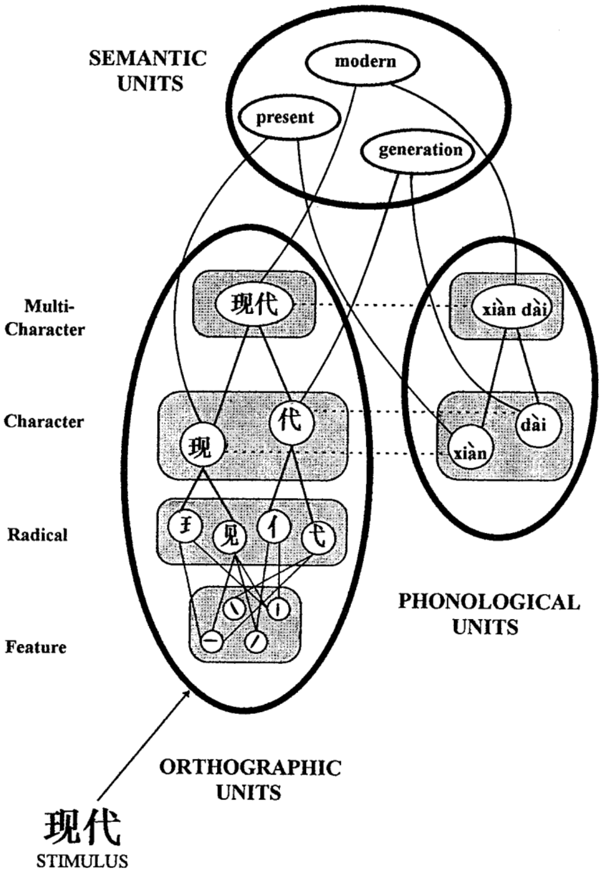


Figure 10. The multilevel-interactive model. The example of 现代 (/xian4 dai4/, *modern*) is used to illustrate the different levels of representational units. (Taft, Zhu & Peng, 1999)

We can still find some common roots in this multilevel-interactive model with the interactive model of alphabetic word recognition, that is, the phonological units would be activated as the orthographic units are accessed.

In fact, the phonological effect during Chinese word processing interests many researchers, due to the weak correspondence between the written form and pronunciation, phonological effects are less expected in Chinese compared to alphabetical languages. But still, some researchers claim that phonology is very important to Chinese character recognition. Weekes, Chen, and Lin (1998) investigated how phonological priming influences the recognition of two types of Chinese characters: compound targets containing separate radical units and integrated targets without separate radicals. The results revealed that phonology was activated only when processing Chinese compound characters. Also, Tan and Perfetti (1999) supported an obligatory role of phonological processing in the identification of Chinese characters, and Ziegler and his colleagues (2000) found that characters with a high phonological frequency elicited faster reading responses in comparison to characters with a low phonological frequency.

Additionally, there have been researchers who have found that phonology may not play such a significant role in the process of recognition. For instance, Wu and Chen (2000) did not provide support for an obligatory role of phonological processing in the identification of Chinese characters, but they found a facilitation effect of homophone priming in the naming task. Similarly, in a study conducted by Chen, Vaid, and Wu (2009), there was no facilitatory effect of phonological frequency.

Due to the differences between the processing of printed words and spoken words, it is important to mention one model of spoken word recognition of tonal language - the TTRACE model (Tong et al., 2014), which was modified based on the traditional TRACE model (see section 2.1.3). This model aims to integrate tone processing into spoken word recognition of Cantonese, a dialect within the Chinese branch. Most models for spoken word processing have actually been proposed for nontonal alphabetical languages, but the tone is the main characteristic of the Chinese language and greatly influences spoken word recognition (Lee, 2007; Zhao, Guo, Zhou, & Shu, 2011).

As Figure 11 shows, the TTRACE model is composed of a feature level, a phoneme-toneme level, and a word level, with competitive relations at each level, and with interactive relations between levels, just like the original TRACE model. What is different from the traditional TRACE model is that the TTRACE model integrates segmental (power, vocalic, acute, consonantal, voiced, burst) and suprasegmental dimensions (contour, height, onset, and offset) in a distributed network, and phonemes and tonemes are connected with each other; in addition, the similarity between target and nontarget words is determined by varied segmental and suprasegmental dimensions, as well as symbolized by the thickness of curved dotted lines as described in Figure 11; and these factors together determine the degree of word activation (Tong et al., 2014). For example, the target word in Figure 11 is /fu1/ (skin), and there are other words with different degrees of phonological similarities with the target word, such as sharing two of the three aspects with /fu1/: the same vowel and tone (/wu1/, *black*), the same consonant and vowel (/fu6/, *father*), or the same consonant and tone (/fa1/, *flower*), are most strongly activated; in decreasing order of activation strength, the words which share only one aspect with the target are less activated, and which share no similarity consequently display the least activation.

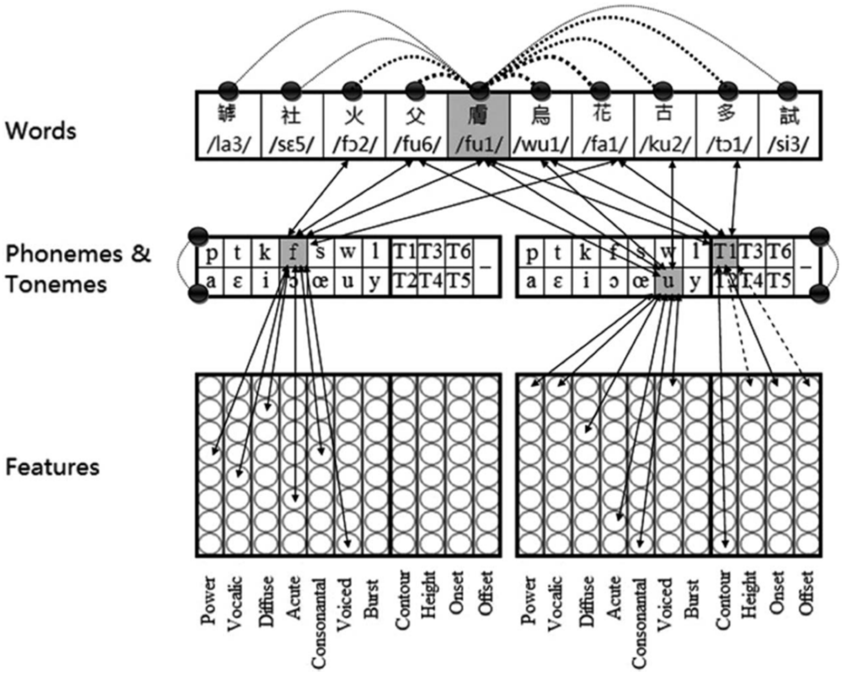


Figure 11. The TRACE model for speech perception of tonal languages (TTRACE). (Tong et al., 2014)

Having introduced the model of lexical processing of Chinese, it is evident that tone plays an important role in Chinese. This is due to the fact that distinct tones can generate entirely different characters, even when sharing the same combination of letters — a phenomenon uncommon in alphabetic languages. The presence of homophony in

Chinese leads to ambiguity in auditory word recognition, implying the importance of relying on orthographic forms to differentiate characters from one another. As a consequence, a question about whether orthographic information influences Chinese word recognition becomes particularly interesting.

2.4.2.2 Evidence of orthographic effect during Chinese word processing

Studies investigating the impact of orthography on spoken language have primarily focused on alphabetic writing systems. Hence, early studies neglected to explore the influence of orthographic similarity on Chinese spoken word recognition. However, in recent decades, an increasing number of researchers have shown interest in investigating orthographic activation in logographic writing systems like Chinese. However, whether the orthographic effect exists, remains controversial.

On the one hand, some studies indicated the important role of orthographic information in auditory Chinese word recognition. Zou and collegeus (2012) conducted an auditory lexical decision task, wherein they manipulated orthographic and phonological overlap between the first syllable of prime-target pairs in four conditions: P+O+, P+O-, P-O+, P-O- ("P±" means phonologically similar/dissimilar, "O±" means orthographically similar/dissimilar). The findings revealed that N400 amplitudes were significantly reduced when a target was primed by an orthographically similar word. Furthermore, Qu and Damian (2017) investigated the orthographic influence with a semantic relatedness judgment task, and native Chinese speakers need to judge whether or not the word pairs they heard were related in semantics. Word pairs were categorized as either semantically related, orthographically related, or unrelated. The result showed that judgments were faster for semantically related word pairs compared to unrelated ones. Importantly, when there was orthographic overlap in semantically unrelated word pairs, it led to a notable increase in response latencies. Besides, Mok, Lee, Li, and Xu (2018) suggested that orthographic effects are not just related to alphabetic systems, they also depend on the nature of the task and the language skill of the learner.

On the other hand, there are several studies that provide evidence for meaning access during Chinese spoken word recognition without orthographic activation. For instance, Wu and Thierry (2010) conducted a study with Chinese-English bilingual participants, and in the study, the unrelated English word pairs had either a sound or a spelling repetition in their corresponding Chinese translations. Results showed that sound repetition elicited smaller ERP amplitude but spelling repetition was the same as the control group, and suggested that processing a second language activates the phonological information, but not the orthographic information of their native language. Similarly, Wang, Li, Ning, and Zhang (2012) found an inhibitory homophone density effect for Chinese monosyllabic homophones, which is related to semantic processing rather than orthographic processing.

In summary, the orthographic effect was not always found in Chinese spoken word recognition.

2.4.3 The role of orthography during spoken word recognition in L2

Numerous studies have already explored the influence of orthography on word processing in L1. However, it is worth noting that native speakers generally encounter more spoken language than written language, while L2 learners typically receive more exposure to written language, particularly at the initial stages of their learning journey. As a result, we can hypothesize that orthographic information may play an even more pronounced role in L2 word processing compared to L1. There are indeed some studies that support the important role of orthography in spoken word processing. For instance, Pytlyk (2017) found that L2 phoneme awareness is influenced by L2 orthography in an auditory phoneme counting task. Similarly, Hao and Yang (2018) observed that written characters are more effective than *Pinyin* in assisting advanced English learners of Chinese in encoding the tones of unfamiliar Chinese words. Since there is no direct one-to-one grapheme-to-phoneme correspondence and each character corresponds to one syllable, encompassing both segments and tone, the input of characters likely promotes a holistic representation of sounds that encompasses both segmental and tonal dimensions. As a result, advanced English learners of Chinese could gain advantages from the integrated representation of sounds in characters (Hao & Yang, 2018). Türk and Domahs (2022) reported a facilitative orthographic priming effect in English spoken word recognition when participants are late German-English bilinguals.

Additionally, language proficiency might be a critical factor to consider when investigating word recognition in L2. Only a few studies have examined the effects of varying proficiency levels on L2 orthographic processing. Veivo and Järvikivi (2013) used a masked cross-modal priming paradigm to investigate orthographic and phonological processing in L2 spoken word recognition by Finnish learners of French. The study's findings indicated that in cases where there was no phonological or semantic overlap, the high-proficiency group exhibited stronger repetition priming effects and significant facilitation from shared orthography between L1 Finnish and L2 French. Conversely, the group with lower proficiency demonstrated no orthographic effect but showed a significant L1 pseudohomophone facilitation instead. Similar results are presented by Veivo, Järvikivi, Porretta, and Hyönä (2016). They used an eye-tracking experiment to observe the orthographic activation in L2 spoken word recognition and found no general orthographic effects in the L2 group. However, they did uncover an important finding related to language proficiency in L2 during spoken word orthographic processing, that is, only higher proficiency L2 learners used orthographic information in the matching task. Besides, consistent results were also found in Mitsugi's (2018) study, where the activation of orthography during speech processing was supported among native Japanese speakers. In this research, L2 proficiency was identified as a crucial factor influencing word recognition performance for participants who were learners of Japanese with an L1 English background. These findings suggest whether orthography indeed plays a role in spoken word recognition among L2 learners appears to be contingent upon the learners' proficiency level.

On account of the language-specificity, the question of how L1 influences L2, especially when L1 is a non-alphabetic language and L2 is an alphabetic language, remains to be investigated. In the next section, we will report results about the influence of L1 writing system on L2 word recognition.

2.5 The influence of L1 orthographic knowledge on L2 word recognition

Currently, there is considerable debate among researchers concerning the impact of L1 orthographic background on L2 word recognition.

Several studies have presented evidence suggesting that L1 orthographic knowledge does indeed influence the processing of non-native words. In the study by Chikamatsu (1996), English participants employed phonological information in Japanese *kana* words more frequently than Chinese participants did, but Chinese partcipants relied more on the visual information in Japanse *kana* in contrast to English participants; the results suggested that the different word recognition strategies rely on different L1 characteristics of orthography. Similarly, Akamatsu (2003) conducted a study to explore the effects of L1 orthographic features on L2 reading and the results indicated that both the Chinese and Japanese groups demonstrated lower efficiency in processing English words compared to the Persian group. Besides, In Martin's (2017) study, a comparison was made between the English spelling knowledge of L2 learners from three different L1 backgrounds (French, Hebrew, Chinese) and English native speakers. The findings of the study demonstrated that participants from non-alphabetic languages (Hebrew and Chinese) performed notably worse on items containing vowel-related misspellings when compared to those with consonant-related misspellings; the accuracy in distinguishing between vowels and consonants varied across L1 speakers, indicating that the L1 writing system of learners influences the development of their L2 orthographic proficiency and their ability to detect diverse types of word misspellings. These studies offer valuable insights into the crucial role of the writing system in influencing word processing through orthographic representation. When recognizing spoken or written words, listeners associate each phoneme with the corresponding spelling stored in their minds, enabling access to the word’s meaning. This process is known as the orthographic mapping of a word. For listeners who are unfamiliar with phoneme-based languages or are learning an L2 with deep orthography, they may encounter challenges in orthographic mapping and may rely on processing strategies from their L1. Conversely, individuals with highly proficient phonological awareness and a strong grasp of letter-sound correspondence can decode words by identifying their sounds letter by letter (Ehri, 2014; Kilpatrick, 2015), leading to more successful orthographic mapping.

Other researchers claim that L1 orthographic knowledge does not influence non-native word processing. According to Sun-Alperin and Wang (2011), orthographic patterns are often language-specific and not likely to have the same role in spelling performance. They found that Spanish orthography (L1) facilitated English (L2) reading, but it did not have the same effect on spelling, despite the visual similarities between the two orthographies. One possibility for this observation could be attributed to the highly regular letter-phoneme mappings in Spanish. As a result, native Spanish speakers may encounter difficulties in extracting English phonology due to the highly irregularity of letter-phoneme correspondences prevalent in the English language (Vokic, 2011), so the Spanish orthography (L1) could not facilitate English (L2) spelling.
Furthermore, Pytlyk (2017) conducted a study investigating L1 and L2 orthographic effects on L2 phoneme perception among English learners of Russian and Chinese with intermediate-level proficiency. Both Russian and Chinese are classified as deep orthography with non-transparent phoneme-letter correspondences. The results of the study revealed that learners exhibited greater success in phoneme counting for L2 words with consistent letter-phoneme correspondences when compared to words with inconsistent correspondences. This finding indicated that L1 orthography did not significantly impact L2 phoneme perception.

Driven by the viewpoint that the orthographic form plays an important role in spoken word recognition not only in alphabetic but also in non-alphabetic languages, and considering the limited evidence available from L2 groups with non-alphabetic language backgrounds, my study attempted to determine the extent of orthographic activation during L1 and also in L2 spoken word processing, focusing on Chinese learners of German with different L2 proficiency. Moreover, while most existing studies on word recognition have centered on phonological effects in visual word processing, my research also considers the influence of phonological priming in the context of spoken word processing.

To summarize, several unresolved questions surround the roles of orthography and phonology in spoken word recognition. Firstly, it remains unclear how different the involvement of orthographic and phonological information is in both L1 and L2 word recognition, particularly concerning the influence of proficiency in L2 word processing. Additionally, the comparison between different writing systems and their impact on the L1 orthographic role during word recognition requires further investigation. Moreover, the effects of immersion in an L2-spoken environment for L1 speakers and the potential reduction of orthographic knowledge due to living abroad have not been explored from the perspective of orthographic activation. The forthcoming experiments, detailed in the following chapter, are designed to address these research questions comprehensively.

Chapter 3. Empirical investigation of the influence of orthography on spoken word recognition

As shown in Chapter 2, the models of processing and correspondence between phonological and orthographic representations, which have been found in the L1 with an alphabetic language background, may not suit L2 word recognition when the L2 listeners have a non-alphabetic language background. Based on that, this study aims to address this gap by investigating a group of Chinese–German late bilinguals with distinct L2 proficiency levels, residing in an L2-dominant environment.

In this chapter, I will first present the overall design of the empirical research, subsequently followed by a description of the series of experiments. The description will encompass various methodological aspects, the obtained results, and subsequent discussions.

3.1 The overall design of the empirical research

The research reported here will be presented as four separate experiments.

Experiment 1 examines the impact of orthographic and phonological similarities on L1 spoken word recognition, with a specific emphasis on the alphabetic language German. Therefore, the participants involved in this experiment are native speakers of German.

In Experiment 2, the focus shifts to the orthographic influence on L2 spoken word recognition. The participants in this experiment are Chinese non-native speakers of German, representing different levels of language proficiency. To facilitate a meaningful comparison between groups, the stimuli and paradigm used in both Experiment 1 and Experiment 2 remain identical, allowing for a comparison of the performance differences among various participant groups, including those with German as their L1 and L2, as well as those with intermediate and high proficiency in German.

The selection of these two participant groups (native German and Chinese speakers) was guided by several considerations. Firstly, the majority of previous research investigating orthographic effects has predominantly used English stimuli in monolingual studies. Similarly, studies on bilingualism have often focused on English as the second language choice. Hence, there exists a gap in the literature regarding the orthographic effects in languages other than English, as well as in bilinguals with different language backgrounds. As mentioned in section 2.3.1, in contrast to English with deep alphabetic orthography, German has a rather shallow orthography, meaning that grapheme-to-phoneme mappings are mostly regular and predictable, and Chinese employs a distinct writing system that differs from English. By including German and Chinese native speakers, we aimed to address this gap and provide valuable insights into orthographic effects in a non-English context. Another motivation behind selecting German and Chinese native speakers as participants stems from the observation that most L2 studies have primarily examined languages involving alphabetic writing systems, neglecting the inclusion of non-alphabetic languages. Therefore, it is very meaningful to know whether individuals with a non-alphabetic language background exhibit differential recognition patterns of alphabetic orthographic information. In Experiments 1 and 2, the relevant linguistic and non-linguistic confounding variables, such as word frequency, language proficiency, age of L2 acquisition, length of residence in the German-speaking country, etc., were controlled for.

The aim of Experiments 3 and 4 is to investigate the role of orthography and phonology on Chinese spoken word recognition. Both experiments used the same materials and instructions, but the participants were divided into two groups: those living in Germany (Experiment 3) and those residing in China (Experiment 4). Participants in Experiment 3 were a subgroup of the participants from Experiment 2, particularly those who had been in Germany for at least one year and had achieved intermediate to high proficiency in the German language. During their time in a German-speaking environment, their exposure to and usage of Chinese was heavily limited. In contrast, participants in Experiment 4 had no prior experience learning German and had not lived abroad, relying exclusively on Chinese for their daily communication. Given the crucial role of orthography in Chinese lexical processing and the potential impact of reduced Chinese language usage on the quality of orthographic representation, we aimed to explore whether these two groups exhibited differences in their L1 spoken word processing.

To explore the orthographic influence on L1 and L2 spoken word recognition, I conducted semantic judgment tasks employing auditory presentation to manipulate the association between phonologically and orthographically related primes and targets.

Actually, the lexical decision task is a commonly used method for examining the influence of various linguistic factors on spoken word recognition research. But recently, the lexical decision task has been questioned for testing orthographic activation in alphabetic languages. It has been claimed that it may create an orthographic image of the spoken word, affecting reaction times during word recognition (Pattamadilok et al., 2009; Qu & Damian, 2017). In comparison to the lexical decision task, semantic judgment tasks are commonly considered strategy-free, as participants need to focus on the meaning of spoken words without explicitly analyzing the form representations (Pattamadilok et al., 2009; Qu & Damian, 2017). Considering the aim of the research, the semantic judgment task was selected as the main method in the present studies, allowing for implicit manipulation of orthographic and phonological information. In experiments 3 and 4, the method presented by Qu and Damian (2017) has been adopted. The original study only focused on orthographically and semantically related words without incorporating a phonological-related condition. In the present study, word pairs with exclusive phonological similarity were included in the stimuli to investigate whether phonological similarity influences spoken word recognition to the same extent as orthographic overlap.

Based on previous findings described in Chapter 2, orthography plays an important role in the Chinese language, due to the deep orthography with an arbitrary mapping between written characters and phonological form. Additionally, the L1 orthography might impact the L2 processing. Consequently, we predicted stronger orthographic effects in the Chinese L2 group compared to the German L1 group. In contrast, we anticipated that phonological similarities would have a more pronounced impact on spoken word recognition in the L1 group compared to the L2 group. This hypothesis was based on the assumption that the L2 group had greater exposure to the written form but less exposure to spoken language during their language learning process. Nevertheless, it could be the case that the orthographic and phonological effects are especially prominent in L2 learners with high language proficiency.

Concerning Chinese native speakers, it was predicted that the stronger impact of orthographic information on spoken word recognition was found in the Chinese group living in China compared to those who live in Germany. If the current study's findings suggest that the phonological and orthographic effect sizes differ from those shown in Chapter 2, spoken word processing may differ in languages with different writing systems and proficiency levels.

3.2 Experiment 1: Experiment for German participants as L1 group

The present experiment was performed with German native participants who only took part in the German semantic judgment task. This group is regarded as a control group in the whole research.

3.2.1 Method

Participants

The L1 group contains thirty-seven native German speakers with normal hearing. Among them, 27 were university students in Germany, and 10 were employed. Their mean age was 25 years (ranging from 18 to 38 years) and the group consisted of 20 women and 17 men. All participants indicated that German was their L1 and none of them claimed an early bilingual background. 29 participants from this group were tested in the behavioral laboratory at the University of Marburg, and eight participants took part in this experiment in Frankfurt.

Materials and design

Before we further describe the stimuli used in the semantic judgment task, we need to talk about *semantic relatedness*. The terminology *semantic relatedness* can also be defined as the *semantic proximity* or *semantic association*, i.e., how strong connections between two concepts can be drawn (Taieb, Zesch & Aouicha, 2020). Another terminology that frequently occurs in many studies is *semantic similarity*, which is often confused with *semantic relatedness*. Actually, *semantic similarity* has been identified as a particular subset of *semantic relatedness* and involves any relation between two expressions, but *semantic similarity* is defined as a "is a"-relation (Ballatore, Bertolotto & Wilson, 2014). For example, *Zug (train)* and *Flugzeug (airplane)* are both meanings of KIND OF PUBLIC TRANSPORTATION, so we can say that these two words are semantically similar. In contrast, *Zug (train)* and *Gleis (track)* are semantically related but not semantically similar, because *Gleis (track)* is a facility that keeps the train running. Both meanings are in an associative connection, but they are not denoting similar things. The word pairs employed in the experiment were selected to either demonstrate obvious semantic relatedness or lack thereof, ensuring that participants could make unambiguous judgments.

The critical stimuli used in this experiment consist of 91 word-sets, each consisting of one target word and three prime words. Prime-target pairs are semantically unrelated but include the following relations with regard to orthographic or phonological overlap:

- O+P+: orthographically and phonologically related (e.g., Kleid, *dress* - Neid, *jealousy*);
- O-P+: phonologically related, but orthographically unrelated (e.g., Kleid, *dress* - Streit, *quarrel*);
- O-P-: unrelated in phonology or orthography (e.g., Kleid, *dress* - Stuhl, *chair*).

Note that the related part is in the position of rhymes. Most stimuli were taken from the stimulus list used by Türk and Domahs (2022), but considering that Experiment 2 was run with non-native speakers of German, some low-frequency words had to be replaced by words that were more suitable for the L2 learners. In total, we selected 273 word pairs as critical stimuli, in addition to 210 word pairs that functioned as fillers. From the fillers, 160 pairs were semantically related (e.g., *Leiter* and *Führer* mean *leader* in English) and 90 pairs semantically unrelated (e.g., Kurs*, course* - Lachs*, salmon*). Note that across all filler sets, the primes had no orthographic or phonological overlap with the respective targets. The full list of stimuli is presented in Appendix A.

In order to assess the degree of semantic relatedness, I conducted a rating task with native German speakers. A questionnaire was created using SoSci Survey (Leiner, 2019), a professional tool for online surveys, and was completed by 30 German native speakers who did not participate in the reaction time experiments. These participants were presented with multiple word pairs and were asked to make subjective judgments regarding the semantic relatedness of each pair. Specifically, they were instructed to rate the degree of connection between the words on a scale ranging from 0 to 6. A rating of 0 indicated no connection, while a rating of 6 signified a very strong connection. The results are summarized in Table 1.

What can be seen in Table 1 is the difference between semantically related and unrelated stimuli. Moreover, the correlation between the semantic relatedness scores and reaction times of critical stimuli in Experiments 1 and 2 was calculated and the results showed that there was no correlation between these two variables (in Experiment 1: *Pearson's* $r = 0.019$, $p = 0.754$; in Experiment 2: *Pearson's* $r = -0.076$, $p = 0.212$). Hence, we would not regard the semantic relatedness score as a dependent factor in our data analysis.

Table 1. Summary of the semantic relatedness scores rated by German native speakers

condition	**N**	**Mean**	**SD**	**SE**
O+P+	91	2.12	0.707	0.074
O-P+	91	1.77	0.488	0.051
O-P-	91	1.43	0.444	0.047
Filler_S+	150	4.55	0.627	0.051
Filler_S-	60	1.34	0.253	0.033

condition	**N**	**Mean**	**SD**	**SE**

S+/- : semantically related/unrelated
O+/- : orthographically related/unrelated
P+/- : phonologically related/unrelated

All the words were mono- or bi-syllabic German nouns/verbs and were matched as well as possible for word frequencies. Word frequency measures were taken from the dlexDB corpus (Heister, Würzner, Bubenzer, Pohl, Hanneforth, Geyken & Kliegl, 2011), which is based on the reference corpus of the German language of the 20th century compiled by the Digital Dictionary of the German Language (DWDS) at the Berlin-Brandenburg Academy of Science (BBAW). The dlexDB corpus includes 100 million tokens and provides counts on orthographic neighborhood size. The mean frequency across conditions was per million words of the corpus. Measures refer to normalized type frequency, which is calculated as the count of distinct word forms divided by 1 million tokens in the dlexDB corpus. These values ranged from 0.03 to 336 (Mean = 21.6; SD = 38.5). The mean frequency of the target items was 26.7 per million and the primes in every condition are matched with that in target items (see Table 3).

Additionally, the analysis included factors such as word length and the size of the orthographic and phonological neighborhoods (also known as neighborhood density). Orthographic neighborhood size which is denoted as N, is commonly defined as the number of words that can be formed by altering a single letter of a target word while maintaining letter positions (Coltheart, Davelaar, Jonasson, & Besner, 1977). The method used to measure phonological neighborhood density is similar to calculating orthographic neighborhood size, but instead of letter substitution, it involves the addition, deletion, or substitution of a single phoneme (Luce & Pisoni, 1998). A large number of studies have reported the impact of phonological and orthographic neighborhood density on word processing in alphabetic languages (Andrews, 1997; Davis & Perea, 2005; Grainger et al., 2005; Siakaluk et al., 2002), so matching the stimuli with their neighborhood size is obviously necessary.

The neighborhood size in our study was calculated by using the CLEARPOND matching tool (Marian et al., 2012). CLEARPOND is the abbreviation for "Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities" and provides an interface for many languages. All characteristics of the stimulus materials are displayed in Table 2 and Table 3, and the results of post-hoc tests (with Bonferroni-corrected *p*) are shown in Table 4. As can be seen from Table 4, the differences between the targets' and primes' frequency, orthographic and phonological neighborhood size and word length counts were all non-significant across conditions.

Table 2. Overall characteristics of the stimulus materials

	Frequency_ dlexDB	**Ortho. neighbourhood**	**Phon. neighbourhood**	**Word length**
Mean	21.6	6.29	13.7	4.85
SD	38.5	4.11	9.82	1.17
Min.	0.03	0.00	0.00	3
Max.	336	22.0	57.0	9

Table 3. Characteristics of the stimuli in every condition

	condition	**Frequency_ dlexDB**	**Ortho. neighbourhood**	**Phon. neighbourhood**	**Word length**
Mean	**Target**	26.7	6.55	13.7	4.80
	O+P+	17.6	6.54	15.2	4.84
	O-P+	14.3	5.21	13.9	4.76
	O-P-	27.8	6.67	12.2	5.02
SD	**Target**	40.4	4.30	9.31	1.08

	O+P+	41.0	4.38	10.4	1.27
	O-P+	27.6	3.48	9.23	1.19
	O-P-	42.2	4.06	10.1	1.13
Min.	**Target**	0.0300	0.00	0.00	3
	O+P+	0.0300	0.00	1.00	3
	O-P+	0.0700	0.00	1.00	3
	O-P-	0.740	1.00	0.00	3
Max.	**Target**	173	22.0	40.0	7
	O+P+	315	16.0	45.0	9
	O-P+	148	16.0	48.0	8

Table 4. Post hoc tests results of the stimulus materials

	condition	**condition**	**Mean Difference**	**SE**	**df**	**t**	***p***
Frequency_dlexDB	**Target**	**O+P+**	9.12	5.67	360	1.607	0.653
	Target	**O-P+**	12.43	5.67	360	2.191	0.174
	Target	**O-P-**	-1.01	5.67	360	-0.178	1.000
Ortho. neighbourhood	**Target**	**O+P+**	0.0125	0.646	318	0.0194	1.000
	Target	**O-P+**	1.3387	0.666	318	2.0105	0.271
	Target	**O-P-**	-0.1203	0.626	318	-0.1923	1.000
Phon. neighbourhood	**Target**	**O+P+**	-1.463	1.55	318	-0.943	1.000
	Target	**O-P+**	-0.244	1.60	318	-0.152	1.000
	Target	**O-P-**	1.491	1.50	318	0.992	1.000
Word length	**Target**	**O+P+**	-0.0330	0.173	360	-0.191	1.000

	Target	**O-P+**	0.0440	0.173	360	0.254	1.000
	Target	**O-P-**	-0.2198	0.173	360	-1.271	1.000

273 word pairs (91 x 3 conditions) were distributed in a balanced way in three versions. In this way, each participant received 31 orthographically and phonologically related prime-target pairs, 31 phonologically related but orthographically unrelated prime-target pairs, and 31 pairs that were unrelated in all properties, to ensure that each version consisted of only one of the three priming conditions per target and participants heard each target item only once during the experiment. They also received 150 fillers matched in pairs with semantic relatedness, as well as 60 fillers without semantic relatedness, so that the whole word pairs were almost equally distributed over semantically related and unrelated pairs. The participants were randomly assigned to one of the three versions and the selection of versions was based on the order in which they participated in the experiment.

Procedure and Apparatus

The participants were individually tested in a quiet room. The entire experiment session was run on a 24-inch desktop (Dell) and lasted approximately 30 minutes. Stimuli were presented using OpenSesame presentation software (Mathôt et al., 2012). The stimuli presented via BeyerDynamics DT 900 headphones were produced by a female native speaker of German, using Behringer Xenyx X2442 mixing console with Audacity audio recording software (Audacity Team, 2019). The stimuli were edited afterward using PRAAT software (Boersma & Weenink, 2020) to remove noise and to normalize the sound intensity.

At the beginning of the experiment, participants were instructed to determine whether the two spoken words they heard were semantically related or not, and to indicate their response as soon as possible by pressing two distinct keys. “Yes”-responses were given by pressing a green button of a response box (LOBES version 5/6) and “no” by a red button. The assignment of the two response buttons to the left or right hand was counterbalanced across participants (the response box was head against or reversely placed to participants), to avoid faster key presses caused by right- or left-handedness. The stimuli were preceded by 12 practice trials, half of which were in the semantically related condition. The experimental block consisted of five blocks, with each block containing 60 trials (the last block contains 63 trials), and participants were allowed to take a short break at the end of each block. Figure 12 presents the timeline of the experiment. In each trial, participants saw firstly a fixation cross at the center of the screen for 300 ms. Following the fixation cross, they heard the prime word, and after 20 ms the fixation cross turned to a question mark coinciding with the presentation of the target word. The participants were allowed to make a judgment as soon as the target word started to play. A timeout of 3500 ms forced the end of the trial in cases of null responses. After the key press, a blank screen was displayed for a duration of 1000 milliseconds before the onset of the next trial. Response times and judgments were recorded for later data analyses.

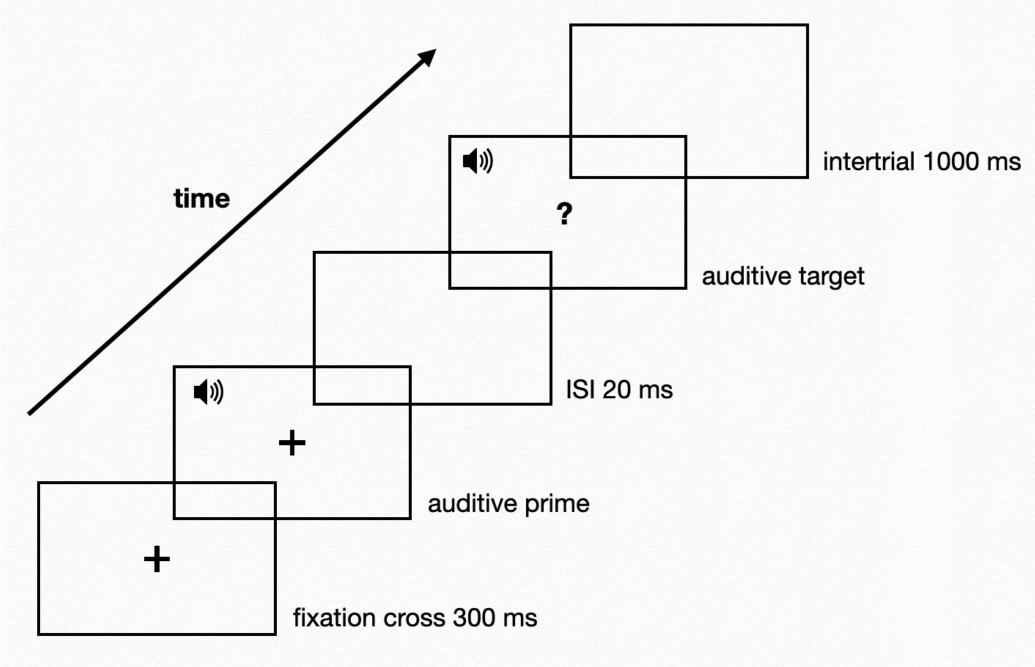


Figure 12. The timeline of the experimental trials in Experiment 1

3.2.2 Results

Prior to statistical analyses, a comparison was made between semantically related fillers and semantically unrelated fillers, as well as critical word pairs. The results of pairwise comparisons indicated that semantically related fillers elicited significantly faster responses ($t = -4.673$, $p < 0.001$) and a significantly higher error rate ($z = -13.130$, $p < 0.001$) compared to the other conditions. This outcome validates the effectiveness of the semantic relatedness task assigned to the participants.

For the response time analysis of critical stimuli, all incorrect responses (less than 10.3%) were excluded from further analyses. Furthermore, RTs smaller than the first quartile minus 3 times the IQR (Interquartile range: the difference between the first and the third quartiles) and larger than the third quartile plus 3 times the IQR of correct responses were discarded (less than 1.2%) (Tukey, 1977). Within the condition O+P+, a word pair (Fuchs – Luchs, *fox – lynx*) exhibited a remarkably high semantic relatedness score (4.73) and yielded the lowest accuracy rate (0.23) when evaluating the correctness of semantic judgments made by all participants for each critical word pair. Consequently, in Experiments 1 and 2, all word pairs associated with the target word "Fuchs" were excluded from further analysis. Also, one trial with responses faster than 100 ms was excluded from the response time analysis. Table 5 summarizes the results from Experiment 1. It shows that participants displayed the slowest yet most accurate responses when the prime and target words were both phonologically and orthographically unrelated. On the other hand, they demonstrated the fastest responses, albeit with lower accuracy, when the prime and target words were solely phonologically related.

Table 5. Mean reaction times (RTs in s) and accuracy rate in Experiment 1

	Condition	**RT**	**Accuracy**
Mean	O+P+	1.074	0.910
	O-P+	1.070	0.917
	O-P-	1.166	0.950

SD	O+P+	0.358	0.287
	O-P+	0.375	0.276
	O-P-	0.397	0.219

3.2.2.1 Response times analysis

The response times were analyzed using a linear mixed model with participants and items as crossed-random factors (e.g., Baayen, 2008), and condition (O+P+, O-P+, O-P-) as the fixed-effect predictor. The O-P+ condition was set as the referene level in the model, so that the phonological priming effect on words could be assessed by contrasting the O-P+ with the O-P- condition, then the orthographic priming effect could be examined by contrasting the O+P+ with the O-P+ condition. For model selection, Barr et al. (2013) pointed out the importance of varying slopes in linear mixed-effects models (LMEMs) and indicated that random-intercept-only LMEMs inflate Type I error rates when experimental designs include within-subjects or within-items manipulations. Barr and colleagues (2013) didn’t support *data-driven* random effects structures (i.e., model selection based on comparison of AIC/BIC value and significance-test), due to the fact that there is little or no power advantage of this approach over maximal models. Following this approach, some researchers postulate to fit ‘maximal’ models. However, Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017) criticized the ‘keep it maximal’ credo. They demonstrated that maximal models may have lower statistical power. In particular, for small sample sizes, a model with a parsimonious random effect structure is most suitable for interpreting factorial experiment results. The simulation in this paper also showed that the parsimonious model has the best chance to detect a true fixed effect as significant. So, there is no strict rule that can be followed in all psycholinguistic studies when it comes to statistical modeling. I chose to use hypothesis-driven models (i.e., model selection based on the hypothesis of the study) for the general response latency and accuracy analyses, meanwhile keeping statistical power in mind.

Likelihood ratio tests were employed to evaluate the models (function *anova* in R) and the lme4 package (Bates et al., 2015) was used in R Studio to perform the model analysis. The effect size estimates for mixed model predictors were computed using the *MuMIn* package in R (Bartón, 2022), which calculates marginal and conditional R-squared values for mixed models based on the approach proposed by Nakagawa and Schielzeth (2013).

Adding the fixed-factor *condition* improved the model fit significantly ($\chi2(2) = 92.919$, $p < 0.001$, $R^2_{marginal} = 0.015$), so we keep *condition* as the fixed-effect predictor in the model. The best model showed a significant effect for the phonological priming condition (*Estimate* = 0.099, *SE* = 0.012, $t = 8.433$, $p < 0.001$); however, the orthographic prime condition did not result in a significant facilitation of the response time (*Estimate* = 0.0004, *SE* = 0.012, $t = 0.030$, $p = 0.976$). The results of the final model are presented in Table 6.

Table 6. The model with the best fit for RTs data in Experiment 1. The reference level for factors was as follows: Condition – O-P+.

German Group—RT: Mixed-Effect-Model with random intercepts and slopes for subject/for item.

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.068	0.261
Item	(Intercept)	0.008	0.091
Residual		0.069	0.262

Fixed effects	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	1.078	0.045	24.124	
O-P- : O-P+	0.099	0.012	8.433	< 0.001
O+P+ : O-P+	0.0004	0.012	0.030	0.976

For post-hoc analysis, a pairwise comparison using *emmeans* package in R was performed between the O+P+ and the O-P- conditions, which demonstrated a significant difference in mean reaction times ($t = -8.376$, $p < 0.001$). To control for family-wise error rates, Bonferroni correction was applied.

3.2.2.2 Accuracy analysis

Analyses of accuracy were conducted with a model of the binomial family due to the binary nature of the data (Jaeger, 2008), which was coded as 1 (correct) or 0 (incorrect) in the experiment. I fitted a mixed logit model with subjects and items as a crossed-random effect, and condition (O+P+, O-P+, O-P-) as a fixed effect (e.g., Baayen, 2008). Similar to the response time analysis, the phonological priming effect on words was assessed by contrasting the O-P+ with the O-P- condition, and the orthographic priming effect by contrasting the O+P+ with the O-P+ condition. The initial model reached a better fit by the inclusion of the fixed-factor condition ($\chi2(2) = 18.751$, $p < 0.001$, $R^2_{marginal} = 007$). Thus, the optimal model incorporated random effects for both subjects and items, while considering the condition as a fixed-effect predictor. Table 7 provides the summary of the model, which shows a significantly reduced accuracy rate for targets that were preceded by phonological primes (*Estimate* = 0.649, *SE* = 0.185, $z = 3.505$, $p < 0.001$) compared with O-P- condition, whereas targets preceded by orthographic primes were not significantly influenced in accuracy rates (*Estimate* = -0.115, *SE* = 0.160, $z = -0.719$, $p = 0.472$) compared with only phonological overlaps.

Table 7. The model with the best fit for accuracy data in Experiment 1. The reference level for factors was as follows: Condition – O-P+.

German Group—Accuracy: Mixed-Effect-Model with random intercepts for subject and item.

Random effects	Name	Variance	SD
Subject	(Intercept)	0.929	0.964
Item	(Intercept)	0.495	0.704

Fixed effects	Estimate	Std. Error	z-value	p-value
(Intercept)	2.971	0.220	13.536	
O-P- : O-P+	0.649	0.185	3.505	< 0.001
O+P+ : O-P+	-0.115	0.160	-0.719	0.472

Additionally, a pairwise comparison using *emmeans* package in R was performed as post-hoc analysis and we found a significant difference in overall accuracy rates between the O+P+ and the O-P- conditions (z = -4.184, p = 0.001). Bonferroni correction was applied.

3.2.3 Discussion

Experiment 1 involved the participation of German native speakers. In this experiment, the targets were preceded by types of primes: those with phonological overlaps but different written forms, those with both phonological and orthographic similarities, and those without any orthographic or phonological relations (O-P+ vs. O+P+ vs. O-P-). We observed that in the German native group, participants exhibited significantly faster and less accurate responses when the prime and target words were only phonologically related than unrelated, but primes in O+P+ condition did not enhance the recognition of the spoken target word compared with in O-P+ condition. This finding supported the important role of phonology during German spoken word recognition with the absence of orthographic effects as shown in Türk & Domahs (2022). They explored the influence of orthography in spoken word recognition among German native speakers, using a lexical decision task. Part of the critical material of my study was taken from

Türk and Domahs (2022). Nonetheless, both their study and mine did not reveal a noteworthy impact of orthography.

The result that the German natives showed significantly faster reactions with phonologically related pairs is in line with the TRACE model (McClelland & Elman, 1986). This model allows for both bottom-up processing, such as from features to phonemes and from phonemes to words, and top-down processing, such as from words to phonemes; phonemes to activate words, and then the activated words activate their phonemes so that the spoken priming word could preactive sublexical representations which facilitate the activation of the following target when the prime and target share the similar rhyme.

The participants in this study exhibited shorter response times when making semantic judgments for the word pairs with phonological overlaps. This can be attributed to the facilitative effect of top-down processing, which enhances the efficiency of perceiving spoken words when the acoustic features of the prime and target words shared similar rhymes. However, contrary to expectations based on the bimodal interactive activation model (Grainger & Ferrand, 1994, 1996; Grainger et al., 2003), the consistency in phonology and orthography between primes and targets did not significantly enhance the activation of matched orthographic representations compared to the phonologically related pairs.

It is important to note that our findings regarding the absence of an orthographic effect are inconsistent with those reported by Perre and colleagues (2009), who employed lexical decision tasks and observed a facilitatory orthographic priming effect. However, it should be acknowledged that their study was performed with English native speakers and used English stimuli, whose grapheme-to-phoneme mappings are considerably more inconsistent compared to German, which has a shallower orthographic system. Also, in studies on other languages such as French (Pattamadilok et al., 2009) and Portuguese (Ventura et al., 2004), the orthographic inconsistency affected spoken word recognition, which didn't occur in my study with German. But in the study by Türk and Domahs (2022), which focused on the German language, the orthographic influence was not evident. Therefore, we suggest the absence of the orthographic effect to be a language-dependent result, as explained by the psycholinguistic grain size theory (Muneaux & Ziegler, 2004; Ziegler & Goswami, 2005). Due to the high orthographic consistency in German, native speakers tend to focus primarily on the smaller psycholinguistic grain size of the phoneme, while English speakers cannot rely as heavily on smaller grain sizes due to the inconsistency in their orthography, so they more depend on higher levels of orthographic consistency to recode the phonological units in smaller grain sizes. Consequently, larger grain sizes such as rimes in English may exhibit significantly higher activation compared to smaller grapheme units in German. Moreover, within shallow orthographies, the number of homophonic heterographs is significantly reduced. If the majority of homophones are also homographs, the application of bimodal processing may not be helpful and could potentially result in increased processing costs because it might activate representational units that do not contribute additional information beyond what is already activated by the phonological units through articulatory features (Türk & Domahs, 2022). As a result, German native speakers can activate phonological units at the whole word level very fast through the acoustic features of spoken words, and they don’t require additional information to distinguish one word from another. Subsequently, semantic information will be activated very quickly.

It is also possible that orthography is not coactivated in online processing, but orthographic similarities could have influenced the nature of the phonological representations during the process of reading acquisition, according to the restructuring hypothesis (Muneaux & Ziegler, 2004; Perre, Pattamadilok, et al., 2009; Taft, 2006). This hypothesis is built on the assumption that separate representations for orthographic and phonological codes may not exist; rather, phonological representations experience some changes throughout language development, so during the learning process, the consistency between spelling and sound at the sublexical level could serve as a mechanism for changing phonological representations at the lexical level, leading to that orthography influences the abstract phonological representations (Muneaux & Ziegler, 2004; Taft, 2006; Taft & Hambly, 1985; Ziegler & Goswami, 2005; Veivo & Järvikivi, 2013). As a result, word pairs with orthographic overlap are considered to be more “similar” due to their shared rhymes. But if word pairs with orthographic overlap became more “similar” in terms of similar rhymes, we should expect obviously faster response times compared to those with only phonological overlap. But the shared orthographic units in Experiment 1 didn’t create facilitation as observed in previous studies. Thus, the restructuring hypothesis does not appear to be an appropriate explanation for the results obtained in Experiment 1.

Moreover, whether the orthographic effect occurs may depend on the type of tasks used in experiments. Grainger and Ferrand (1996) reported significant orthographic and phonological priming effects in both the lexical judgment and perceptual identification tasks, while no such effects were observed in the word naming task. Their explanation was based on the observation that the word naming task, within the masked priming paradigm, is highly sensitive to shared onsets between primes and targets, leading to facilitation only when such shared onsets exist; however, in the lexical judgment task, priming effects were evident regardless of the presence of similar onsets between the prime and target words. Hence, we believe that the orthographic effect may be contingent upon the specific task being performed.

Let’s move on to investigate the role of phonological and orthographic information during spoken word recognition among L2 groups. In the next section, we sought to assess the influence of language proficiency levels on the recognition process and determine how it differed from the L1 group.

3.3 Experiment 2: Experiment with Chinese participants as L2 group

In Experiment 1, the role of orthography and phonology during spoken word recognition was examined in German native speakers with a semantic judgment task. So far, the impact of these effects on non-native word processing has not been investigated in other L2 groups, particularly when their L1 orthography is highly opaque. For this reason, Experiment 2 was designed to investigate orthographic and phonological effects among Chinese native speakers. In Experiment 2, participants were required to complete not only the German semantic judgment task but also a word-learning task and a questionnaire to assess their language experience and proficiency. In contrast to the findings from Experiment 1, we anticipate potential differences in the phonological and orthographic influences on spoken word processing between the L1 and L2 groups.

Moreover, it is investigated whether different orthographic representations influence spoken word recognition in two L2 subgroups with different proficiency. If the size of the orthographic effect is strongly dependent on their L2 proficiency according to previous studies (Veivo et al., 2016; Veivo & Järvikivi, 2013), we should expect a stronger connection between phonological and orthographical representations in the mental lexicon among the subgroup with higher proficiency. In addition to that, when comparing to German native speakers, we hypothesized that the L2 group exhibits worse performance in processing phonological similarities during spoken word recognition, as in the L1 group, top-down processing carries a greater weight compared to the L2 group.

3.3.1 Method

Participants

The L2 group consists of sixty-eight native Chinese speakers, including 48 females and 20 males. The participants had an average age of 26.79 years, falling within a range of 21 to 33 years. Among the participants, 47 were registered as students at the University of Marburg, while 21 were employed by various companies in Germany. Thirty-three of them are considered to have intermediate language proficiency (B1-B2) and thirty-five demonstrate a high level of proficiency in the German language (C1-C2). The classification of language levels, namely B1, B2, C1, and C2, adheres to the language proficiency framework established by the Council of Europe (2018). In the context of the Common European Framework of Reference (CEFR) for language proficiency, B1 and B2 correspond to the third and fourth level, respectively, indicating that L2 learners are able to use German independently; C1 and C2 denote the fifth and sixth level, respectively, expressing that L2 learners possess a high degree of proficiency in German. Participants in this study with intermediate language proficiency were assigned to Group A, while those with high proficiency were assigned to Group B. Neither participant had a documented history of hearing or neurological disorders.

The language proficiency was assessed with a questionnaire, which encompassed several factors: age of acquisition for German, duration of L2 learning and residence in a German-speaking country, number of languages they have learned, and frequency of usage across four domains (speaking, listening, reading, and writing). Moreover, the questionnaire collected information about the specific language examinations undertaken by the participants and their corresponding results, and the participants' self-evaluation about their competence in speaking, listening, reading, and writing as well as an overall assessment of their language proficiency level. Details about the questionnaire design are described in the section *Language experience questionnaire*. According to the questionnaire, all participants were classified as late bilinguals with Chinese as their L1. They had begun learning German at the mean age of 19 and had been residing in Germany for an average duration of 4 years and 10 months. See Table 8 for further details.

Materials and design

The experiment was conducted using the same software and equipment as in Experiment 1, the semantic relatedness task also employed the same set of experimental stimuli. However, in contrast to the German native speakers in Experiment 1, the L2 group was required to complete additional tasks to assess their L2 proficiency level and comprehension of the German words presented during the experiment. To this end, besides the semantic relatedness task and the language experience questionnaire, the test LexTALE (Lemhöfer & Broersma 2012) was selected, and a word-learning task was designed and employed to comprehensively investigate the aforementioned factors.

1) Language Experience Questionnaire

There are two existing questionnaire tools widely used in research to assess the language proficiency of bilinguals – the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) and the Language History Questionnaire (LHQ 2.0; Li, Sepanski, & Zhao, 2006; Li, Zhang, Tsai, & Puls, 2014). The LEAP-Q is designed to collect language background information on bilingual and multilingual speakers, including questions about the age of language acquisition, the participants' language proficiency, language dominance in different situations, and current/past exposure to their languages across settings, etc. The Language History Questionnaire (LHQ) is a web-based questionnaire and includes questions about the participants' age of second language acquisition, length of second language learning, self-rated first- and second-language proficiency, language usage in the home environment, etc. Nevertheless, it should be noted that both questionnaires used in the study were not entirely suitable for our Chinese participants, as they primarily targeted immigrants or individuals with multilingual backgrounds. Therefore, I choose to select relevant questions from these questionnaires, then modified them into a newly developed, reliable questionnaire specifically for Chinese participant group.

The final questionnaire consists of three parts. The first part is about the social background of participants, which gathers demographic information such as age, gender, education, and job. The second part focuses on the language background of participants, including the questions about which language is their first language, which dialects they are able to speak, whether they grew up in a monolingual environment, length of German learning, where they have learned German, at what age, and how long they have resided in Germany. Moving on, the third part delves into language usage and proficiency among participants, including questions that assess the frequency of German language usage in listening, speaking, reading, and writing domains; the response options range from "Never" (0) to "Always" (6), with each choice accompanied by an associated percentage annotation (1: 0%, 2: 0%-20%, 3: 20%-50%, 4: 50%-70%, 5:70%-90%, 6: >90%). Furthermore, participants were requested to provide self-rated proficiency levels in listening, speaking, reading, and writing

German, as well as an overall evaluation of their language proficiency. These self-ratings were categorized into six levels (A1, A2, B1, B2, C1, C2) following the established proficiency scale, where A1 denotes beginners and C2 represents highly proficient speakers. Additionally, participants were asked to list any German language exams they had taken, along with their corresponding scores and dates of completion. Besides, participants were also asked to indicate any other language(s) they have learned, specifying their proficiency level in each language. The completion of the questionnaire was followed by the aggregation of scores. For example, the duration of German learning and residence in Germany were recorded in years, and the number of years served as a score reflecting their language experience. Similarly, the approximate language usage was assessed on a scale ranging from 0 to 6, and self-rated proficiency in listening, speaking, reading, and writing German was rated on a scale of 1 to 6, representing the six different language levels (A1, A2, B1, B2, C1, C2). These aggregated scores were used as predictors in estimating participants' language proficiency.

2) LexTALE

Except for the questionnaire, I also used the German version of LexTALE (Lexical Test for Advanced Learners of English) (Lemhöfer & Broersma, 2012) as a vocabulary knowledge test for intermediate to highly proficient speakers of German as a second language. The LexTALE test is user-friendly and can be completed in approximately 5 minutes. It comprises 60 trials, wherein participants are presented with uppercase letter strings and required to determine whether each string constitutes a valid German word. The final score will be shown at the end of the test. The validity of LexTALE was tested in a study on Dutch and Korean advanced learners of English, and the results suggested that LexTALE scores gave a good indication of general English proficiency and also correlated substantially with experimental word recognition data (Lemhöfer & Broersma, 2012). While the German version of the test has not been validated yet, a correlation analysis was used to examine the relationship between self-rated proficiency levels and LexTALE scores among the Chinese participants. The results revealed a significant correlation (*Pearson's* $r = 0.700$, $p < 0.001$), indicating a strong correlation between self-reported language proficiency ratings and LexTALE scores. In addition, the aggregated score from the Language Experience Questionnaire was combined with the LexTALE score, enabling correlation analyses to be conducted between self-rated proficiency and the final score. The findings revealed a significant and positive correlation between participants' self-ratings of proficiency and the overall score (*Pearson's* $r = 0.823$, $p < 0.001$). Only two participants displayed a discrepancy between their self-reported ratings of language proficiency and the aggregated score, but they provided recent German language exam results, which can prove the validity of their self-ratings. Therefore, based on these considerations, it is deemed appropriate to keep the self-reported ratings of all participants in the study.

See Table 8 for the language experience and proficiency of Groups A and B.

Table 8. Summary of the information of the participants in Experiment 2

Participant background factor	**Group A**		**Group B**	
	Mean	Range	Mean	Range
Age	26.70	21–33	26.89	21–32
Age of acquisition for German	19.45	16–24	18.54	16–25
Length of residence in Germany (years)	4.85	1–10	4.76	1–12
Length of German learning (years)	2.95	1–5	3.63	0.5–6
Frequency of German using in Listening*				
Speaking*	4.33	2–6	4.54	2–6
Reading*	3.48	2–6	4.23	2–6
Writing*	4.36	2–6	4.46	3–6

	3.45	1–5	3.86	2–6
Number of languages learned	1.27	1–3	1.77	1–3
L2-proficiency overall**	3.91	3–4	5.17	5–6
Listening**	4.03	3–5	5.20	4–6
Speaking**	3.91	2–5	4.97	4–6
Reading**	4.36	3–5	5.20	3–6
Writing**	3.70	2–5	4.86	4–6
LexTALE score	53.87	42.5–62.5	66.05	50–88.75

* Participants self-evaluated a rough frequency of German using on a 7-point scale ranging from 0 (“Never”) to 6 (“Always”).

** Participants self-evaluated L2 language proficiency on a 6-point scale ranging from 1 (A1) to 6 (C2).

3) Word learning task

Given the reason that a substantial number of low-frequency German words were used in the experiment and the German language proficiency varied among these Chinese participants, a vocabulary list of the corresponding experimental version will be provided to the participants. The vocabulary list was generated as an online link through SoSci Survey (Leiner, 2019) and was given to the participants a week before the experiment. This list was accompanied by Chinese translations and was presented audibly. And participants were asked to read over the list and remember the meaning of these German words. By employing this approach, it can be ensured that participants possess a clear understanding of the words used in the experiment. However, it was emphasized that they were not permitted to access the list on the day of the experiment. Upon the completion of all experimental procedures, participants were required to undertake a lexical test to examine their comprehension of the German words used in the study.

In the lexical test, the list of stimuli consisted of German words that had been previously presented in an online vocabulary list and presented auditorily in randomized order. During the test, participants heard a spoken word, followed by two Chinese translation options that appeared in the center of the screen. Participants were allowed to determine the correct translation from the given options for 2000 ms.

Example:

- Please match the meaning of the word you just heard.
  (Play the sound: Feier *celebration*)
- A. 庆祝 (*celebration*) B. 火车 (*train*)

Following the incorrect response, the right answer will be marked in green. At the end of the entire test, feedback regarding the accuracy rate was provided. The results showed that all participants achieved at least an accuracy rate of 80 percent.

Procedure and Apparatus

The entire experiment lasted approximately 40 to 50 minutes. Participants were instructed to finish the language experience questionnaire firstly, followed by the semantic relatedness judgment task, LexTALE, and the lexical test. The procedure of the semantic relatedness judgment task was the same as in Experiment 1. Note that the learning test of words was completed at home and was not included in the total time of the experiment.

3.3.2 Results

Consistent with the data analysis approach employed in Experiment 1, any incorrect responses (less than 27% of the total) were excluded from further analyses of response time. Similarly, response times shorter than the first quartile minus 3 times the IQR and response times exceeding the third quartile plus 3 times the IQR of correct responses were considered outliers and discarded (less than 0.1% of the dataset) (Tukey, 1977). The dataset includes two distinct proficiency-level groups: Group A, representing participants with intermediate proficiency, and Group B, comprising participants with high proficiency. Subgroup data for all three conditions were calculated separately, allowing for a comprehensive analysis of each group's performance within the respective conditions.

Table 9 summarizes the results from Experiment 2, and Figure 13 visualized the response data of the two subgroups. As can be seen from the table and diagrams above, the unrelated condition (O-P-) revealed a longer response time than the other two conditions, and the orthographically-phonologically similar condition (O+P+) yielded the fastest response times. This pattern was consistent across both proficiency groups. Additionally, the accuracy rate results showed the same pattern of response time, wherein the O-P- condition exhibited lower accuracy compared to the other two conditions, and the O+P+ condition displayed the best accuracy, regardless of the participants' proficiency level.

Table 9 Mean reaction times (RT) and accuracy rates in Experiment 2.

			RT (s)		**Accuracy**	
			Mean	SD	Mean	SD
All	Condition	O+P+	1.747	0.670	0.843	0.364
		O-P+	1.788	0.644	0.824	0.381
		O-P-	1.922	0.646	0.749	0.434
			Mean	SD	Mean	SD
Group A	Condition	O+P+	1.883	0.687	0.809	0.393
		O-P+	1.913	0.671	0.801	0.399
		O-P-	2.037	0.656	0.698	0.460
			Mean	SD	Mean	SD
Group B		O+P+	1.628	0.630	0.875	0.331

Condition	O-P+	1.675	0.597	0.845	0.362
	O-P-	1.829	0.623	0.797	0.403

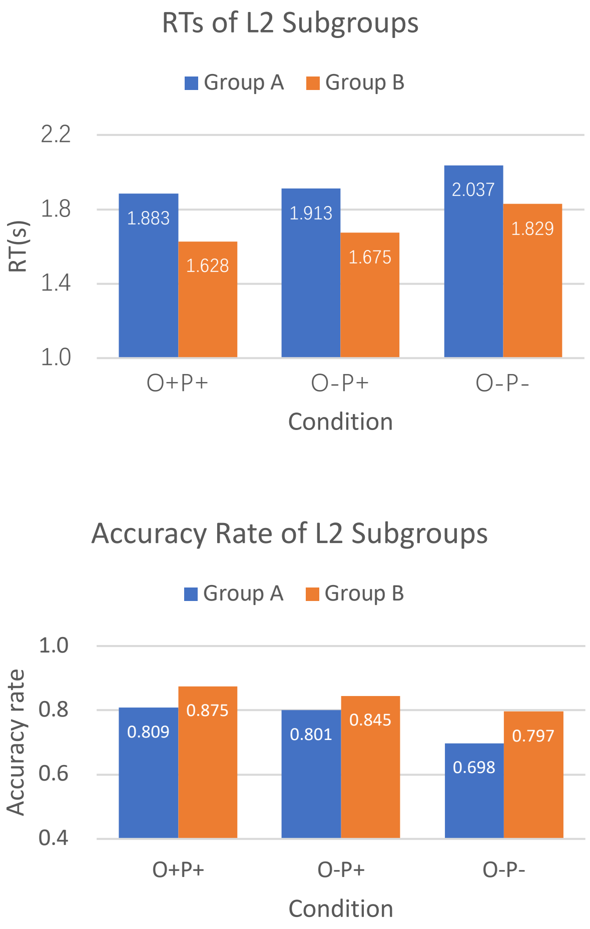


Figure 13. Statistical diagrams of RTs and accuracy of the L2 subgroups

Again, analyses of the RTs for correct responses and accuracy rates were conducted using generalized linear mixed-effects models (Baayen, 2008) in R with the lme4 package (Bates et al., 2015). The two contrasts of condition used in Experiment 1 were replicated in this study as well. The first contrast examined the phonological priming effect on target words, achieved by contrasting the O-P+ condition with the O-P- condition. Subsequently, the second contrast explored the orthographic priming effect by comparing the O+P+ condition against the O-P+ condition. In both contrasts, the O-P+ condition was designated as the baseline against which the other conditions were compared. For model selection, I also used likelihood ratio tests to evaluate the models in R with function *anova*. Effect size estimates for predictors in mixed models were computed using the *MuMIn* package (Bartón, 2022). This package calculates marginal and conditional R-squared values for mixed models, following the methodology of Nakagawa and Schielzeth (2013). Followed by the analysis of response time and accuracy rate in Experiment 2, the contrasting outcomes between the L1 group and L2 group, along with their respective subgroups, will also be delineated.

3.3.2.1 Response time analysis

Considering the diversity of language proficiency levels among participants and the varying familiarity of the items used in the experiment, there is considerable variability in participants' lexical comprehension, resulting in a faster or slower response. It is plausible that these factors have an impact on the observed priming effects. As a result, in this study, both participants and items were considered as random effects and included in the linear mixed-effects model for data analysis. Considering the limited sample size, it is advisable to employ a parsimonious model featuring solely random intercepts, as it offers greater statistical power.

Therefore, a model was constructed to incorporate random intercepts for both participants and targets in the data analysis. Likelihood-ratio test accordingly showed a significant effect when adding *condition* as a fixed-effect predictor ($\chi 2(2) = 114.73$, $p < 0.001$, $R^2_{\text{marginal}} = 0.014$). Adding *proficiency* as fixed-effect predictor also increased the model fit significantly ($\chi 2(1) = 6.5354$, $p = 0{,}01$, $R^2_{\text{marginal}} = 0.048$), so we keep *proficiency* as the fixed-effect factor in our model. Next, we added an interaction between condition and proficiency to the model. However, the addition of this interaction did not yield a substantial improvement in model fit ($\chi 2(2) = 1.63$, $p = 0.44$, $R^2_{\text{marginal}} = 0.048$), indicating that different priming conditions didn't elicit significantly faster or slower responses in Group A as compared to Group B. This lack of significant interaction might result from either a true absence of an effect or insufficient statistical power to detect such an interaction (Veivo & Järvikivi, 2013). Thus, the interaction term was dropped from the model.

Table 10 displays the information of the final model. The results revealed significant differences between O-P- and O-P+ (*Estimate* = 0.150, *SE* = 0.018, $t = 8.385$, $p < 0.001$) and for proficiency (*Estimate* = -0.250, *SE* = 0.095, $t = -2.619$, $p = 0.011$), and showed a trend between O+P+ and O-P+ (*Estimate* = -0.031, *SE* = 0.017, $t = -1.788$, $p = 0.074$).

Table 10. The model with the best fit for RTs in Experiment 1. The reference levels for factors were as follows: Condition – O-P+, Proficiency – Level A

Chinese Group —RT: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.151	0.388
Item	(Intercept)	0.018	0.135
Residual		0.252	0.502

Fixed effects	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	2.194	0.153	14.330	
O-P- : O-P+	0.150	0.018	8.385	< 0.001
O+P+: O-P+	-0.031	0.017	-1.788	0.074
proficiency	-0.250	0.095	-2.619	0.011

After that, pairwise comparisons on RTs were run with *emmeans* package in R (Lenth et al., 2021). The results revealed a significant effect of proficiency on the overall mean RTs ($t = 12.697$, $p < 0.001$) as well as on the RTs of the three priming conditions respectively (Condition – O-P+: $t = 7.560$, $p < 0.001$; Condition – O-P-: $t = 6.303$, $p =$

< 0.001; Condition – O+P+: *t* = 8.199, *p* < 0.001), indicating that participants in the high proficiency group responded significantly faster to the stimuli compared to those in the intermediate proficiency group.

Then separate analyses for the two subgroups with different language proficiency levels were conducted. Because the number of observations was reduced in subgroup analysis, the statistical power decreased. Therefore, I chose to construct the parsimonious models, which included only by-participant and by-target random intercepts for priming conditions, and with *condition* as a fixed-effect predictor.

The resulting models are depicted in Table 11 and Table 12. The results of Group A (the intermediate proficiency group) revealed that O+P+ did not have any significant influence on response time compared with O-P+ (*Estimate* = -0.022, *SE* = 0.026, *t* = -0.836, *p* = 0.403) whereas O-P+ showed a significant difference compared with O-P- (*Estimate* = -0.131, *SE* = 0.027, *t* = -4.776, *p* < 0.001). On the other hand, Group B (the high proficiency group) exhibited significant facilitation in response time by contrasting O+P+ and O-P+ (*Estimate* = -0.049, *SE* = 0.023, *t* = -2.104, *p* = 0.035) and O-P+ and O-P- (*Estimate* = -0.161, *SE* = 0.024, *t* = -6.762, *p* < 0.001).

Table 11. Results from the RT analyses for Group A.

Group A — RT: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.168	0.410
Item	(Intercept)	0.014	0.118
Residual		0.273	0.523

Fixed effects	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	1.944	0.075	25.993	
O-P- : O-P+	0.131	0.027	4.776	< 0.001
O+P+: O-P+	-0.022	0.026	-0.836	0.403

Table 12. Results from the RT analyses for Group B.

Group B — RT: Mixed-Effect-Model with random intercepts for subject and item			
Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.133	0.365
Item	(Intercept)	0.022	0.150
Residual		0.232	0.482

Fixed effects	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	1.697	0.066	25.787	
O-P- : O-P+	0.161	0.024	6.762	< 0.001
O+P+: O-P+	-0.049	0.023	-2.104	0.035

Besides, in the post-hoc analysis using *emmeans* package in R, a pairwise comparison was conducted between the O+P+ and the O-P- conditions, revealing a significant difference in mean reaction times (t = -10.177, p < 0.001), which is the same as when examining the subgroup with intermediate proficiency (t = -5.590, p < 0.001) as well as the subgroup with high proficiency (t = -8.904, p < 0.001). To control for family-wise error rates, a Bonferroni correction was applied.

3.3.2.2 Accuracy analysis

Similar to the accuracy analyses conducted in Experiment 1, a mixed-effects regression model of the binomial family was fitted to the accuracy data in R. The accuracy data were coded as 1 for correct responses and 0 for incorrect responses during the experiment. Subjects and items were conducted as crossed-random effects, and condition (O+P+, O-P+, O-P-) and proficiency level (Group A and Group B) as fixed effects (e.g., Baayen, 2008). Again, the exploration of the phonological priming effect was conducted by contrasting the O-P+ condition against the O-P- condition, while the orthographic priming effect was examined by contrasting the O+P+ condition against the O-P+ condition. The fit of the model was getting better fitted with the inclusion of the fixed-factor *condition* ($\chi2(2) = 72.02$, $p < 0.001$, $R^2_{marginal} = 0.011$). Additionally, adding the fixed-factor *proficiency* slightly enhanced the model fit, as indicated by the likelihood-ratio test ($\chi2(1) = 5.999$, $p = 0.014$, $R^2_{marginal} = 0.023$), but the interaction between *condition* and *proficiency* did not increase the fit ($\chi2(2) = 2.277$, $p = 0.32$, $R^2_{marginal} = 0.023$). It thus occurred that the proficiency did not modulate the accuracy rate in phonological or orthographic overlap conditions. Hence, a model without the interaction term was selected.

Table 13 presents the resulting model with the best fit of the data. Overall, the phonological overlap condition significantly reduced the accuracy rate (*Estimate* = -0.524, *SE* = 0.083, *z* = -6.335, *p* < 0.001), and the orthographic overlap condition increased the accuracy rate slightly but did not reach the significance level (*Estimate* = 0.152, *SE* = 0.242, *z* = 2.493, *p* = 0.087).

Table 13. The model with the best fit for accuracy data in Experiment 1. The reference levels for factors were as follows: Condition – O-P+, Proficiency – Level A

Group AB—Accuracy: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.888	0.942
Item	(Intercept)	0.152	0.390

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	0.951	0.389	2.445	
O-P- : O-P+	-0.524	0.083	-6.335	< 0.001
O+P+ : O-P+	0.152	0.090	1.710	0.087
proficiency	0.604	0.242	2.493	0.013

We next moved on to analyze the accuracy in greater detail within the two subgroups. Response accuracy in the subgroups was further examined by fitting a mixed-effects logistic regression model in R. The model included crossed-random effects for subjects and items, with the O-P+ condition serving as the baseline. Different patterns of accuracy results in subgroups were revealed: in Group A, orthographic overlap didn't modulate significantly the facilitation on accuracy rate (*Estimate* = 0.049, *SE* = 0.118, z = 0.414, p = 0.679), but phonological overlap did (*Estimate* = -0.621, *SE* = 0.111, z = -5.583, p < 0.001), whereas Group B benefitted significantly from both the phonological overlaps (*Estimate* = -0.405, *SE* = 0.124, z = -3.258, p = 0.001) and the orthographic similarities (*Estimate* = 0.281, *SE* = 0.135, z = 2.077, p = 0.038). The results of the accuracy analysis for Group A and B are summarized in Table 14 and Table 15 respectively.

Table 14. Results from the accuracy analyses for Group A.

Group A — Accuracy: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.627	0.792
Item	(Intercept)	0.137	0.370

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	1.606	0.168	9.569	
O-P- : O-P+	-0.621	0.111	-5.583	< 0.001
O+P+ : O-P+	0.049	0.118	0.679	0.679

Table 15. Results from the accuracy analyses for Group B.

Group B — Accuracy: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	1.195	1.093
Item	(Intercept)	0.216	0.464

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	2.120	0.217	9.759	
O-P- : O-P+	-0.405	0.124	-3.258	0.001
O+P+ : O-P+	0.281	0.135	2.077	0.038

Through the pairwise comparison result, Group B demonstrated significantly higher accuracy in the semantic judgment task compared to Group A (overall: $z = -6.677$, $p < 0.001$; condition O-P+: $z = -2.565$, $p = 0.010$; condition O+P+: $p < 0.001$; condition O-P-: $z = -5.094$, $p < 0.001$). Moreover, across both groups, the O+P+ condition yielded the highest accuracy rate, while the O-P- condition resulted in the highest error rate.
For post-hoc analysis, a pairwise comparison using *emmeans* package in R was performed between the O+P+ and the O-P- conditions, revealing a significant difference in overall accuracy rates ($z = 7.975$, $p < 0.001$), and a similar result was also found when examining the subgroup with intermediate proficiency ($z = 5.972$, $p < 0.001$) as well as the subgroup with high proficiency ($z = 5.274$, $p < 0.001$). Bonferroni correction was also applied.

3.3.2.3 Differences between L1 group and L2 group

In contrast to the L2 group which revealed the fastest response and the best accuracy in O+P+ condition, the L1 group exhibited the fastest response time in the O-P+ condition and the highest accuracy in the O-P- condition (see Figure 14). Upon conducting pairwise comparisons between the L1 group and L2 group, it was observed that the L1 group exhibited significantly faster mean response times (condition O-P+: *Estimate* = 0.718, *SE* = 0.023, $t = 31.865$, $p = < 0.001$; condition O+P+: *Estimate* = 0.673, *SE* = 0.022, $t = 29.975$, $p < 0.001$; condition O-P-: *Estimate* = 0.757, *SE* = 0.023, $t = 33.302$, $p < 0.001$) and significantly better accuracy (condition O-P+: *Estimate* = -0.864, *SE* = 0.123, $z = -6.996$, $p = < 0.001$; condition O+P+: *Estimate* = -0.629, *SE* = 0.121, $z = -5.190$, $p < 0.001$; condition O-P-: *Estimate* = -1.842, *SE* = 0.146, $z = -12.589$, $p < 0.001$) compared to the L2 group across all conditions.

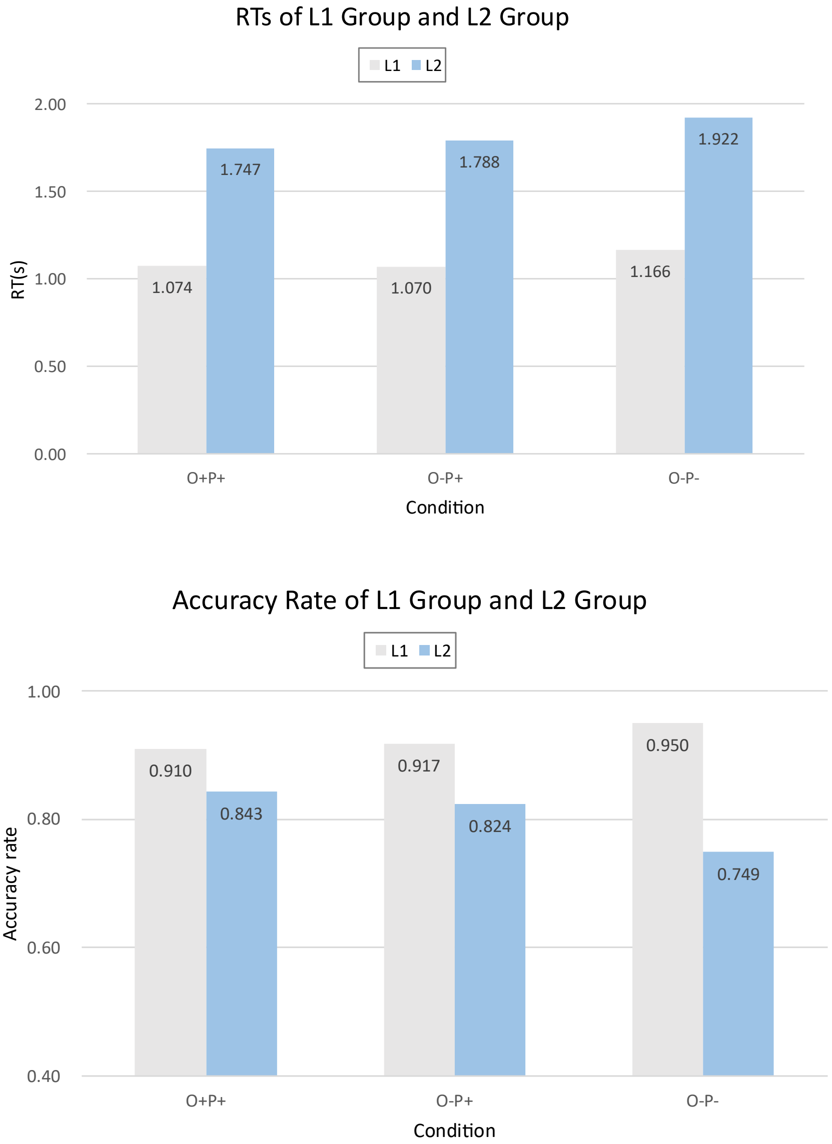


Figure 14. Statistical diagrams of RTs and accuracy in L1 and L2 Groups

Then the pattern of orthographic and phonological effects in L1 and L2 groups was explored. Numerically, the p-value for the orthographic effect in RTs was 0.976 for the L1 group, while it was 0.074 for the L2 group. This implies that the orthographic effect was stronger in the L2 group compared to the L1 group. On the opposite, the phonological effect appears to be quite similar in size in L1 and L2 groups, the p-value of each phonological effect was smaller than 0.001. Moreover, a joint analysis on the RT results from both experiments was performed with a linear mixed model (*lmer* in R), *subject* and *item* were conducted as crossed-random effects, and *condition* and *group* as fixed effects. The model exhibiting the best fit to the data indicated a noteworthy interaction between *condition* and *group* ($\chi2(2) = 11.784$, $p = 0.003$, $R^2_{\text{marginal}} = 0.290$), indicating that the facilitation of different priming conditions on RTs varied between the two groups. The effects of *condition* and *group* were also tested for accuracy rate in L1 and L2 groups with a linear mixed model (*lmer* in R). The interaction of *condition* and *group* was found to be significant ($\chi2(2) = 56.581$, $p < 0.001$, $R^2_{\text{marginal}} = 0.050$), demonstrating that the facilitation of orthographic and phonological effects on accuracy rate was distinct in L1 and L2 groups. These results suggest that the behavior of the L1 group significantly differed from that of the L2 group in terms of orthographic and phonological effects.

Finally, the L1 group was compared with two subgroups of L2 learners. Numerically, we found that RTs for O+P+ condition in L1 group were slightly slowed down compared to O-P+ condition, whereas in Groups A and B the reaction times for O+P+ condition was decreased (see Figure 15). However, the size of the orthographic effect observed in the L1 group resembled that of the Chinese participants with intermediate proficiency (Group A), more than that of the high proficiency group (Group B). Regarding accuracy rates, a significant orthographic priming effect was only observed in Group B, although the effect was weaker compared to the solely phonological priming effect.

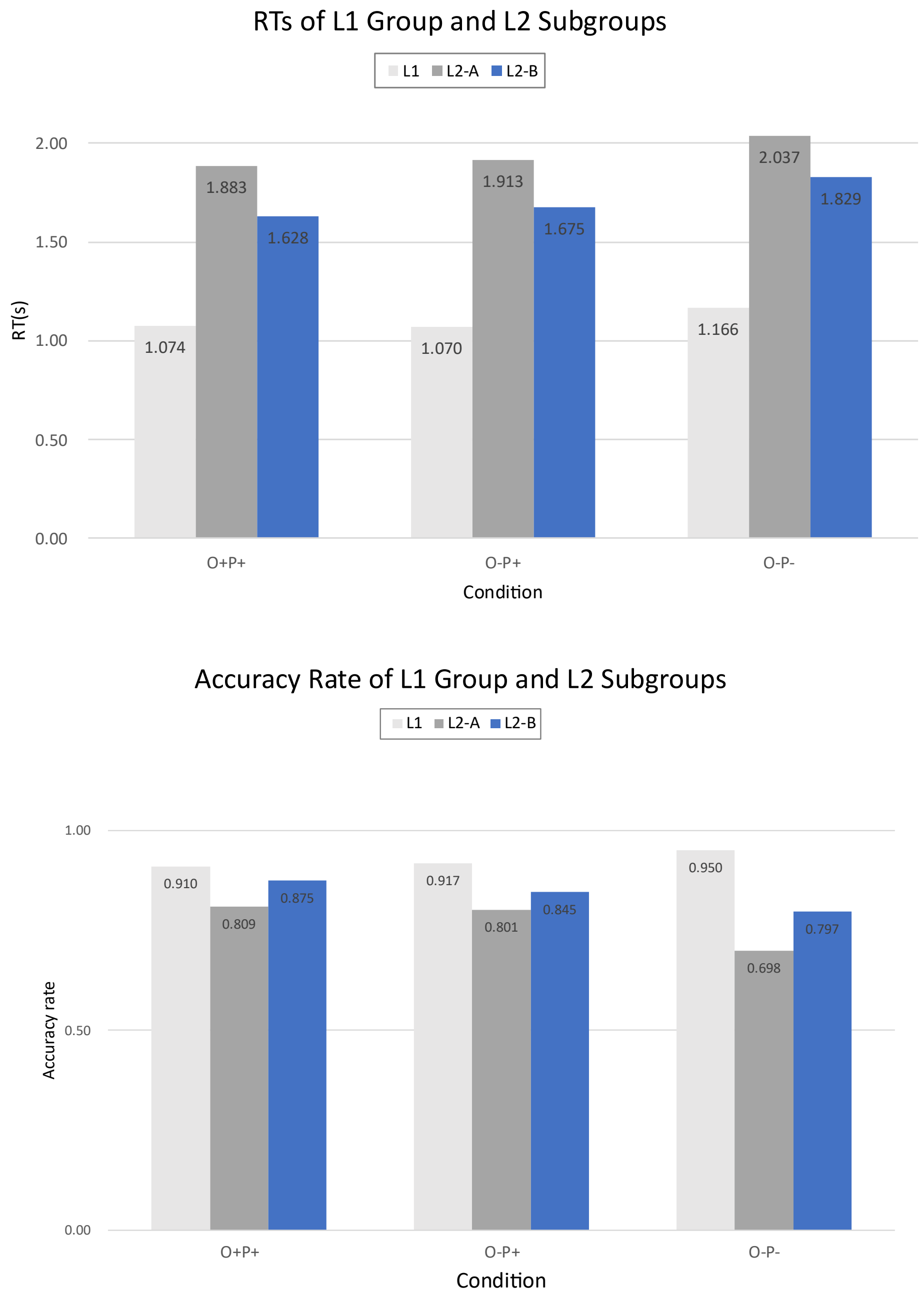


Figure 15. Statistical diagrams of RTs and accuracy in L1 and L2 subgroups

3.3.3 Discussion

Experiment 2 replicated the experimental design of Experiment 1 but focused on an L2 group consisting of participants with two different language proficiency levels:

intermediate and high proficiency. The targets in Experiment 2 were contrasted with three types of primes which shared the same phonological rhyme but differed in their written forms, share both phonological and orthographic rhymes, or were neither orthographically nor phonologically related with the target (O-P+ vs. O+P+ vs. O-P-). Overall, proficiency in the L2 did influence the response time and accuracy rate: the higher the proficiency in the L2 is, the faster and more accurate the judgment was made. Interestingly, Chinese participants exhibited significantly faster response times when presented with word pairs that shared orthographic and phonological similarity, compared to pairs without orthographic overlap. Additionally, the phonological priming effect was highly significant. These findings align with previous research by Perre et al. (2009), which suggests that native English speakers respond more quickly to word pairs with both orthographic and phonological overlap compared to pairs with only phonological similarity.

The results from the L2 group revealed a prominent phonological priming effect, with the similar rhymes affecting the mapping in both Group B and Group A to a similar extent. These findings demonstrate that higher language proficiency did not result in a stronger effect of phonological overlap but influenced how fast and how accurate the response was. The participants with high language proficiency have more stable correspondences between phonemes and graphemes in the mental lexicon, so they can map the spoken word to written form more quickly and accurately.

We can interpret the findings of the phonological priming effect in support of the bimodal interactive activation model (Grainger & Ferrand, 1996; Grainger & Ziegler, 2007; Ziegler, Muneaux, & Grainger, 2003). According to this model, sublexical representations are preactivated by primes and subsequently activated by targets, so targets will be easier reactivated, resulting in the facilitatory rhyme priming.

In terms of another research question about orthographic influence, it can be found that when the orthography of the target was similar to that of the prime, the prime would have a more positive impact on the processing of the target compared to cases where the written forms were different. But we didn’t observe the same significant effect of the orthographic overlap on the target in Group A and B, which suggests that L2 listeners with varying language proficiency levels may not be using the same type of information during spoken word recognition and the orthographic effect could be dependent on their proficiency in L2 to a great extent. The psycholinguistic grain size theory (Goswami, 2010; Ziegler & Goswami, 2005) could provide an explanation for this observed different performance. Chinese L2 speakers with intermediate language proficiency tend to rely more on large psycholinguistic grain size, just like their approach in their native language. As a result, they may experience difficulties in quickly and accurately recognizing smaller grain sizes. It is possible that they subdivide the phonological input into onset and rhyme or perceive the individual phonemes within a spoken word as part of a larger unit. For instance, when Chinese participants heard two German words *Kleid (dress)* and *Streit (quarrel)*, they might subdivide *Kleid* into *Kl* + *eid*, *Streit* into *Str* + *eit*, and consider *eid* in *Kleid* and *eit* in *Streit* as phonologically as well as orthographically similar and ignore the different written form.

The facilitatory orthographic effect observed in this experiment was consistent with previous research investigated in other alphabetical languages (Pattamadilok et al., 2009; Perre et al., 2009; Ventura et al., 2004). During the process of word recognition, spoken and written words have different grain sizes, but they are dynamically and interactively linked, when one word has a very consistent sound-spelling relation, it is likely to be selected more easily compared to a word with inconsistent sound-spelling relation (Ziegler et al., 2003; Ziegler & Muneaux, 2007). Therefore, in the case of Chinese participants, judgments were made faster when there were orthographic overlaps between the primes and targets, indicating the facilitatory nature of the orthographic effect.

With respect to the varying orthographic influences on spoken word recognition observed in participants with high language proficiency and intermediate language proficiency, it is plausible that the characteristics of their native language might play an important role. Chinese participants begin learning German after attaining literacy in their native language. Consequently, the phonological representations of words are already established through the *pinyin* system, which directly represents sounds, and each character represents a monosyllabic unit. The relationship between spelling and sound is – however – not as straightforward as is the case for German spelling, so the weak correspondence between spelling and sound poses a specific challenge in Chinese. As a result, Chinese learners do not need to rely heavily on complex phonological decoding and the written form contrasts between some phonemes is difficult to distinguish for Chinese learners of German. Participants with an intermediate language level, who primarily learned and used these words in written form rather than through auditory means in their daily lives, appeared to encode them as having identical orthographic units. Consequently, their limited experience with L2 resulted in segmental errors during spoken word recognition. Conversely, the group of participants with high proficiency in the L2 demonstrated greater familiarity with the phonological and orthographic representations of these words, exhibiting a more stable phoneme-grapheme correspondence in their mental lexicon. As a result, they were able to discern words in the orthographic priming condition (O+P+) as having not only phonological but also orthographic similarities. These findings indicated that language proficiency in a second language influences the orthographic and phonological representations of spoken words.

A detailed analysis of the results from Experiments 1 and 2 revealed contrasting patterns in the processing of orthographic and phonological information during spoken word retrieval between the L1 and L2 groups. Generally, L1 participants' responses to the targets were significantly more accurate and faster than L2 participants.
It is noteworthy that both the L1 and L2 participants demonstrated a significant facilitation effect of phonological overlap compared to unrelated distractors. However, the impact of orthographic similarity differed between the two groups. In the L2 group, O+P+ accelerated response times and improved accuracy rates, whereas in the L1 group, it resulted in slower response times and decreased accuracy rates compared to O-P+ condition. Türk and Domahs (2022) supported that orthographic overlap counteracts the facilitation effect caused by phonologically identical rimes among German native speakers. And the results from our L2 participants suggested this kind of counteraction

might not exist in L2 learners.

Moreover, different responses were observed in relation to orthographically similar versus dissimilar pairs. In the L1 group, we observed slightly longer response times for orthographically related pairs during the semantic judgment task. In other words, the effect of orthography in Experiment 1 can be considered as a null effect. Conversely, in the L2 group, responses were faster when targets were preceded by orthographically similar primes. These findings imply that orthographic similarity exhibit stronger facilitation in word recognition in the L2 group compared to the L1 group. On one hand, in the L1 context, phonological knowledge is typically acquired at an earlier stage than orthographic knowledge (Qu & Damian, 2017). In contrast, in the L2 context, orthographic forms are learned concurrently with or even prior to the sound of words. This is particularly evident in Chinese learners of German, whose native language does not belong to an alphabetic writing system. In the initial stages of learning German, these learners tend to prioritize the written form and establish connections between phonological features and orthographic information through the Chinese *pinyin* system. As a result, orthographic and phonological codes may be not much interconnected in non-native speakers. On the other hand, native speakers are able to rapidly activate semantic information through the acoustic features of spoken language. This activation follows a bidirectional process, as described by the BIAM model. In contrast, non-native speakers tend to rely more on bottom-up processing. When the auditory input matches their phonological representations, both orthographic and phonological lexical, as well as sublexical representations are activated, after that, the semantic information achieved the activation. However, the experimental method used in the study is the semantic judgment task, which elicits more higher-level activation as participants are instructed to respond based on word meaning. Just in the O+P+ and O-P+ conditions, the presence of orthographic and phonological overlapping facilitates greater bottom-up activation compared to the unrelated condition. And as the BIA+ model reviewed, bottom-up recognition process will usually be much faster (Dijkstra & van Heuven, 2002), which is consistent with the fastest response results in the O+P+ condition and the longest response time in the O-P- condition in L2 group.

There is also another view that because lexical codes are often less integrated and stable in L2 than in L1, orthographic effects in the L2 group may be less prominent (Qu et al., 2018). But the result of my study was the opposite. One possible explanation for this discrepancy is that the stronger orthographic effects in L2 group may not be attributed to online orthographic activation alone. Instead, it is possible that the orthographic similarity between words has influenced the quality of phonological representations when we learned to read, so words with similar spellings may have become more phonologically similar than those with only phonological similarity, leading to the observed orthographic effects (Muneaux & Ziegler, 2004; Taft, 2006; Taft & Hambly, 1985). In this case, it is suggested that facilitatory orthographic priming can occur in auditory speech perception without orthographic presentation (Perre et al., 2009). When two words share similar spellings, such as "Neid" (*jealousy*) and "Kleid" (*dress*), they are likely to have stronger associations with their respective phonological representations compared to words with dissimilar spellings, such as "Streit" (*quarrel*) and "Kleid" (*dress*). As a result, the facilitatory orthographic effect emerges due to the enhanced connections between orthographic and phonological representations in the former case.

Our findings in L2 are consistent with that of Veivo and Järvikivi (2013), which suggest that the impact of orthographic information on L2 spoken word recognition is contingent upon L2 proficiency. The study conducted by Veivo and Järvikivi (2013) specifically focused on Finnish (L1) and French (L2), both of which employ alphabetical writing systems. Other studies from Pytlyk (2017), Türk and Domahs (2022), Hao and Yang (2018) also showed the influence of orthography in L2 word processing. Based on these findings and results of my study, which showed the absence of orthographic influence in native speakers but in L2 groups, we can assume that the influence of orthography on spoken word recognition in L2 might be proficiency-dependent, but not associated with the target language.

To determine whether the stronger orthographic effect observed in the L2 group is attributed to the inherent characteristics of participants' L1 orthography (Chinese) or simply because L2 processing emphasizes orthography to a greater extent, further investigation is necessary to clarify if orthography indeed plays a significant role in phonological processing within the Chinese language. This hypothesis will be subjected to empirical testing in subsequent experiments.

3.4 Experiment 3: Experiment for Chinese participants living in Germany

To examine the impact of orthography on L1 spoken word recognition, particularly in the context of logographic languages such as Chinese, the semantic priming paradigm, similar to Experiment 1, was employed. Given the weak grapheme-phoneme correspondence in Chinese, it allowed for the creation of an exclusively orthographic overlap condition without phonological overlaps between the orthographically related primes and targets that existed in German. Moreover, we wanted to contrast the performances in response to L1 and L2 of Chinese participants. Considering that Chinese, in comparison to German, exhibits inconsistent correspondence between its written and spoken forms, we hypothesized that Chinese participants may display different patterns influenced by orthographic similarities when perceiving Chinese words.

3.4.1 Method

Participants

A group of 42 Chinese participants, drawn from the same pool as Experiment 2, was included in this study. These participants had a mean age of 26.3 and resided in Germany for an average duration of 4 years and 10 months. To prevent potential response biases due to the excessive length and repetition of the experimental process, the order of Experiments 2 and 3 was counterbalanced among the Chinese participants who took part in both experiments.

Materials and design

Several stimuli in this experiment are chosen from Qu & Damian (2017), but the original experiment didn't incorporate phonological overlap in all conditions. Since the aim of the study was to investigate how Chinese listeners map incoming acoustic information onto the meaning of spoken words when primes and target share similar pronunciations, a new stimuli list with phonological overlap was generated.

The critical stimuli employed in this experiment comprise 90 sets of words, with each set comprising one target word and three prime words. All prime-target pairs were chosen to be semantically unrelated, while incorporating orthographic or phonological overlap to establish the following conditions (the overlap is in the position of the first character):

- O+P-: orthographically related, but phonologically unrelated (e.g., 童年, /tong2nian2/, *childhood* – 撞击, /zhuang4ji1/, *crash,* the first characters of these two words share the same radical "童", but are pronounced differently);
- O-P+: phonologically related, but orthographically unrelated (e.g., 童年, /tong2nian2/, *childhood* – 铜牌, /tong2pai2/, *bronze medal,* the first characters of both two words pronounced as /tong2/, but they are written differently);
- O-P-: and unrelated in phonology and orthography (e.g., 童年, /tong2nian2/, *childhood* – 破旧, /po4jiu4/, *old and shabby*).

Based on the previous studies and the features of Chinese characters, some independent variables should be controlled for:

1. Word formation: All of the characters chosen for this study consisted of two-character words, as single-spoken Chinese characters often have numerous homophones, leading to possible ambiguity in meaning. The phonological and orthographic relatedness of the Chinese word pairs referred to matches or mismatches that varied in terms of phonemes, tone, and orthography within the first

character of each word pair, rather than the second character.

2. Character frequency: The frequency of the selected characters was determined by the Chinese Lexical Database (CLD) (Sun et al., 2018). CLD is a lexical database for simplified Chinese, which comprises not only one-character words but also two-character, three-character as well as four-character words. As we used two-character words as stimuli in this experiment, CLD can be a valuable resource. The frequency of the first character and the whole word was calculated. The mean frequency of target words was 25.7 per million, and prime words in three conditions are matched with targets ($F = 1.695$, $p = 0.170$). The mean frequency of the first character in target words was 198 per million, the first characters in primes are also matched with that in target words ($F = 1.850$, $p = 0.139$).
3. The stroke numbers: In consideration of the visual complexity of Chinese words, prime words are matched with targets on the stroke numbers of the first character and the whole word (Mean stroke numbers of targets: 18.2; Mean stroke numbers of primes: 17.2).
4. Phonological and orthographic neighborhood: The phonological and orthographic neighborhood sizes of Chinese characters are very important in word recognition, the influence has been demonstrated in several studies (Chang et al., 2016; Huang et al., 2006; Q. L. Li et al., 2011), so these two factors were included as well. Phonological neighborhood density was also calculated through CLD (Sun et al., 2018). In alphabetic languages, phonological neighborhood density is typically assessed based on shared phonemes. However, in the case of Chinese, phonological forms are determined not only by constituent phonemes but also by tones. Therefore, when calculating neighborhood density in CLD, both phonemic and tonal differences were taken into account (Sun et al., 2018). Specifically, the phonological neighborhood density size was derived from the count of words or characters that exhibited differences of either one phoneme or one tone when compared to the target word or character. However, calculating orthographic neighbors for Chinese is not as straightforward as it is for languages like German, where orthographic neighbors are determined based on shared letters. The unique

nature of the Chinese writing system has been demonstrated in Section 2.3.2, which highlights that most Chinese characters are compound characters, constructed from a combination of a phonetic radical and a semantic radical (Perfetti & Tan, 1998). In line with the findings of Chang et al. (2016), the size of the orthographic neighborhood in Chinese was defined in terms of phonetic combinability or phonetic radical frequency by type, which means that we could consider the number of characters that share the same phonetic radical as the orthographic neighborhood size. But not all the Chinese stimuli used in the experiment have phonetic radicals. Hence, in this research, we couldn't directly treat orthographic neighborhoods as phonetic radical neighborhoods in Chinese. Sun et al. (2018) gave us a solution for the calculation of orthographic neighborhood size, which is called OLD Pixels. OLD represents orthographic Levenshtein distance and denotes mean distance between a character and its *n* closest neighbors, and *n* was set to 20 for the OLD Pixels measures in the CLD (Sun et al., 2018). Character distances were computed using PNG image files, with each pixel being defined as either white or non-white. Subsequently, the distance between the character and all other characters was computed (Sun et al., 2018). It is important to note that in this study, the measurement of orthographic neighborhood size focused solely on the first character, as previous studies showed that during the reading of Chinese two-character words, the neighborhood size of the initial character influenced the early stage of lexical access more strongly than the second character did (Huang et al., 2006).

Table 16 and 17 shows the characteristics of the material used in the experiment. The results of post hoc tests (with Bonferroni-corrected *p*) are shown in Table 18. As can be seen from Table 18, the differences between targets and primes' frequency, orthographical and phonological neighborhood sizes, and stroke numbers were all non-significant.

Table 16. Overall characteristics of the stimulus materials

	frequency	**frequency C1**	**stroke_ Word**	**stroke_ C1**	**phono.NB Word**	**phono.NB C1**	**OLD C1**	**OLD C2**
Mean	36.3	264	17.4	9.35	1.72	14.2	2524	2469
SD	93.5	437	4.48	2.85	1.94	5.44	263	314

*C1 means the first character, C2 means the second character.

**NB means neighborhood size.

Table 17. Characteristics of the stimuli in every condition

	condition	**frequency**	**frequency C1**	**stroke_ Word**	**stroke _C1**	**phono.NB Word**	**phono. NB C1**	**OLD C1**	**OLD C2**
Mean	Target	25.7	198	18.2	9.69	1.78	13.5	2500	2501
	O+P-	29.5	304	17.6	9.76	1.62	14.7	2488	2477
	O-P+	42.6	251	16.5	8.62	2.03	13.5	2555	2457
	O-P-	47.5	304	17.5	9.37	1.43	14.9	2551	2440
SD	Target	44.0	283	4.48	2.61	1.74	5.32	232	297
	O+P-	79.9	629	4.29	2.65	2.00	5.16	276	308
	O-P+	137	384	4.63	3.10	2.02	5.32	283	312
	O-P-	88.3	372	4.43	2.92	1.95	5.87	254	340

Table 18. Post hoc tests results of the stimulus materials

	condition	**condition**	**Mean Difference**	**SE**	**df**	**t**	***p***
frequency	**Target**	**O+P-**	-3.75	13.9	356	-3.75	1.000
		O-P+	-16.93	13.9	356	-1.215	1.000
		O-P-	-21.80	13.9	356	-1.565	0.711
frequency C1	**Target**	**O+P-**	-105.770	65.0	356	-1.626	0.629
		O-P+	-53.510	65.0	356	-0.823	1.000
		O-P-	-106.530	65.0	356	-1.638	0.614
stroke_N. Word	**Target**	**O+P-**	0.600	0.665	356	0.903	1.000
		O-P+	1.678	0.665	356	2.524	0.072
		O-P-	0.656	0.665	356	0.986	1.000
stroke_N. C1	**Target**	**O+P-**	-0.067	0.422	356	-0.158	1.000
		O-P+	1.067	0.422	356	2.530	0.071
		O-P-	0.344	0.422	356	0.817	1.000
phono.NB Word	**Target**	**O+P-**	0.156	0.288	356	0.540	1.000
		O-P+	-0.256	0.288	356	-0.887	1.000
		O-P-	0.344	0.288	356	1.196	1.000
phono.NB C1	**Target**	**O+P-**	-1.189	0.809	356	-1.470	0.855
		O-P+	6.55e-15	0.809	356	8.10e-15	1.000
		O-P-	-1.422	0.809	356	-1.759	0.477
OLD C1	**Target**	**O+P-**	11.59	39.1	356	0.296	1.000
		O-P+	-55.57	39.1	356	-1.421	0.937
		O-P-	-51.74	39.1	356	-1.323	1.000
OLD C2	**Target**	**O+P-**	23.3	46.9	356	0.497	1.000
		O-P+	44.0	46.9	356	0.939	1.000
		O-P-	60.7	46.9	356	1.294	1.000

In addition to the critical word pairs, a set of 150 word pairs was generated as fillers, with 120 of them being semantically related and the remaining 30 being semantically unrelated. The reason why there were 60 fillers (of which 30 were semantically related and 30 semantically unrelated) less than in the semantic relatedness task in German is the large dissociation between orthography and phonology in Chinese. In the orthographically related condition, the phonological overlap can be completely separated from orthography. Conversely, in the German semantic relatedness task, orthographically related word pairs are also phonologically related, necessitating a larger number of fillers to minimize the possibility of strategic influences on the response. The full list of stimuli is provided in Appendix B.

Like the German stimuli used in Experiments 1 and 2, there is also no appropriate database concerning the degree of semantic relatedness in Chinese. Consequently, a comparable approach was adopted, wherein 29 Chinese native speakers who did not take part in the experiments were asked to rate the semantic relationship of all Chinese word pairs employed in this study through a questionnaire. The questionnaire was administered again through SoSci Survey (Leiner, 2019) and was specifically designed to gather rating scores assigned by the Chinese native speakers in terms of semantic relatedness. The results are summarized in Table 19. What can be clearly seen in Table 19 is the clear difference between semantically related and unrelated stimuli. According to the post-hoc test results, no distinction was evident between the orthographically/phonologically related and unrelated conditions. This suggests that the stimuli were well-controlled for semantic factors across both orthographically/phonologically related and unrelated word pairs (O+P- vs. O-P-: t = 2.593, p = 0.098; O-P+ vs. O-P-: t = 1.854, p = 0.654). Furthermore, a correlation analysis was conducted to examine the relationship between the mean reaction times in Experiments 3 and 4 and the semantic relatedness scores of the critical stimuli. The results indicated a lack of significant correlation between these two variables (in Experiment 3: *Pearson's r* = -0.015, p = 0.810; in Experiment 4: *Pearson's r* = 0.110, p = 0.070). Hence, we would not regard the semantic relatedness score as a dependent factor in our following data analysis.

Table 19. Summary of the semantic relatedness scores rated by Chinese native speakers

condition	**N**	**Mean**	**SD**	**SE**
O+P-	90	1.82	0.510	0.0538
O-P+	90	1.77	0.431	0.0454
O-P-	90	1.64	0.389	0.0410
Filler S+	120	5.00	0.414	0.0378
Filler S-	30	1.93	0.565	0.1032

A total of three versions of the experiment were created, with each version consisting of 90 critical word pairs and 150 filler word pairs. All items were randomly distributed across five blocks, wherein each block encompassed 48 word pairs. During the experiment, each participant received 30 orthographically related but phonologically unrelated word pairs, 30 phonologically related but orthographically unrelated word pairs, and 30 word pairs that were unrelated in all conditions. This design ensured that participants heard the target item only once throughout the entire experiment. Same to the semantic relatedness task in German, the participants were randomly assigned to one of the three versions and the selection of versions was based on the order in which they participated in the experiment.

Procedure and Apparatus

The auditory stimuli were recorded by a female native Chinese speaker, using the same sound recording equipment employed in Experiment 2. Participants were individually tested in the same room and used identical facilities as described in Experiment 2. To ensure consistency, the experimental software *OpenSesame (*Mathôt, Schreij, &

Theeuwes, 2012) was also used to present stimulus and collect data.

Like Experiment 2, participants were instructed to give the response as quickly as possible to determine whether the two words they hear are semantically related by pressing the buttons on the response box. All instructions were given in Chinese. The experiment lasted for approximately 25 minutes.

3.4.2 Results

Before conducting statistical analyses, a comparison was conducted between semantically related fillers and other semantically unrelated word pairs. The pairwise comparison results revealed that semantically related fillers yielded significantly faster responses ($t = -4.386$, $p < 0.001$) and a significantly higher error rate ($z = -4.367$, $p < 0.001$) compared to the other conditions. This outcome confirmed the effectiveness of the semantic relatedness task given to the participants.

Similar to the analysis conducted in Experiments 1 and 2, all incorrect responses (less than 9%) were excluded from further analyses of response times. RTs that fell below the first quartile minus 3 times the IQR and those exceeding the third quartile plus 3 times the IQR of correct responses were discarded (less than 0.4%) (Tukey, 1977). All results were summarized in Table 20.

Table 20. Mean reaction times (RT) and accuracy rates in Experiment 3.

	Condition	**RT**	**Accuracy**
Mean	O+P-	1.462	0.927
	O-P+	1.391	0.950
	O-P-	1.447	0.933
SD	O+P-	0.410	0.260
	O-P+	0.401	0.218
	O-P-	0.404	0.251

3.4.2.1 Response time analysis

Response time results were analyzed with linear mixed-effects models with participant and item as random factors. Due to the small sample size of Chinese participants, I didn't use the maximal model with by-participant and by-item random intercepts and slopes.

The inclusion of the fixed-factor *condition* significantly improved the model fit compared to a model without this variable ($\chi2(2) = 37.106$, $p < 0.001$, $R^2_{\text{marginal}} = 0.006$), so the variable *condition* will be retained in the final model. Compared to the unrelated condition, response times in the phonologically related condition were significantly faster (*Estimate* = -0.061, *SE* = 0.013, *t* = -4.664, *p* < 0.001); by contrast, the orthographically related condition elicited an inhibition on response time (*Estimate* = 0.014, *SE* = 0.013, *t* = 1.079, *p* = 0.28). But note that the inhibition effect here was not significant.

Table 21. The model with the best fit for RT data in Experiment 3. The reference level: Condition – O-P-

Group in Germany — RT: Mixed-Effect-Model with random intercepts for subject and item				
Random effects	**Name**	**Variance**	**SD**	
Subject	(Intercept)	0.053	0.230	
Item	(Intercept)	0.013	0.114	
Residual		0.102	0.319	

Fixed effects	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	1.460	0.039	37.886	
O-P+ : O-P-	-0.061	0.013	-4.664	< 0.001
O+P- : O-P-	0.014	0.013	1.079	0.28

3.4.2.2 Accuracy analysis

A mixed-effects regression model with a binomial family in R to the accuracy data was fitted, which was coded as 1 (correct) or 0 (incorrect) in the experiment. Participants and items were conducted as crossed-random effects, and condition (O-P+, O+P-, O-P-) as fixed effect (e.g., Baayen, 2008). The investigation of the phonological priming effect was conducted by contrasting the O-P+ condition against the O-P- condition, while the orthographic priming effect was examined by contrasting the O+P- condition against the O-P- condition. The fit of the model increased marginally with the fixed-factor *condition* ($\chi2(2) = 5.818$, p = 0.054, $R^2_{\text{marginal}} = 0.002$), so we kept the *condition* as fixed-effect factor in our model. The resulting model is depicted in Table 22, which showed that phonological priming elicited significantly more accurate responses (*Estimate* = 0.351, *SE* = 0.176, *z* = 1.988, *p* = 0.047), and orthographic priming slightly elicited more errors (*Estimate* = -0.063., *SE* = 0.163, *z* = -0.389, *p* = 0.698).

Table 22: The model with the best fit for accuracy data in Experiment 3. The reference level: Condition – O-P-

Group in Germany — Accuracy: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	1.097	1.048
Item	(Intercept)	0.957	0.979

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	3.413	0.241	14.171	
O-P+ : O-P-	0.351	0.176	1.988	0.047
O+P- : O-P-	-0.063	0.163	-0.389	0.698

3.4.3 Discussion

The present study was an effort to investigate the possible effect that orthographic and phonological overlap might have on Chinese spoken word recognition. It revealed that phonological similarities can be very influential on spoken word recognition, but the presence of orthographic similarities only resulted in a response delay without reaching statistical significance. These results go in the same direction as those reported in Qu and Damian's (2017) study, where orthographic overlap demonstrated a significant increase in response latencies.

The facilitatory phonological effects observed in this study can be attributed to the influence of bottom-up connections from phonological representations, which are modulated by the shared initial phonemes during Chinese spoken word recognition. As mentioned in Chapter 2, one of the most important characteristics of the Chinese language is the large number of homophones. This complex nature of homophones in Chinese can activate multiple word candidates, leading to increased difficulty in generating a response. However, the disyllabic word context serves as a crucial factor in narrowing down the activated candidates, resulting in facilitatory phonological effects during word recognition.

The facilitatory phonological effect observed in my study contrasts with the findings of Zou et al. (2012), which demonstrated that word-initial overlaps led to later N400 responses. It is worth noting that the first character of disyllabic Chinese primes used in their study also shares the same pronunciation with that of targets, and the words do not differ from other homophone characters until the second syllable is encountered. This characteristic can be viewed as the word-initial overlap in some alphabetical languages. Zou et al. (2012) ascribed their findings to the delayed point of uniqueness between word pairs and suggested that the response would be delayed when the initial phonemes of a semantically incongruous word matched the expected word. One possible explanation for the discrepancy may lie in the interstimulus interval (ISI) used in the respective studies. While Zou et al. employed an ISI of 150 ms in a lexical decision task, my study settled a shorter ISI of only 20 ms in a semantic judgment task. Consequently, participants of my study may not have enough time to generate sublexical and lexical conflict during this brief interval, which will slow down the response time. As a result, we did not observe the delayed response that would be expected according to the predictions of the Cohort model. Moreover, the primes used in the phonologically related condition were with great tonal similarity to targets (e.g., 铜牌, /tong2pai2/, *bronze medal* – 童年, /tong2nian2/, *childhood*), so they can be activated more strongly than those exhibiting low tonal similarity (e.g., 通知, /tong1zhi1/, *notification* – 童年, /tong2nian2/, *childhood*) (Shen et al., 2021).

In contrast, the presence of orthographic overlap did not facilitate auditory word processing. This may be attributed to the fact that orthographic overlap between word pairs tends to bias participants towards a “yes” response, whereas the correct response for the semantic relatedness task is “no”, and these incompatible responses consequently lead to a conflict which results in longer response times (Qu & Damian, 2017).

Another plausible explanation for the inhibitory effect of similar orthography could be attributed to top-down cognitive processes. The initial word is processed based on participants’ general knowledge and subsequently connected to new information. When two words exhibit orthographic similarities, such as sharing common radicals and stroke combinations, competition arises at the word level among these similar orthographic units. This competition between homographs activations and the consequent slowdown in feedback-activation increased the response time required by Chinese participants in semantic judgment tasks. Furthermore, according to TTRACE model (Tong et al., 2014), two words without phonological similarity consequently display the least activation. Therefore, words without phonological relations did not facilitate the response time.

However, in this task, although shared orthography between the prime and target resulted in longer response times than those in the unrelated condition, it didn’t reach a significant level. The inhibitory effect from our results is not as strong as in other studies (Mok et al., 2018; Qu & Damian, 2017; Zou et al., 2012), but the present result is consistent with previous studies that indicate the access to meaning without concurrent orthographic activation during the recognition of Chinese spoken words (Wang et al., 2012; Y. J. Wu & Thierry, 2010).

The present result is possibly related to the sample of subjects I selected. The Chinese participants included in this study were individuals who learned German as a second language and currently reside in Germany, where their daily usage of the Chinese language may be limited. With the prevalent use of computers and mobile devices, Chinese handwriting is gradually being replaced by typing and speech input devices. Native speakers of logographic languages, like Chinese, often rely on phonetic input methods and do not necessarily need to memorize the exact character. Furthermore, the acquisition of an alphabetic second language and immersion in an alphabetic L2-speaking environment may enhance the participants’ sensitivity to phonological information. As a result, the effects of orthography on spoken word recognition may be diminished, while phonological effects may become more prominent.

Therefore, in the next experiment, I intended further to investigate whether the homophone and homograph characters influence Chinese spoken word recognition, in particular, whether factors such as residence and experience with L2 are associated with individuals’ performance in processing their native language words.

3.5 Experiment 4: Experiment for Chinese participants living in China

The Chinese participants in Experiment 3 reported difficulties with writing Chinese characters, primarily due to their extensive use of the German language in their daily lives. When they want to write something, they will turn to computers and mobile phones with alphabet-based input systems (*pinyin* system), so they still can recognize the character, but they don’t necessarily need to recall and reproduce the orthographic forms. In other words, the emphasis shifts from the ability to write characters to the ability to pronounce them accurately. The weakened ability to recall the written form of the characters right away could eventually affect reading ability, which is relative to the orthographic effect (Zou et al., 2012). The impact is not limited to long-term migrants alone, speakers who use a second language while residing in an L2-speaking environment are also affected by that to some extent. Moreover, the memorization of character writing is so crucial to logographic languages, whereas it is comparatively less important in alphabetic languages. Chinese participants residing in Germany have limited exposure to their native language, particularly in terms of orthographic input and output. While they comprehend the meaning of Chinese words by hearing, they may be not necessary to recall the corresponding orthographic information. Therefore, I selected one group of Chinese participants who live in China, can't speak German, exclusively use Chinese in their daily lives, and have no experience of living abroad.

It can be hypothesized that Chinese participants living in China have sufficient exposure to orthography, so they rely equally on their knowledge of character pronunciation and writing forms, and the more stable phoneme-grapheme correspondence in their minds will result in faster and more accurate responses during L1 spoken word recognition. Also, we expect that Chinese participants who reside in China exhibit worse performance in phonological processing compared to those living in Germany. This may be attributed to that participants in Germany have long-term exposure to and usage of German as a second language may enhance their phonological sensitivity. Several studies have provided evidence that acquiring and employing a second language with an alphabetic script can enhance phonological awareness. For instance, it has been observed that children who receive extensive English instruction demonstrate better phonological awareness in Chinese (X. Chen et al., 2010). Therefore, it can be assumed that the participants' improvements in Chinese phonological performance may correspond to their learning experiences in their second language.

3.5.1 Method

Participants

Thirty-eight Chinese native speakers took part in this experiment. All participants had a monolingual background, hadn't learned German as L2, exclusively used Chinese in their daily lives, and had no experience living abroad. The participant group consisted of adults, with 26 females and 12 males, with a mean age of 24.9 years (ranging from 18 to 34 years old). They had either completed a college education or were college students currently. It was noted that they began learning English in school after the age of 9. None of them reported any hearing problems or a history of language problems.

Materials and design

For this experiment, the same semantic relatedness task as in Experiment 3 was employed. Because these participants were recruited in China, this experiment was conducted online. This online experiment was presented using *PsychoPy* software (Peirce et al., 2019). *PsychoPy* is a widely used, freely available software platform designed for running behavioral studies and supports online experiments through integration with *Pavlovia.org*. *PsychoPy* was selected for its ability to ensure precise and accurate timing of auditory stimuli and response measurements in online studies, compared to other software like *E-Prime*®, *NBS Presentation*®, *Psychophysics Toolbox*, *OpenSesame*, etc. (Bridges et al., 2020).

Procedure and Apparatus

Participants received written instructions to perform a semantic relatedness task in Chinese. They were specifically instructed to use the *Chrome* browser when accessing the experiment link. The *Chrome* browser was recommended due to its compatibility with running online studies from the *PsychoPy* software on both Windows and Linux platforms, ensuring sub-millisecond precision in timing measurements (Bridges et al., 2020). They have also been told that other software in the background should be closed to prevent random noise or reduced time resolution caused by computer overload. After that, participants completed an online questionnaire providing demographic information and details about their language experience. The procedure for the web-based version of the experiment was identical to the lab-based semantic relatedness task conducted in Experiment 3. Participants finished the experiment individually in a quiet room in their homes in China and were asked to wear headphones during the experiment. They were first presented with 12 practice trials. Following the practice session, the experiment comprised five blocks, each containing 48 trials. There were brief breaks between blocks, and the next block would not present until participants pressed any keys to indicate that they were ready to continue. During the task, participants were instructed to quickly determine whether the two words they heard were semantically related by pressing either the "x" or "m" key on the keyboard. The key “x” is just like the left button on the response box and the key “m” is like the right button. Either the “x” or “m” keys represent the answer “yes” was counterbalanced between participants, just like the setting of the response box in lab-based experiments. The response time was measured starting from the onset of the target word and ending at the participants’ response. The present experiment lasted approximately 25 minutes.

3.5.2 Results

Same as the analysis in Experiment 3, all incorrect responses (less than 14%) were excluded from further analyses for response time. And RTs smaller than the first quartile minus 3 times the IQR and greater than the third quartile plus 3 times the IQR of correct responses were discarded (less than 0.5%) (Tukey, 1977). Also, one trial with a response below 100 ms was excluded from the response time analysis. All results were shown in Table 23.

Table 23. Mean reaction times (RT) and accuracy rates in Experiment 4.

	Condition	RT	Accuracy
Mean	O+P-	1.482	0.820
	O-P+	1.408	0.907
	O-P-	1.434	0.886
SD	O+P-	0.398	0.385
	O-P+	0.360	0.290
	O-P-	0.382	0.318

3.5.2.1 Response time analysis

The response times in Experiment 4 were analyzed using a linear mixed model, with participants and items as a crossed-random factor (e.g., Baayen, 2008), and condition (O+P-, O-P+, O-P-) as the fixed-effect predictor. The unrelated condition was set as the baseline in the mixed-effects model and explored the phonological priming effect on words by contrasting the O-P+ with the O-P- condition as well as the orthographic priming effect by contrasting the O+P- with the O-P- condition.

This model showed a significantly better fit in comparison to a model lacking the fixed-factor *condition* ($\chi 2(2) = 30.627$, $p < 0.001$, $R^2_{\text{marginal}} = 0.009$); and compared to the unrelated condition, response time in the phonologically related condition was significantly faster (*Estimate* = -0.051, *SE* = 0.016, *t* = -3.277, *p* = 0.001), whereas it elicited a significant inhibition (*Estimate* = 0.037, *SE* = 0.016, *t* = 2.295, *p* = 0.022) in the orthographically related condition.

Table 24. The model with the best fit for RTs data in Experiment 4. The reference level for factors was as follows: Condition – O-P-.

Group in China — RT: Mixed-Effect-Model with random intercepts for subject and item				
Random effects	**Name**	**Variance**	**SD**	
Subject	(Intercept)	0.033	0.181	
Item	(Intercept)	0.008	0.092	
Residual		0.104	0.322	

Fixed effects	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	1.458	0.033	44.029	
O-P+ : O-P-	-0.051	0.016	-3.277	0.001
O+P- : O-P-	0.037	0.016	2.295	0.022

3.5.2.2 Accuracy analysis

Due to the binary nature of the data, an analysis was carried out for accuracy using a binomial family in R (Jaeger, 2008).

From the likelihood ratio test we could see that this model demonstrated a significantly better fit in comparison to a model without the fixed-factor *condition* ($\chi2(2) = 31.154$, $p < 0.001$, $R^2_{marginal} = 0.011$); and in contrast to the unrelated condition, the accuracy rate in the phonologically related condition was significantly higher (*Estimate* = 0.230, *SE* = 0.153, *z* = 1.961, *p* = 0.050), whereas in the orthographically related condition, the accuracy rate decreased significantly (*Estimate* = -0.496, *SE* = 0.138, *z* = -3.593, *p* < 0.001). The results of the final model are summarized in Table 25.

Table 25. The model with the best fit for accuracy data in Experiment 4. The reference level for factors was as follows: Condition – o-p-.

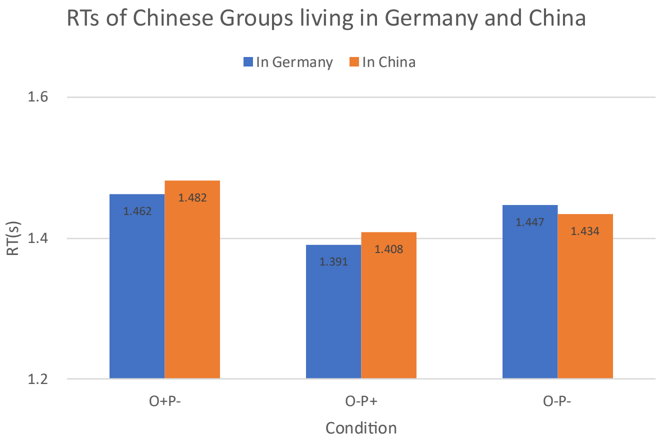
Group in China — Accuracy: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.677	0.823
Item	(Intercept)	0.304	0.551

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	2.343	0.183	12.789	
O-P+ : O-P-	0.230	0.153	1.961	0.050
O+P- : O-P-	-0.496	0.138	-3.593	< 0.001

3.5.2.3 Differences between Chinese groups living in China and Germany

To investigate the influence of immersion in an L2 spoken environment on activation of L1 orthographic and phonological information, the results from Experiment 4 were compared with that from Experiment 3. Generally speaking, both Chinese groups living in China and Germany showed the fastest response and the best accuracy rate in the phonologically related condition, as well as the slowest response and the most errors in the orthographically related condition (see Figure 16). Planned comparisons on response time in Chinese groups living in China and Germany showed that there were no significant differences between these two groups on response time either in general ($t = 0.960$, $p = 0.337$) or in separate conditions (O-P+: $t = 1.084$, $p = 0.278$; O+P-: $t = 1.071$, $p = 0.284$; O-P-: $t = -0.497$, $p = 0.619$). With respect to accuracy rate, we found that the group in Germany exhibited significantly better accuracy than the group in China either in general ($z = -8.687$, $p < 0.001$) or in separate conditions (O-P+: $z = -3.909$, $p < 0.001$; O+P-: $z = -7.551$, $p < 0.001$; O-P-: $z = -4.013$, $p < 0.001$).



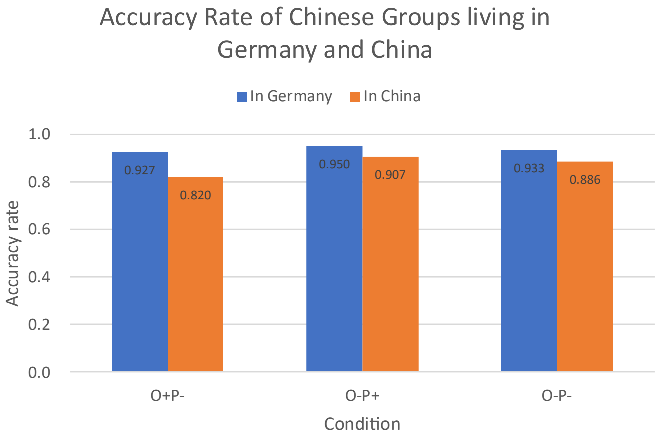


Figure 16. Statistical diagrams of RTs and accuracy in groups living in Germany and China.

Then the results of the two groups were combined and an analysis including residence place (Germany vs. China) in interaction with priming condition (O-P+ vs. O+P- vs. O-P-) was carried out on response time as well as accuracy rate. The analysis results didn’t show a significant interaction of priming condition and residence place for the response time ($\chi2(2) = 0.726$, $p = 0.696$) and accuracy rate ($\chi2(2) = 4.886$, $p = 0.087$), indicating that different priming conditions didn’t significantly impact RTs and accuracy rate in the group in China as compared to the group in Germany. Hence, the interaction term was removed from the model and the condition and residence were kept as the fixed-effect factors in the final model. The results are indicated in Table 26 and Table 27. It can be observed that the influence of both phonological priming (RT: *Estimate* = -0.053, *SE* = 0.010, $t = -5.354$, $p < 0.001$; Accuracy: *Estimate* = 0.315, *SE* = 0.117, $z = 2.683$, $p = 0.007$) and orthographic priming (RT: *Estimate* = 0.021, *SE* =

0.010, *t* = 2.046, *p* = 0.041; Accuracy: *Estimate* = -0.315, *SE* = 0.106, *z* = -2.969, *p* = 0.003) is very significant.

Table 26. Results from the RT analyses for Chinese groups in China and Germany.

Chinese Group — RT: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.043	0.208
Item	(Intercept)	0.010	0.101
Residual		0.103	0.322

Fixed effects	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	1.465	0.036	30.783	
O-P+ : O-P-	-0.053	0.010	-5.354	< 0.001
O+P- : O-P-	0.021	0.010	2.046	0.041
residence	-0.009	0.048	-0.181	0.857

Table 27. Results from the accuracy analyses for Chinese groups in China and Germany.

Chinese Group — Accuracy: Mixed-Effect-Model with random intercepts for subject and item

Random effects	**Name**	**Variance**	**SD**
Subject	(Intercept)	0.841	0.917
Item	(Intercept)	0.456	0.676

Fixed effects	**Estimate**	**Std. Error**	**z-value**	**p-value**
(Intercept)	2.333	0.191	12.231	
O-P+ : O-P-	0.315	0.117	2.683	0.007
O+P- : O-P-	-0.315	0.106	-2.969	0.003
residence	0.948	0.231	4.109	< 0.001

Then the effect of orthography and phonology in two subgroups with different residence places was inspected separately. In Experiment 4, the orthographic effect was significant (RT: *Estimate* = 0.037, *SE* = 0.016, t = 2.295, p = 0.022; Accuracy: *Estimate* = -0.496, *SE* = 0.138, z = -3.593, $p < 0.001$), but in Experiment 3 the orthographic effect was not obvious (RT: *Estimate* = 0.014, *SE* = 0.013, t = 1.079, p = 0.28; Accuracy: *Estimate* = -0.063., *SE* = 0.163, z = -0.389, p = 0.698). Hence, one of the hypotheses that the orthographic effect was stronger for the Chinese group living in China than in Germany received statistical support. However, the influence of phonological similarity on spoken word recognition didn't show a clear distinction between the two subgroups.

3.5.3 Discussion

Briefly, the findings of Experiment 4 revealed an inhibitory effect of orthographic similarity and a facilitatory effect of phonological overlap on Chinese disyllabic word recognition. More specifically, compared with unrelated condition, responses were slower and less accurate when the targets were preceded by orthographically similar primes, while they were faster and more accurate when word pairs exhibited phonological overlap. Further, compared to the results of Experiment 3, the effect size of orthography on spoken language is different between the participants in China and Germany: participants in China demonstrated pronounced orthographic effects, whereas participants in Germany only presented moderate orthographic effects using the same auditory task.

Generally, the observed orthographic effects in Experiment 4 might be resulted from the bidirectional connections between the sublexical phonological representations and orthographic representations during spoken word recognition, as proposed by TTRACE model (Tong et al., 2014). Different tones and phonemes in the primes and targets activated their respective orthographic candidates, then the best-matched candidate was activated by the disyllabic word context, meanwhile, similar written forms in the primes and targets competed at the word level to some extent. This competition influenced the judgment of the semantic relation between the primes and targets, indicating that orthography played a role during the overall spoken word recognition process.

But why was a significant inhibitory effect seen in the orthographically related condition in Experiment 4 but not in Experiment 3? One crucial factor is the difference in participants' exposure to printed Chinese characters. Participants living in China have greater exposure to printed Chinese characters in their daily lives compared to participants in Germany, who may have limited exposure to reading and writing Chinese characters due to their long-term residence abroad. As a result, the representations of orthographic information are likely to be stronger in participants living in China, leading to more pronounced orthographic effects in spoken word recognition. Conversely, for participants in Germany, who have less exposure to printed Chinese characters, orthographic effects in spoken word recognition may be relatively weak. However, participants in both Experiment 3 and Experiment 4 exhibited a significant benefit of L1 homophones, suggesting the early point of uniqueness between word pairs during hearing Chinese. This finding supports the idea that phonological information of the initial syllable significantly facilitates word recognition, regardless of the writing system employed. Furthermore, the increasing use of computers and mobile devices has resulted in a shift from handwriting to typing in logographic languages. By installing the appropriate keyboard settings on their phones, they can simply type the *pinyin*, and the characters will appear automatically. Hence, there has been a reduced emphasis on character writing skills among native speakers and an increased reliance on pronunciation and character recognition, potentially leading to impaired orthographic effects in spoken word recognition. Moreover, in view of arguments from Grosjean (1989) and Cook (1997, 2002), learning a second language changes the way our first language is represented in our minds and the L1 can be harmed by the use of an L2. Supporting this argument, Zou et al. (2012) conducted a study that revealed a positive correlation between the orthographic effect in Chinese spoken word recognition and the basic reading performance of Chinese native speakers. This suggests that sufficient exposure to a second language, particularly one with a sound-based writing system like German, may undermine the reading skills of Chinese characters, thereby influencing the orthographic effect when hearing Chinese words.

However, there was another argument suggesting that the L1 can be enhanced through the use of the L2. Yelland et al. (1993) conducted a study involving English children who received one hour of Italian language course per week, and they found that these children exhibited better reading skills in English compared to those who were not taught Italian. One possible explanation for this finding could be the linguistic closeness between the two languages (Yelland et al., 1993). The languages employed in this study (English and Italian) are alphabetic languages, which may contribute to cross-linguistic transfer and the beneficial effects observed in the participants' L1 proficiency.

Nevertheless, the homophone character in the disyllable words in both Experiments 3 and 4 reduced the response time, which is inconsistent with the result from Zou et al. (2012) that longer response times were taken for phonologically related words during the lexical decision task. The reason for this different result was already discussed in Experiment 2.

It's noteworthy that there is no significant difference in mean response time under every condition between the two groups. However, the group in Germany showed slightly faster responses in both phonologically and orthographically related conditions and demonstrated significantly higher accuracy rates in the semantic relatedness judgment tasks compared to the participants in China. The lower accuracy rate observed in the group in China may be attributed to the different testing environments. In Germany, participants were tested in a controlled, quiet lab setting, whereas in China, they were tested online, so it would not be guaranteed whether they will remain focused throughout the whole process of the experiment, leading to increased error rates.

In summary, the behavioral data directly speak to the issue that the Chinese native speakers who have lived in L2-spoken countries for a long time exhibit reduced reliance on orthography and greater dependence on phonological overlap during Chinese spoken word recognition, and those who reside in China all the time and use Chinese in their daily lives are easily influenced by orthographic similarities during hearing words, even though the Chinese language is characterized by a deep orthography with limited correspondence between graphemes and phonemes.

Chapter 4. General Discussion and Conclusion

In this chapter, I will begin by presenting a comprehensive summary of the main findings obtained from the experiments in Chapter 3. Subsequently, a general discussion of the results will be conducted, mainly focusing on the research questions addressed in the study, accompanied by a new bilingual model of spoken word recognition. Furthermore, a brief conclusion of the dissertation will be provided. Finally, an exploration of the limitations of the present study as well as potential directions for future research will be revealed.

4.1 Summary of results

To explore orthographic effects in L1 and L2 auditory word processing, four experiments were conducted in Chapter 3.

Experiment 1 investigated the orthographic role in German spoken word recognition among German native speakers. In the study, German participants were recruited to participate in a task that explored the implicit processing of orthographic and phonological information during spoken word retrieval. To achieve this, prime-target pairs with varying degrees of orthographic and phonological overlap between primes and targets were used in a semantic relatedness task. The findings revealed that participants responded significantly faster when the prime and target words were solely phonologically related than were unrelated. However, the presence of phonological primes led to a notable reduction in accuracy rates. Conversely, the orthographic overlap did not result in a significant facilitation of accuracy rates or response times. Moreover, word pairs that were orthographically and phonologically unrelated elicited the longest response times and yielded the highest error rates. These results suggested that the orthographic effect during spoken word recognition might be language selective.

Experiment 2 employed the same experimental design as Experiment 1, with a specific focus on Chinese participants possessing intermediate and high language proficiency in German. Generally, the results revealed significant effects in the phonological priming condition, while the effects in the orthographic priming condition were not significant. When we inspected the performance in two subgroups with different language proficiency separately, we did find a clear influence of language proficiency on overall mean response time and accuracy rate. Moreover, compared to the O-P+ condition, we observed that primes in the O+P+ condition facilitated the spoken word recognition of the target in highly proficient Chinese learners of German significantly, but only moderately in the group with intermediate language proficiency. And in both subgroups, the phonological overlap condition led to a noteworthy decrease in reaction times and a significant improvement in accuracy rate compared to the unrelated condition. Moreover, the O+P+ condition exhibited a notable rise in accuracy compared to the O-P+ condition, but this effect was observed exclusively in the highly proficient group; in the intermediate group, the difference did not attain statistical significance. The findings of Experiment 2 provide insights into the interplay between language proficiency and orthographic information processing during spoken word recognition among Chinese learners of German. The results suggested the influence of orthography in L2 spoken word recognition is proficiency-dependent but not associated with the target language.

Experiment 3 examined the orthographic role in Chinese spoken word recognition among Chinese learners of German residing in Germany for a long time. In Chinese, the orthographic form has the function of distinguishing homophonic morphemes, making orthography a critical factor in auditory word recognition. Similar to Experiment 2, Chinese participants were tasked with judging the semantic relatedness of prime and target pairs, but this time the pairs consisted of words from their native language – Chinese. The prime-target pairs were categorized as either orthographically related, phonologically related, or unrelated. The results of Experiment 3 revealed that processing Chinese spoken word pairs with orthographic similarities led to slightly longer reaction times compared to unrelated word pairs whereas the phonological similarities took significantly shorter reaction times than unrelated word pairs. Furthermore, phonological priming significantly elicited more accurate responses, while orthographic overlap only slightly increased the error rates. These findings suggested that orthographic similarities could not really affect Chinese spoken word recognition among individuals with extensive exposure to an L2-dominent environment.

Experiment 4 was designed as a contrast to Experiment 3, with the recruitment of Chinese participants who reside in China, never learned German, and have no experience of living abroad. In contrast to the unrelated condition, judgment in the phonologically related condition was significantly quicker and more accurate, while in the orthographic overlapping condition was significantly slower and less accurate. The results of Experiment 4 revealed a significant inhibitory orthographic effect and a significant facilitatory effect when prime-target word pairs shared the same homophone initial character. By comparing these results with those of Experiment 3, we observe a more pronounced influence of orthographic similarities when Chinese participants consistently reside in a monolingual environment and are continuously exposed to their native language. This finding suggested that the acquisition of an alphabetic L2 and the long immersion in an alphabetic L2-spoken environment can reduce speakers' sensitivity to orthographic information during spoken word recognition, especially among individuals from logographic L1 backgrounds.

4.2 General discussion

By investigating the different performances of L1 and L2 groups in the four experiments, this dissertation provides a better understanding of the impact of orthographic similarity on spoken word recognition. The general discussion will be structured into the following aspects and a newly modified model for L2 spoken word recognition will be described at last.

4.2.1 The influence of orthography on spoken word recognition in different writing systems

Based on the findings, it is suggested that the absence of the orthographic effect in the experiments is likely a language-dependent result. Previous studies exploring the orthographic role in spoken word recognition primarily involved English native speakers and used English stimuli, whose grapheme-to-phoneme mappings are considerably inconsistent. The results of these studies demonstrated the important role of orthography in spoken word processing in English. Also, in studies on other languages such as French (Pattamadilok et al., 2009) and Portuguese (Ventura et al., 2004), the influence of orthographic inconsistency or complexity on spoken word recognition was evident. However, in our study involving German, this effect was not observed. This distinction may be attributed to the fact that German has a highly transparent and consistent orthographic system, with highly predictable grapheme-to-phoneme correspondence rules. Consequently, in our behavioral data, orthographic similarities did not significantly impact the participants' responses, which is consistent with the results from Türk and Domahs (2022) in the German language, where no significant orthographic effects were observed in the behavior data. These findings supported that orthographic depth could influence how orthographic and phonological information interact during spoken word recognition. Because of the consistency in the shallow orthography, native speakers can rapidly activate phonological units at the whole-word level through the acoustic features of spoken words, without requiring additional information such as sublexical orthographic forms to differentiate between words. This, in turn, leads to the quick activation of semantic information.

But how is this effect in a language with another orthographic system which has been classified to be very deep? In the Chinese experiment conducted with monolingual Chinese native speakers, a significant inhibitory orthographic effect was observed. This result can be attributed to the unique characteristics of the Chinese writing system, which differs from the German's grapheme-phoneme correspondence rules. Unlike German, where phonological information is encoded within single graphemic units, the phonology of Chinese syllables is represented in graphemic characters. However, in multiple-unit characters (known as "compound characters"), phonology is indicated by one of the internal elements, known as the phonetic radical. For example, 晴 (/qing2/, *sunny*) as a compound character has a semantic radical 日 which delicates the meaning of sun, and a phonetic radical 青 /qing1/ which shares the same *Pinyin* with 晴 but with a different tone. Chinese characters do not have phoneme-to-grapheme correspondences but rather syllable-to-character correspondences. Thus, in Chinese spoken word recognition, participants must determine which radical or character represents the phonological information they heard, making the orthographic form of the character crucial for their comprehension of the words.

Note that similar orthographies in Chinese words showed an inhibition whereas in English, French, and Portuguese similar orthographies reduced difficulty in lexical access (Pattamadilok et al., 2009; Perre et al., 2009; Ventura et al., 2004). One possible explanation might be that the bidirectional connections between orthographic and phonological representations during spoken word recognition differ in strength in the two different writing systems. In alphabetical languages, such as English, the written form of each word is directly represented by a sequence of phonemes that correspond to acoustic signals, allowing for automatic prelexical processing when a prime and a target share orthographic overlaps. This facilitates the recognition of spoken words and leads to the observed facilitatory orthographic effects. But in logographic languages like Chinese, the relationship between orthography and phonology is not as direct as in alphabetical languages, because their written form is represented by characters, and the acoustic signals are transformed into syllables, which then activate their corresponding characters; but due to the nature of Chinese characters (e.g., homographs and homophones), preactivated sublexical phonological representations arbitrarily mapped to different written forms, and the similar written forms in the primes and targets result in competition at the word level to some extent, leading to inhibitory effects. Meanwhile, the best-matched orthography is chosen through the disyllabic word context.

Moreover, the comparison between the results from participants living in Germany and China told us that the activation of Chinese orthography might be weakened when the language environment is changed. Indeed, the size of orthographic effects in spoken word processing also depends on the quality of orthographic representations, which means that individuals with stronger orthographic knowledge tend to experience a higher degree of interference in spoken language recognition (Dich, 2011; Ziegler & Muneaux, 2007). Participants living in China have greater exposure to printed Chinese characters in their daily lives compared to those in Germany, who may have limited exposure to reading and writing Chinese characters due to their long-term residence abroad. As a result, the representations of orthographic information are likely to be stronger in participants living in China, leading to more pronounced inhibitory orthographic effects during spoken word recognition. In contrast, the participants living in Germany, with reduced exposure to Chinese orthography, may exhibit weaker orthographic effects in the same spoken language processing task.

The results from both Experiment 3 and Experiment 4 demonstrated a significant advantage for L1 homophones in Chinese spoken word recognition, suggesting the early point of uniqueness between word pairs while hearing Chinese. This finding supported that phonological information of the initial syllable contributes significantly to word recognition, regardless of the writing system employed. Besides, the primes used in the phonologically related condition were with great tonal similarity to targets, so they can be activated more strongly than those with low tonal similarity (Shen et al., 2021).

In sum, the impact of orthography on spoken word recognition might vary across different writing systems and might be influenced by changes in first language use frequency and living environment. Whether the contact with alphabetical L2 plays a role in altering the orthographic effect on L1 auditory word processing, remains to be investigated in future research. As an example, we could conduct a study comparing a group that learned an alphabetical L2 with a group that learned a logographic L2, having them both performing the same task in their L1.

4.2.2 The orthographic effects during spoken word recognition in L1 vs. L2

Comparing the results of spoken word recognition in L1 with that in L2, we can find some interesting differences.

Chinese participants were impacted by similarly written forms during their L1 and L2 spoken word recognition. But the similar written forms in their L1 reduced the response efficiency whereas in their L2, the similar written forms accelerated the response.

The absence of inhibition in Chinese participants when responding to orthographically related words in their L2 suggests that the way they processed their L1 orthography did not influence their L2 orthography processing, especially when the writing systems of L1 and L2 are entirely different. This finding supported that L2 users with L1 writing systems similar to that of their L2 can potentially draw on their L1 system to some extent, whereas L2 users with entirely distinct L1 backgrounds must develop a new system for their L2 (Cook, 2006).

The facilitatory influence of orthography in L2 is consistent with previous findings. For instance, Veivo and Järvikivi (2013) observed a beneficial effect of orthographic onset overlap when French-Finnish bilinguals responded to L2 word pairs sharing similar orthographic forms. Similarly, Türk and Domahs (2022) reported a facilitative orthographic priming effect in spoken word recognition for late German-English bilinguals in English, consistent with the findings of Perre and colleagues (2009) who investigated English native speakers.

Contrary to previous studies conducted in alphabetical languages (Pattamadilok et al., 2009; Perre et al., 2009; Ventura et al., 2004), our results do not support the facilitatory nature of the orthographic effect in L1. In our study, German and Chinese participants' judgments were slightly slower when there were orthographic overlaps between the primes and targets in their L1.

During L2 spoken word recognition, the role of orthography is influenced by language proficiency, as evidenced by the significant orthographic priming effect observed only in the highly proficient participants group, but not in the intermediate proficiency group. This finding may be explained by the association between the strength of orthographic influence and the quality of orthographic representations, in other words, stronger orthographic knowledge might result in greater interference in spoken language processing (Dich, 2011; Ziegler & Muneaux, 2007). This finding suggests that the efficiency of integrating orthographic, phonological, and semantic information may be less robust in the group with intermediate language proficiency compared to those with high proficiency.

However, why didn't German native speakers show an orthographic effect on German spoken words like the Chinese participants with high language proficiency? We can attribute the orthographic effect in the L2 group with high language proficiency to L2 learning, regardless of the language difference, which results in an orthographic effect that is not present in the L1 group. This explanation is consistent with findings from Veivo and Järvikivi (2013) who investigated French-Finnish bilinguals responding to word pairs with similar or dissimilar orthographic forms. They found that only participants with high language proficiency were influenced by orthography, and they concluded that proficiency modulated the orthographic influence on spoken word recognition. Other studies by Mitsugi (2018) (with English learners of Japanese), Qu et al. (2018) (with Tibetan learners of Chinese), and Türk and Domahs (2022) (with German learners of English) also found orthographic effects in L2 spoken word recognition. The languages investigated in these studies encompassed syllabic, logographic, and alphabetical systems and ranged from very shallow to very deep orthographic depth. Hence, these findings suggest that the orthographic effect may be stronger in the L2 group compared to the L1 group, and the influence of orthography on spoken word recognition in L2 might be independent from the target language. In other words, the proficiency level in L2 may play an important role in determining the degree to which orthographic information influences spoken word processing, regardless of the specific characteristics of the L1 and L2.

It is noteworthy that both the German and Chinese participants demonstrated a significant facilitation effect of phonological overlap compared to unrelated distractors. However, the impact of orthographic similarity differed between the two groups. In the German native group, orthographic and phonological similarity resulted in slightly slower response times and decreased accuracy rates compared to only phonological overlapping conditions, whereas in the L2 group, orthographic similarity accelerated response times and improved accuracy rates. The results from Türk and Domahs (2022) also suggest that orthographic overlap counteracts the facilitation effect caused by phonologically identical rimes among German native speakers. But similar orthography really helps Chinese learners of German to perceive the meaning of words. It is possible that bilinguals are better at separating written form from language meaning (Cook, 1997). The orthographic overlap between word pairs can lead to an erroneous bias toward a “yes” response, while the correct response for critical stimuli should be “no”, thus generating a conflict between incongruent responses (Qu & Damian, 2017). This conflict can result in longer response times and more errors, particularly observed in native participants. But our L2 participants may possess an enhanced ability to separate meaning from similar orthography, thus the orthographic overlap didn’t lead to significant conflicts in semantic judgment compared to the German L1 group. Moreover, native speakers are able to rapidly activate semantic information through the acoustic features of spoken language. This activation follows a bidirectional process, as described by the TRACE model. In contrast, non-native speakers tend to rely more on bottom-up processing. When the auditory input matches their phonological representations, both orthographic and phonological lexical, as well as sublexical representations are activated, ultimately leading to the activation of semantic information. Just in the O+P+ and O-P+ conditions, the presence of orthographic and phonological overlapping makes non-native listeners focus so much attention on identifying sounds and respective phonemes, leading to greater bottom-up activation compared to the unrelated condition.

The results we found indicate that the orthographic effect could be more prominent in the L2 group when it compared with the L1 group, and the influence of orthography on spoken word recognition in L2 might be language non-selective.

4.2.3 New model of L2 spoken word recognition for Chinese-German bilinguals

Combining all findings, I want to propose a new model that accounts for the interactive processes of spoken word recognition in L2, which is called the new interactive activation model for Chinese-German bilinguals. This new model is based on the previous TTRACE model, BIA+ model, and BIMOLA model and is illustrated in Figure 17. There are still some similar features that are inspired by previous monolingual and bilingual models, such as three levels of structure: acoustic features, phonemes, and words. However, given that Chinese is a tonal language, additional annotation (T1, T2, T3, T4) was generated for tonemes, alongside the phonemes level. Furthermore, the activation connections between the phoneme and word level are bidirectional, but the phonemes only received bottom-up activation from the features level.

As reviewed in the BIA+ model (Dijkstra & van Heuven, 2002), the bottom-up recognition process tends to be considerably faster. This is consistent with the fastest response times observed in the orthographic overlapping condition and the longest response times in the O-P- condition in the L2 group. However, it is essential to acknowledge that the assumption of the BIA+ model is that both languages involved should have similar orthographic characteristics, and activation is considered language non-selective, with access to language nodes occurring only after lexical phonological and orthographic units have been activated. But Chinese and German have significant orthographic differences and the results demonstrate that Chinese native speakers exhibit distinct processing effects for their L2 German and L1 Chinese. Thus, the BIA+ model might not fully account for the findings observed in our Chinese L2 group.

An alternative model, the BIMOLA model, offers another assumption, that is, L1 and L2 are represented by distinct subsets of phonological units at the features level, the phonological units of L1 and L2 exist as one large system (or cluster), and the two languages are restricted to when L2 shares its alphabet with L1, so the number of activated orthographic or phonological codes is determined by neighborhood density and word frequency, which does not hold true for Chinese-German bilinguals due to those phonological units of the Chinese language contain tonemes. Furthermore, it is important to note that our stimuli were exclusively presented in either German or Chinese and did not include cognates, so cross-language inhibition was not explored in our experiments. As a result, the route of L1 and L2 will be independently illustrated, and connections between phonemes & tonemes level and words level will be considered within each respective language. In other words, the new model does not simulate the cross-language effect, but this aspect can be subject to investigation in future research.

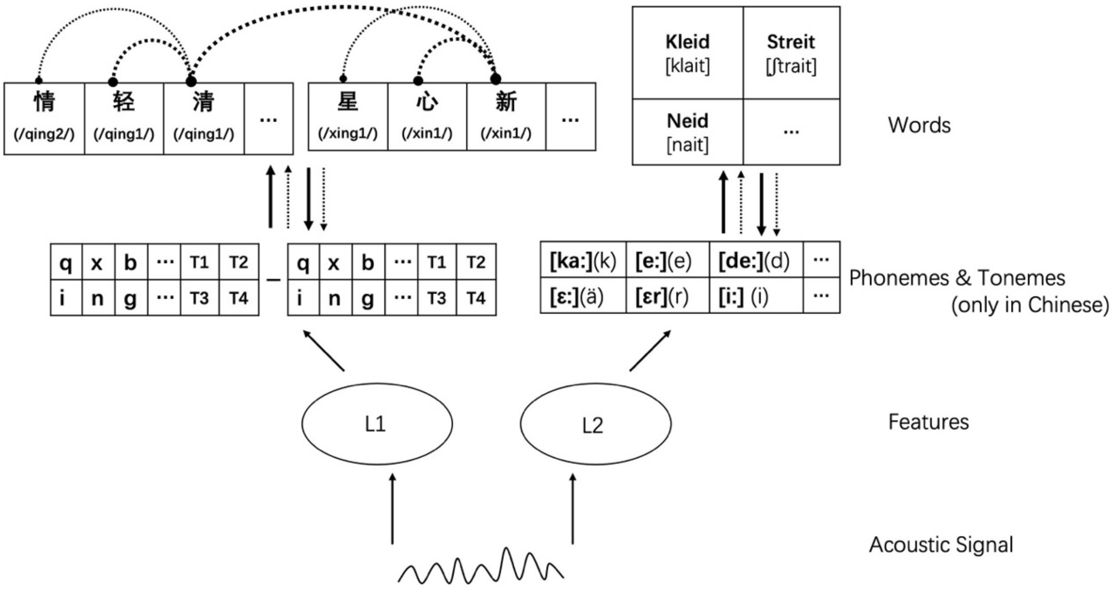


Figure 17. The new interactive activation model for Chinese-German bilinguals

In the L1 route, the arrows between orthographic units and sublexical phonological units are bidirectional, indicating that the top-down and bottom-up processing interact during the auditory processing. It is important to note that the thickness of the arrows between these two levels in the model represents the activation of orthography in L1, which is determined by the extent of exposure to L1: the fine dotted line indicates limited exposure to L1 in an L2-spoken environment and the thick full line represents sufficient exposure to L1. In the case of Chinese listeners, the impact of L1 orthography becomes less pronounced when their exposure to the language is limited, such as in the context of living abroad. Additionally, since many Chinese single words are represented by homophone characters, the written form and meaning of the word cannot be identified until the second character is heard, further influencing the orthographic processing in the model. Take characters 晴(/qing2/), 轻(/qing1/), and 清(/qing1/) as examples. They share the same syllable (/qing/) but have different written forms (轻 is written differently from 晴 and 清) or different tonemes (晴/qing2/ is with the second tone, and 清/qing1/ is with the first tone), so listeners can identify the meaning of the word only when the second character 新(/xin1/) comes into play, allowing the orthography of the word 清新(/qing1xin1/, *fresh*) to be recognized; meanwhile, other characters sharing a similar syllable, such as 星 (/xing1/), or having the same pronunciation, such as 心(/xin1/), may also be considered as candidates, but only 新 (/xin1/) will be activated due to the semantic context provided by the first character. As Figure 17 illustrates, the strength of activation at the word level is represented by the thickness of the curved dotted lines. We could see that 清 (/qing1/) and 新 (/xin1/) gained stronger activation compared with other candidates.

In L2 route, the different types of arrows between phonemes & tonemes level and words level symbolize the activation of orthography in L2, which is determined by the proficiency of L2: the dotted line indicates intermediate proficiency, and the full line represents high proficiency. This is an important aspect of the new model, the strength of orthographic activation at the word level is not only determined by neighborhood density and word frequency like in BIMOLA model (in Grosjean, 2008) but also necessarily by the proficiency of L2. As can be seen from Figure 1, the degree of thickness of the arrows between word and phonemes & tonemes level represents the activation strength of orthography in L2 which is determined by the proficiency of L2, which means for Chinese listeners that the impact of L2 orthography becomes stronger when the proficiency of their L2 gets improved. For example, if the target word *Kleid* is heard and the acoustic features match the sublexical phonological representations, listeners will recognize it as a German word. In listeners with high proficiency in L2, the correspondence between phoneme and grapheme is more stable, both phonological and orthographic lexical as well as sublexical representations will be activated quickly, so phonologically related candidates such as *Streit* and orthographically related candidates such as *Neid* will be activated, which subsequently activate the semantic representations. They paid so much attention to identifying the phonemes, leading to enhanced bottom-up processing, which is usually much faster (Dijkstra & van Heuven, 2002). However, in listeners with intermediate L2 proficiency, the orthographic lexical representations may not be strongly activated, and they are more willing to access semantic representations through the phonology-semantics pathway, resulting in weaker bottom-up and top-down activation.

This new model provides a specialized version of bilingual lexical access, distinguishing itself from the BIA, BIA+, and BIMOLA models. It focuses on auditory processing and highlights that the strength of orthographic activation at the word level is not only determined by neighborhood density and word frequency but also necessarily by the proficiency of L2 and exposure to the L1, that is, for Chinese learners of German, the influence of L1 orthography is diminished with limited exposure to their native language, while the impact of L2 orthography is strengthened with improved proficiency in their L2.

4.3 Conclusion

This research addresses the role of orthography in spoken word recognition across different groups, including German native speakers, Chinese L2 learners of German, and Chinese monolinguals. While the significance of orthographic information in spoken word recognition has been observed in both alphabetic and non-alphabetic languages, its impact on L2 auditory word processing, particularly for learners with a non-alphabetic L1 background like Chinese, remains less clear. Unlike German, which features a shallow alphabetic orthography with consistent grapheme-phoneme correspondences, Chinese is a logographic language with a deep orthography and arbitrary mappings between written characters and phonological form. The main aim of my studies is to investigate which role orthographic and phonological representations play during word recognition in both L1 and L2, and to examine how language proficiency in L2 affects the processing of auditory words in Chinese learners of German.

To explore this issue, two groups of Chinese participants with intermediate and high language proficiency in German and a German control group were recruited to perform tasks on implicit processing of orthographic and phonological information in spoken word retrieval. The findings demonstrate that whether the orthographic form influences spoken word recognition is not simply limited by the types of writing system (alphabetic vs. logographic), but by the orthographic depth, because results suggested that orthographic similarity doesn't impact auditory word processing in German native speakers, whereas such influence is found in previous research conducted with other languages (e.g., English, Portuguese, French, Chinese, etc.). The results also tell us that language proficiency in L2 indeed influences the processing of auditory words when the listeners are Chinese learners of German since there is an obvious orthographic effect in the L2 subgroup with high proficiency but not with intermediate proficiency. We suggested that the orthographic effect in L2 spoken word recognition might be proficiency-dependent but does not tightly associate with the target language.

Also, this study applied the idea of the residence binding factor to the Chinese auditory task to explore how orthographic similarities affect spoken word recognition when participants live in the L1 or L2 spoken environment. We conducted a comparison of the orthographic effect in Chinese spoken word recognition between intermediate to highly proficient Chinese-German bilinguals living in Germany and a group of Chinese monolinguals living in China. The findings of this study revealed a clear influence of orthographic similarity only in Chinese speakers who have been living in China, but not in Germany. These results contribute to a deeper understanding of Chinese auditory word processing, suggesting that the associations between exposure to printed words and the effect of orthography on spoken word recognition are tightly linked, which means that the acquisition of an alphabetic L2 and long-term immersion in an alphabetic L2-spoken environment may improve the sensitivity to phonological information while diminishing the orthographic effect during L1 spoken word recognition for speakers with a logographic L1 background.

These studies provided insights into the influences of phonological and orthographic information on spoken word recognition among German native speakers and Chinese learners of German with two distinct language proficiency levels, thereby contributing to a better understanding of word spoken recognition in an alphabetic and logographic writing system. It emphasizes that we should consider the writing system of each language, the proficiency level, and exposure to printed forms when investigating orthographic effects during auditory processing.

4.4 Limitations and directions for future research

The limitations of the current study are noteworthy.

One notable limitation pertains to the validation of participants' reported language proficiency. Despite employing a language history and proficiency questionnaire, along with the LexTALE test, to assess the second language abilities of the participants, there remains some uncertainty regarding the proficiency levels of certain participants whose test scores fall close to the threshold between intermediate and high proficiency. Furthermore, we did not specifically examine the participants' writing and listening abilities in L2, which could be relevant factors influencing the experimental results.

Secondly, certain Chinese participants who took part in Experiment 2 reported challenges in memorizing the new vocabulary presented before the experiment. They required additional review of the online word list before taking the experiment. This difficulty in recalling specific words during the task might have led to random responses during the experiment, potentially affecting both response time and accuracy rate.

Thirdly, Chinese participants residing in China were engaged in the experiment only online, which presented challenges in maintaining strict control over the testing environment, as would have been achievable in a laboratory setting for behavioral studies. Therefore, they might not have been focused throughout the process of the whole experiment or finished the experimental blocks without breaks, which could have potentially induced fatigue effects.

Moreover, even though the present study controlled for some linguistic factors of the experimental materials, it still lacks control over non-native lexical frequencies, which tend to differ from those of native speakers due to the smaller mental lexicons and the reduced input and exposure to L2 words experienced by L2 speakers (Diependaele, Lemhöfer, & Brysbaert, 2013). Therefore, the used stimuli matched the word frequency from native corpora, but maybe do not suit experiments for L2 participants.

In future studies, it would be valuable to explore the role of orthography in L2 auditory tasks when bilinguals' L1 and L2 are both non-alphabetic languages, such as Chinese-Japanese. Japanese employs Chinese characters (*kanji*) to convey semantic meanings, but the pronunciation differs from Chinese. Thus, the orthographic transfer may occur when Chinese–Japanese bilinguals with different L2 proficiency levels perform an L2 auditory task, and potential orthographic influences during L2 spoken word recognition could be expected.

Furthermore, to investigate whether the closeness of languages affects the activation of orthography on L2 spoken word recognition, we could conduct studies on Spanish or Dutch learners of German compared to Chinese learners of German using the same paradigm. Given that Spanish and Dutch have shallow orthographic depths, similar to German, and Chinese language has very deep orthography, this comparison could be very straightforward. If there is still an orthographic effect on spoken word recognition in Spanish or Dutch L2 groups, the influence of orthography can be a feature of L2 processing, not be limited to the characteristics of L1 of participants.

To gain deeper insights into how orthographic similarity is processed in the brain, employing alternative methodologies such as EEG and fMRI in future research would be advantageous. Additionally, incorporating various types of tasks could further investigate whether orthography is automatically activated or just depends on the specific task at hand.

Reference

Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Language Learning*, *53*(2), 207–231.

Allopenna, PD., Magnuson, JS., & Tanenhaus, MK. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*. 38:419–439.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review*, *4*(4), 439–461.

Baayen, R. H. (2008). *Analyzing Linguistic Data*. Cambridge University Press.

Ballatore, A., Bertolotto, M., & Wilson, D. C. (2014). An evaluative baseline for geo-semantic relatedness and similarity. *GeoInformatica*, *18*(4), 747–767.

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.

Bartón, K. (2022). *Package ' MuMIn .'* (Version 1.47.1) [Software Package].

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in english. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(11), 1835–1842.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1).

Best, C. T., & Tyler, M. D. (2007). *Nonnative and second-language speech perception*. *June 2018*, 13–34.

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer*. (Version 6.1.35) [Computer Software]. https://www.fon.hum.uva.nl/praat/

Bridges, D., Pitiot, A., Macaskill, M. R., & Peirce, J. W. (2020). The timing mega-study : comparing a range of experiment generators , both lab-based and online.

PsyArXiv.

Chang, Y. N., Welbourne, S., & Lee, C. Y. (2016). Exploring orthographic neighborhood size effects in a computational model of Chinese character naming. *Cognitive Psychology*, *91*, 1–23.

Chen, H. C., Vaid, J., & Wu, J. T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes*, *24*(7–8), 967–982.

Chen, W. F., Chao, P. C., Chang, Y. N., Hsu, C. H., & Lee, C. Y. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Language*, *157–158*, 51–62.

Chen, X., Xu, F., Nguyen, T. K., Hong, G., & Wang, Y. (2010). Effects of Cross-Language Transfer on First-Language Phonological Awareness and Literacy Skills in Chinese Children Receiving English Instruction. *Journal of Educational Psychology*, *102*(3), 712–728.

Chikamatsu, N. (1996). The effects of L1 orthography on L2 word recognition: A study of American and Chinese learners of Japanese. *Studies in Second Language Acquisition*, *18*(4), 403–432.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud. *Psychological Review*, *100*(4), 589–608.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI.* (pp. 535–555). Erlbaum.

Coltheart, M., & Rastle, K. (1994). Serial Processing in Reading Aloud: Evidence for Dual-Route Models of Reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1197–1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Cook, V. (1997). The consequences of bilingualism for cognitive processing. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, 279–299.

Cook, V. (2002). Background to the L2 User. In V. Cook (Ed.), *Portraits of the L2 User* (pp. 1–28).

Cook, V. (2003). Introduction: The changing L1 in the L2 user's mind. In *Effects of the Second Language on the First* (pp. 1–18)

Cook, V. (2006). Interlanguage, Multi-Competence and the Problem of the "Second" Language. 1000–1014.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin and Review*, *22*(5), 1216–1234.

Cutler, A., & Davis, C. (2012). An orthographic effect in phoneme processing, and its limitations. *Frontiers in Psychology*, *3*(FEB), 1–7.

Cutler, A., Mehler, J., & Norris, D. (1987). Phoneme Identification and the Lexicon in progress , and it is designed to reflect processing at the phonemic level . developed by Foss (1969) -listen to. *Applied Psychology*, *177*, 141–177.

Cutler, A., Treiman, R., & van Ooijen, B. (2010). Strategic Deployment of Orthographic Knowledge in Phoneme Detection. *Language and Speech*, *53*(3), 307–320.

Damian, M. F., & Bowers, J. S. (2010). Orthographic effects in rhyme monitoring tasks: Are they automatic? *European Journal of Cognitive Psychology*, *22*(1), 106–116.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: a program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*(4), 665–671.

Dich, N. (2011). Individual differences in the size of orthographic effects in spoken word recognition: The role of listeners' orthographic skills. *Applied Psycholinguistics*, *32*(1), 169–186.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*(5), 843–863.

Diependaele, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the bimodal interactive activation model. In *European Journal of Cognitive Psychology* (Vol.

22, Issue 5, pp. 764–778).

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284–301.

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In *Localist connectionist approaches to human cognition.* (pp. 189–225). Lawrence Erlbaum Associates Publishers.

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197.

Dornbusch, T. (2012). *Orthographic influences on L2 auditory word processing*.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin and Review*, *15*(4), 850–855.

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading*, *18*(1), 5–21.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V. H. p., Polyzoe, N., Tsipa, M., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*(4), 438–468.

Elman, J. L., & McClelland, J. L. (1984). *Speech Perception as a Cognitive Process: The Interactive Activation Model* (pp. 337–374).

Forster, K. I. (1976). New Approaches to Language Mechanisms. In *A Collection of Psycholinguistic Studies* (pp. 257–287). North-Holland Publishing Company.

Forster, K. I. (1979). Levels of processing and the structure of the language processes. *Sentence Processing: Psycholinguistic Studies Presented to Merril Garret*, *Lawrence Erlbaum Associates*, 27–84.

Framework, C. E. (2018). Companion Volume. In *New Cambridge Modern History* (Vol. 13).

Frauenfelder, U. H. ., & Tyler, L. K. (1987). *The process of spoken word recognition: an introduction*. *25*, 1–20.

Frost, R. (2008). Orthographic Systems and Skilled Word Recognition Processes in Reading. In *The Science of Reading: A Handbook* (pp. 272–295).

Goswami, U. (2010). A psycholinguistic grain size view of reading acquisition across languages. *Reading and Dyslexia in Different Orthographies*, 23–42.

Goswami, U., Ziegler, J. C., & Richardson, U. (2005). The effects of spelling consistency on phonological awareness: A comparison of English and German. *Journal of Experimental Child Psychology*, *92*(4), 345–365.

Gottardo, A., Siegel, L. S., Yan, B., & Wade-Woolley, L. (2001). Factors related to English reading performance in children with Chinese as a first language: More evidence of cross-language transfer of phonological processing. *Journal of Educational Psychology*, *93*(3), 530–542.

Grainger, J., & Ferrand, L. (1994). Phonology and Orthography in Visual Word Recognition: Effects of Masked Homophone Primes. *Journal of Memory and Language*, *33*(2), 218–233.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, *35*(5), 623–647.

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

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(6), 981–998.

Grainger, J., Spinelli, E., Farioli, F., Diependaele, K., & Ferrand, L. (2003). Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(6), 1256–1269.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm.

Perception & Psychophysics, *28*(4), 267–283.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15.

Grosjean, F. (2008). Studying Bilinguals: Methodological and Conceptual Issues 1. *The Handbook of Bilingualism*, *1*(May 2010), 32–63.

Grosjean, F., & Byers-Heinlein, K. (2018). Speech Perception and Comprehension. In *The Listening Bilingual* (Issue August).

Hadj Taieb, M. A., Zesch, T., & Ben Aouicha, M. (2020). A survey of semantic relatedness evaluation datasets and procedures. In *Artificial Intelligence Review* (Vol. 53, Issue 6). Springer Netherlands.

Hao, Y.-C., & Yang, C.-L. (2018). *The role of orthography in L2 segment and tone encoding by learners at different proficiency levels*. 247–251.

Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB - Eine lexikalische datenbank für die psychologische und linguistische forschung. *Psychologische Rundschau*, *62*(1), 10–20.

Houkema, D. (2001). *The coherent perception of speech within Cognitive Science*. Master's Thesis / Essay, Artificial Intelligence.

Huang, H. W., Lee, C. Y., Tsai, J. L., Lee, C. L., Hung, D. L., & Tzeng, O. J. L. (2006). Orthographic neighborhood effects in reading Chinese two-character words. *NeuroReport*, *17*(10), 1061–1065.

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.

Joanisse, M. F., & McClelland, J. L. (2015). Connectionist perspectives on language learning, representation and processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 235–247.

Katz, L., & Frost, R. (1992). The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. *Advances in Psychology*, *94*(C), 67–84.

Kilpatrick, D. A. (2015). *Essentials of Assessing, Preventing, and Overcoming Reading*

Difficulties. Wiley.

Laufer, B. (2003). The influence of L2 on L1 collocational knowledge and on L1 Lexicaldiversity in free written expression. In V. J. Cook (Ed.), *Effects of the second language on the first* (pp. 19-31). Clevedon, UK: Multilingual Matters.

Lee, C. Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, *50*(1), 101–123.

Leiner, D. J. (2019). *SoSci survey*. (Version 3.1. 06)[Computer Software].

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343.

Lenth, R. ., Buerkner, P. ., Herve, M. ., Love, J. ., Riebl, H. ., & Singmann, H. (2021). *emmeans: Estimated marginal means*. (Version 1.6.2-1). https://github.com/rvlenth/emmeans

Li, C. ., & Thompson, S. . (1981). *Mandarin Chinese: A Functional Reference Grammar*. CA: University of California Press.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism*, *17*(3), 673–680.

Li, Q. L., Bi, H. Y., Wei, T. Q., & Chen, B. G. (2011). Orthographic neighborhood size effect in Chinese character naming: Orthographic and phonological activations. *Acta Psychologica*, *136*(1), 35–41.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. . . (1980). Orthography and the beginning reader. In *Orthography, Reading and Dyslexia* (pp. 137–153).

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, *39*(3), 155–158.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). Clearpond: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, *7*(8).

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940.

Marslen-wilson, W. (1984). *Function and Process in Spoken Word Recognition*. *January 1984*.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71.

Martin, K. I. (2017). The impact of L1 writing system on ESL knowledge of vowel and consonant spellings. *Reading and Writing*, *30*(2), 279–298.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407.

Mitsugi, S. (2018). Proficiency influences orthographic activations during L2 spoken-word recognition. *International Journal of Bilingualism*, *22*(2), 199–214.

Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language*, *85*, 116–134.

Mok, P. P. K., Lee, A., Li, J. J., & Xu, R. B. (2018). Orthographic effects on the perception and production of L2 mandarin tones. *Speech Communication*, *101*(May), 1–10.

Muneaux, M., & Ziegler, J. C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, *19*(5), 641–660.

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134).

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234.

Paradis, Johanne, Genesee, Fred, & Crago, Martha (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning, 2nd Edition.* Baltimore, MD: Brookes.

Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language and Cognitive Processes*, *22*(5), 700–726.

Pattamadilok, C., Perre, L., Dufau, S., & Ziegler, J. C. (2009). On-line orthographic influences on spoken language in a semantic task. *Journal of Cognitive Neuroscience*, *21*(1), 169–179.

Patterson, K., & Coltheart, V. (1987). Phonological processes in reading: A tutorial review. In *Attention and performance 12: The psychology of reading.* (pp. 421–447). Lawrence Erlbaum Associates, Inc.

Peereman, R., Dufour, S., & Burt, J. S. (2009). Orthographic influences in spoken word recognition: The consistency effect in semantic and gender categorization tasks. In *Psychonomic Bulletin and Review*.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Perfetti, C. A., & Liu, Y. (2005). Orthography to phonology and meaning: Comparisons across and within writing systems. *Reading and Writing*, *18*(3), 193–210.

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.

Perre, L., Midgley, K., & Ziegler, J. C. (2009). When beef primes reef more than leaf: Orthographic information affects phonological priming in spoken word recognition. *Psychophysiology*, *46*(4), 739–746.

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research*, *1275*, 73–80.

Pytlyk, C. (2017). Are orthographic effects language specific? the influence of second language orthography on second language phoneme awareness. *Applied Psycholinguistics*, *38*(2), 233–262.

Qu, Q., Cui, Z., & Damian, M. F. (2018). Orthographic effects in second-language spoken-word recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(8), 1325–1332.

Qu, Q., & Damian, M. F. (2017). Orthographic effects in spoken word recognition: Evidence from Chinese. *Psychonomic Bulletin and Review*, *24*(3), 901–906.

Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography Influences the Perception and Production of Speech. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(6), 1588–1594.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60–94.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–

430.

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin and Review*, *22*(6), 1614–1629.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology*, *5*(FEB), 1–16.

Schröter, P., & Schroeder, S. (2018). Differences in visual word recognition between L1 and L2 speakers: The impact of length, frequency, and orthographic neighborhood size in German children. *Studies in Second Language Acquisition*, *40*(2), 319–339.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*(1), 1–30.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(6), 546–554.

Seymour, P. H. K., Aro, M., Erskine, J. M., Wimmer, H., Leybaert, J., Elbro, C., Lyytinen, H., Gombert, J. E., Le Normand, M. T., Schneider, W., Porpodas, C., Ragnarsdottir, H., Tressoldi, P., Vio, C., De Groot, A., Licht, R., Iønnessen, F. E., Castro, S. L., Cary, L., … Olofsson, Å. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Shen, W., Hyönä, J., Wang, Y., Hou, M., & Zhao, J. (2021). The role of tonal information during spoken-word recognition in Chinese: Evidence from a printed-word eye-tracking study. *Memory and Cognition*, *49*(1), 181–192.

Shi, L.-F., & Morozova, N. (2012). Understanding native Russian listeners' errors on an English word recognition test: Model-based analysis of phoneme confusion. *International Journal of Audiology*, *51*(8), 597–605.

Siakaluk, P. D., Sears, C. R., & Lupker, S. J. (2002). Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 661–681.

Sun-Alperin, M. K., & Wang, M. (2011). Cross-language transfer of phonological and orthographic processing skills from Spanish L1 to English L2. *Reading and Writing*, *24*(5), 591–614.

Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (CLD): A large-scale lexical database for simplified Mandarin Chinese. *Behavior Research Methods*, *50*(6), 2606–2629.

Taft, M. (2006). Orthographically influenced abstract phonological representation: evidence from non-rhotic speakers. *Journal of Psycholinguistic Research*, *35*(1), 67–78.

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, *24*(3), 320–335.

Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, *22*(3), 259–282.

Taft, M., & Zhu, X. (1995). The representation of bound morphemes in the lexicon: A Chinese study. In *Morphological aspects of language processing.* (pp. 293–316). Lawrence Erlbaum Associates, Inc.

Taft, M., & Zhu, X. (1997). Using masked priming to examine lexical storage of Chinese compound words. In H.-C. Chen (Ed.), *The cognitive processing of Chinese and related Asian languages* (pp. 233–241). Chinese University Press.

Taft, M., Zhu, X., & Peng, D. (1999). Positional Specificity of Radicals in Chinese Character Recognition. *Journal of Memory and Language*, *40*(4), 498–519.

Tan, L. H., & Perfetti, C. A. (1999). Phonological Activation in Visual Identification of Chinese Two-Character Words. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(2), 382–393.

Tanenhaus, M. K., Carlson, G. N., & Seidenberg, M. S. (1985). Do listeners compute linguistic representations? In *Natural Language Parsing* (pp. 359–408). Cambridge University Press.

Tong, X., McBride, C., & Burnham, D. (2014). Cues for Lexical Tone Perception in Children: Acoustic Correlates and Phonetic Context Effects. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1589–1605.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.

Türk, S., & Domahs, U. (2022). Orthographic influences on spoken word recognition in bilinguals are dependent on the orthographic depth of the target language not the native language. *Brain and Language*, *235*(October).

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, *36*(5), 417–427.

Veivo, O., & Järvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism*, *16*(4), 864–883.

Veivo, O., Järvikivi, J., Porretta, V., & Hyönä, J. (2016). Orthographic activation in L2 spoken word recognition depends on proficiency: Evidence from eye-tracking. *Frontiers in Psychology*, *7*(JUL).

Ventura, P., Morais, J., & Kolinsky, R. (2007). The development of the orthographic consistency effect in speech recognition: From sublexical to lexical involvement. *Cognition*, *105*(3), 547–576.

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, *19*(1), 57–95.

Vokic, G. (2011). When alphabets collide: Alphabetic first-language speakers' approach to speech production in an alphabetic second language. *Second Language Research*, *27*(3), 391–417.

Wang, W., Li, X., Ning, N., & Zhang, J. X. (2012). The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters*, *516*(1), 67–71.

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 387–401.

Weekes, B. S., Chen, M. J., & Lin, Y.-B. (1998). Differential effects of phonological priming on Chinese character recognition. In *An Interdisciplinary Journal* (Vol. 10, pp. 47–68).

Wu, J.-T., & Chen, H.-C. (2000). Evaluating semantic priming and homophonic priming in recognition and naming of Chinese characters. [Evaluating semantic priming and homophonic priming in recognition and naming of Chinese characters.]. *Chinese Journal of Psychology*, *42*(1), 65–86.

Wu, Y. J., & Thierry, G. (2010). Chinese-english bilinguals reading english hear chinese. *Journal of Neuroscience*, *30*(22), 7646–7651.

Yelland, G. W., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics*, *14*(4), 423–444.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761–1770.

Zhou, Y. G. (1978). To what degree are the 'phonetic radicals' of present-day Chinese characters still phonetic radicals? *Zhongguo Yuwen [Chinese Language Journal]*, *146*, 172–177.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review*, *5*(4), 683–689.

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory and Cognition*, *32*(5), 732–741.

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

Ziegler, J. C., & Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: A developmental study. *Psychonomic Bulletin and Review*, *14*(1), 75–80.

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*(4), 779–793.

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to german. *European Journal of Cognitive Psychology*, *12*(3), 413–430.

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology Matters: The Phonological Frequency Effect in Written Chinese. *Psychological Science*, *11*(3), 234–238.

Zou, L., Desroches, A. S., Liu, Y., Xia, Z., & Shu, H. (2012). Orthographic facilitation in Chinese spoken word recognition: An ERP study. *Brain and Language*, *123*(3), 164–173.

Appendix A (Stimuli for experiment 1+2)

Critical Stimuli

Nr.	Target	O+P+	O-P+	O-P-
1	Feier	Leier	Reiher	Höhle
2	Paar	Haar	Zar	Berg
3	Bahn	Wahn	Kran	Brot
4	Spur	Kur	Tour	Leim
5	Not	Rot	Boot	Stein
6	Tor	Chor	Moor	Staub
7	Kuss	Nuss	Bus	Watt
8	Bein	Schein	Hain	Null
9	Mai	Hai	Brei	Wahl
10	Kreis	Reis	Mais	Schuh
11	Fuchs	Luchs	Jux	Seil
12	Spitze	Hitze	Skizze	Drucker
13	Sieger	Flieger	Tiger	Teller
14	Nächte	Mächte	Rechte	Lieder
15	Strähne	Mähne	Däne	Zwiebel
16	Tee	See	Reh	Lob
17	Schal	Gral	Saal	Wind
18	Sohn	Lohn	Ton	Hut
19	Messe	Kresse	Nässe	Wetter
20	Dächer	Fächer	Becher	Brücken
21	Wecker	Stecker	Bäcker	Bogen
22	Scherz	Herz	März	Saft
23	Krieger	Flieger	Tiger	Tasse
24	Damen	Samen	Rahmen	Reste
25	Lord	Nord	Hort	Stift

Nr.	Target	O+P+	O-P+	O-P-
26	Fluss	Nuss	Bus	Post
27	Kind	Rind	Sprint	Pack
28	Park	Mark	Sarg	Hass
29	Mut	Hut	Sud	Burg
30	Haus	Maus	Strauß	Druck
31	Fuß	Ruß	Mus	Stern
32	Kleid	Neid	Streit	Stuhl
33	Pferd	Herd	Schwert	Buch
34	Spaß	Fraß	Gas	Fisch
35	Zahn	Wahn	Schwan	Lied
36	Krone	Zone	Bohne	Kater
37	Wert	Schwert	Herd	Ziel
38	Fleiß	Weiß	Preis	Bach
39	Schar	Bar	Haar	Lack
40	Mord	Nord	Sport	Rand
41	hören	stören	röhren	raten
42	zählen	wählen	quälen	pflegen
43	Leid	Neid	Maid	Holz
44	läuten	häuten	deuten	merken
45	wenden	spenden	schänden	fürchten
46	Gelder	Felder	Wälder	Hefte
47	Bart	Start	Fahrt	Bild
48	wohnen	lohnen	schonen	küssen
49	denken	schwenken	kränken	reisen
50	beugen	zeugen	säugen	tauschen
51	Floh	Stroh	Zoo	Band
52	Krise	Brise	Riese	Maler
53	wellen	bellen	fällen	werfen

Nr.	Target	O+P+	O-P+	O-P-
54	Wachs	Lachs	Fax	Schrank
55	Wiese	Fliese	Brise	Jacke
56	Tal	Qual	Zahl	Schiff
57	Schnee	Klee	Dreh	Huhn
58	Menge	Enge	Länge	Hose
59	Bühne	Sühne	Düne	Kette
60	Rat	Tat	Pfad	Mond
61	Säule	Fäule	Keule	Vogel
62	Kohle	Sohle	Mole	Helfer
63	Hehl	Mehl	Gel	Tuch
64	Bug	Zug	Spuk	Traum
65	Job	Lob	Stopp	Blatt
66	Wut	Hut	Sud	Topf
67	Stahl	Pfahl	Qual	Baum
68	führen	rühren	spüren	tanzen
69	Nähte	Drähte	Räte	Zwiebel
70	Seiher	Reiher	Leier	Käufer
71	Schoß	Floß	Moos	Plan
72	paaren	haaren	fahren	kochen
73	Maat	Saat	Tat	Stern
74	Lot	Rot	Boot	Schirm
75	Laus	Maus	Strauß	Buch
76	krönen	frönen	dröhnen	fliegen
77	Kram	Scham	Rahm	Kern
78	Strahl	Pfahl	Gral	Hand
79	Lehne	Sehne	Vene	Zucker
80	Taxen	Praxen	Achsen	Blätter
81	Kloß	Stoß	Moos	Bad

Nr.	Target	O+P+	O-P+	O-P-
82	Berge	Zwerge	Särge	Münzen
83	Beule	Keule	Fäule	Magen
84	Geier	Leier	Flyer	Vase
85	Fresser	Messer	Fässer	Karten
86	Flug	Krug	Spuk	Stamm
87	Fieber	Schieber	Biber	Lampe
88	fahl	kahl	schmal	klein
89	dämmen	kämmen	hemmen	bieten
90	Bänder	Ränder	Sender	Katzen
91	Mieter	Bieter	Liter	Mantel

Fillers – semantically related

Nr.	Target	S+O-P-
1	Leiter	Führer
2	Zucker	Pfeffer
3	Butter	Käse
4	Seife	Wäsche
5	Wurst	Fleisch
6	Baum	Holz
7	Liebe	Rose
8	Bett	Schlaf
9	Nase	Schnupfen
10	Herd	Topf
11	Wolf	Bär
12	Auto	Taxi
13	Wein	Bier
14	Tuch	Stoff
15	Gabel	Messer

Nr.	Target	S+O-P-
16	Uhr	Zeit
17	Auge	Brille
18	Taube	Vogel
19	Wunsch	Ziel
20	Vase	Blume
21	Schule	Lehrer
22	Kabel	Stecker
23	Kuchen	Sahne
24	Oper	Theater
25	Boss	Chef
26	Obst	Frucht
27	Burg	Schloss
28	Maul	Mund
29	Kuhl	Kalb
30	Schwein	Speck
31	Tante	Nichte
32	Dorf	Ort
33	Macht	Kraft
34	Schule	Note
35	Tisch	Stuhl
36	Knie	Arm
37	Raub	Dieb
38	Muskel	Ader
39	Ei	Huhn
40	Gleis	Zug
41	Hitze	Kühle
42	Tanz	Lied
43	Schiff	See

Nr.	Target	S+O-P-
44	mieten	kaufen
45	Bau	Dach
46	Brot	Mehl
47	Wolle	Seide
48	Licht	Glanz
49	Angst	Furcht
50	Wald	Wolf
51	Gras	Tau
52	Winter	Sommer
53	Ziel	Zweck
54	Brücke	Ufer
55	Schloss	Prinz
56	Mond	Stern
57	Aal	Fisch
58	Jagd	Hirsch
59	Spritze	Patient
60	Korn	Sand
61	Körper	Seele
62	schützen	hüten
63	Geist	Tod
64	sorgen	kümmern
65	Birne	Lampe
66	parken	halten
67	Geld	Markt
68	Wolke	Nebel
69	Kunst	Bild
70	Ampel	Schild
71	Nagel	Finger

Nr.	Target	S+O-P-
72	Stoff	Hemd
73	Glanz	Gold
74	Pein	Qual
75	laufen	joggen
76	Lippe	Zunge
77	Spiel	Sport
78	Schicht	Dienst
79	Deckel	Kiste
80	Antwort	Fehler
81	Ski	Helm
82	Tasche	Reise
83	Trost	Schmerz
84	Schrei	Ruf
85	Klavier	Gesang
86	Kerze	Flamme
87	Knopf	Hemd
88	Stil	Mode
89	Neid	Hass
90	Hose	Jacke
91	Hund	Napf
92	Klima	Wetter
93	Nest	Zweig
94	Schaf	Lamm
95	Zeile	Spalte
96	Phase	Stufe
97	Frosch	Sprung
98	Rauch	Qualm
99	Mond	Schein

Nr.	Target	S+O-P-
100	Mango	Apfel
101	nass	feucht
102	Metal	Eisen
103	Urlaub	Arbeit
104	Pfote	Katze
105	Wind	Sturm
106	Pech	Klee
107	Retter	Hilfe
108	Blei	Gift
109	Perle	Muschel
110	Bluse	Kette
111	Haar	Kamm
112	Ring	Ehe
113	Kuh	Milch
114	Auto	Fahrrad
115	Maus	Tier
116	Blut	Kampf
117	Farbe	Maler
118	Rauch	Dampf
119	Laub	Herbst
120	Becher	Wasser
121	Schrank	Tür
122	Auge	Wimper
123	Pfeffer	Würze
124	Neffe	Junge
125	Regen	Pfütze
126	Zucker	Hefe
127	Frühling	Winter

Nr.	Target	S+O-P-
128	Garn	Schnur
129	Moschee	Kirche
130	Rumpf	Bauch
131	Gurt	Band
132	Bulle	Ochse
133	Strom	Gas
134	Hieb	Schlag
135	Hecke	Garten
136	Tasche	Henkel
137	Wein	Rausch
138	Hase	Möhre
139	Glück	Herz
140	Härte	Eisen
141	Sonne	Urlaub
142	Beil	Griff
143	Druck	Kraft
144	Jagd	Pfeil
145	Damm	Stau
146	Klausur	Prüfung
147	Schirm	Schutz
148	Schuh	Strumpf
149	Stroh	Feld
150	Wange	Kiefer

Filler — semantically related

Nr.	Target	S+O-P-
1	Luft	Rock
2	Block	Lauch

Nr.	Target	S+O-P-
3	Eis	Land
4	Koch	Platz
5	backen	mieten
6	Raum	Gunst
7	Dienst	Saft
8	Biene	Kellner
9	Kurs	Lachs
10	Kasse	Sofa
11	Stern	Mais
12	Stamm	Flur
13	Milch	Wand
14	Gans	Boot
15	Bank	Jagd
16	Stuhl	Saft
17	Tisch	Frosch
18	Reis	Licht
19	Blatt	Arm
20	schicken	kochen
21	Farbe	Schiene
22	Küche	Parfum
23	Schere	Butter
24	Rock	Sieb
25	Korb	Fell
26	Eis	Gut
27	Rohr	Halt
28	Herz	Stab
29	Zucht	Sack
30	Dose	Kachel

Nr.	Target	S+O-P-
31	Schach	Müll
32	Wild	Skript
33	Fell	Norm
34	Lied	Fass
35	Bär	Rauch
36	Watt	Gans
37	Obst	Chor
38	Blick	Farn
39	Angst	Rand
40	Luft	Takt
41	Bast	Wind
42	Hort	Rad
43	Wunsch	Saat
44	meinen	fügen
45	Team	Rauch
46	Gleis	Burg
47	Leine	Woche
48	Stern	Zorn
49	Bauch	Schuh
50	Herd	Baum
51	Seil	Wok
52	Schild	Knall
53	teilen	malen
54	Netz	Pfund
55	zeigen	fallen
56	Hort	Klang
57	Bett	Pilz
58	Loch	Reck

Nr.	Target	S+O-P-
59	Herbst	Schrei
60	Farbe	Woche

Appendix B (Stimuli for experiment 3+4)

Critical Stimuli

Nr.	Target	O+P-	O-P+	O-P-
1	伯父, /bo2fu4/, uncle	柏树, /bai3shu4/, cypress	博士, /bo2shi4/, doctor	晚饭, /wan3fan4/, dinner
2	被子, /bei4zi/, quilt	破裂, /po4lie4/, break	贝壳, /bei4ke2/, shell	酸奶, /suan1nai3/, yogurt
3	童年, /tong2nian2/, childhood	撞击, /zhuang4ji1/, crash	铜牌, /tong2pai2/, bronze medal	破旧, /po4jiu4/, old and shabby
4	诉说, /su4shuo1/, tell	拆迁, /chai1qian1/, demolition	速度, /su4du4/, speed	窗户, /chuang1hu4/, window
5	祝福, /zhu4fu2/, bless	况且,/kuang4qie3/, besides	助手, /zhu4shou3/, assistant	周围, /zhou1wei2/, around
6	凋零, /diao1ling2/, withered	绸缎, /chou2duan4/, silk fabrics	刁民, /diao1min2/, unruly people	脑袋, /nao3dai4/, head
7	促销, /cu4xiao1/, sales promotion	捉弄, /zhuo1nong4/, tease	猝死, /cu4ci3/, sudden death	删除, /shan1chu2/, delete
8	诗词, /shi1ci2/, poetry	待命, /dai4ming4/, standby	师傅,/shi1fu4/, master	玻璃, /bo1li2/, glass
9	待遇, /dai4yu4/, treatment	持续, /chi2xu4/, continue	袋子,/dai4zi/, bag	蚊子, /wen2zi/, mosquito
10	科学, /ke1xue2/, science	料理, /liao4li3/, cuisine	颗粒, /ke1li4/, particle	面具, /mian4ju4/, mask
11	碍事, /ai4shi4/, hinder	得意, /de2yi4/, gloat	爱情, /ai4qing2/, love	酒吧, /jiu3ba1/, bar

12	挣扎, /zheng1zha2/, struggle	净化, /jing4hua4/, purify	蒸汽, /zheng2qi4/, steam	春天, /chun1tian1/, spring
13	帆船, /fan1chuan2/, sailing boat	巩固, /gong3gu4/, strengthen	番茄, /fan1qie2/, tomato	插座, /cha1zhuo4/, plug
14	奋斗, /fen4dou4/, strive	备用, /bei4yong4/, reserve	粪便, /fen4bian4/, shit	唱歌, /chang4ge1/, sing
15	刮痧, /gua1sha1/, scraping	敌人, /di2ren2/, enemy	瓜子, /gua1zi/, melon seeds	阳光, /yang2guang1/, sunshine
16	归来, /gui1lai2/, return	扫帚, /sao4zhou/, broom	闺房, /gui1fang2/, boudoir	饼干, /bing3gan1/, biscuit
17	拒绝, /ju4jue2/, refuse	柜子, /gui4zi/, cupboard	聚会, /ju4hui4/, meeting	宠物, /chong3wu4/, pet
18	晦气, /hui4qi4/, unlucky	海豚, /hai3tun2/, dolphin	绘画, /hui4hua4/, painting	鼻子, /bi2zi/, nose
19	往事, /wang3shi4/, past events	住房, /zhu4fang2/, housing	网络, /wang3luo4/, Internet	压缩, /ya1shuo1/, compress
20	激动, /ji1dong4/, excite	邀请, /yao1qing3/, invite	鸡蛋, /ji1dan4/, egg	列车, /lie4che1/, train
21	脉搏, /mai4mo2/, pulse	泳衣, /yong3yi1/, swimming suit	麦子, /mai4zi/, wheat	语言, /yu3yan2/, language
22	惊动, /jing1dong4/, startle	掠夺, / lüe4duo2/, rob	精华, /jing1hua2/, essence	轮船, /lun2chuan2/, ship
23	韭菜, /jiu3cai4/, leeks	非洲, /fei1zhou1/, Africa	酒瓶, /jiu3ping2/, winebottle	信任, /xin4ren4/, trust
24	袖子, /xiu4zi/, sleeve	油田, /you2tian2/, oil field	秀才, /xiu4cai2/, scholar	暖气, /nuan3qi4/, heater

25	治疗, /zhi4liao2/, cure	始终, /shi3zhong2/, always	志向, /zhi4xiang4/, ambition	钥匙, /yao4shi/, key
26	凉鞋, /liang2xie2/, sandal	惊喜, /jing1xi3/, surprise	粮食, /liang2shi2/, food	忘记, /wang4ji4/, forget
27	聊天, /liao2tian1/, chat	柳树, /liu3shu4/, willow	辽阔, /liao2kuo4/, broad	香水, /xiang1shui3/, perfume
28	流感, /liu2gan3/, flu	梳子, /shu1zi/, comb	留学, /liu2xue2/, study abroad	超市, /chao1shi4/, supermarket
29	路口, /lu4kou3/, intersection	格子, /ge2zi/, cell	录音, /lu4yin1/, recording	简单, /jian3dan1/, easy
30	优秀, /you1xiu4/, outstanding	扰乱, /rao3luan4/, harass	悠闲, /you2xian2/, leisure	帽子, /mao4zi/, hat
31	读书, /du2shu1/, read	卖弄, /mai4nong4/, show off	毒品, /du2pin3/, drugs	流浪, /liu2lang4/, roam
32	驱逐, /qu1zhu2/, banish	呕吐, /ou3tu4/, vomit	趋势, /qu1shi4/, trend	雨水, /yu3shui3/, rainwater
33	幻想, /huan4xiang3/, illusion	幼儿, /you4er2/, infant	换钱, /huan4qian2/, money exchange	桔子, /ju2zi/, orange
34	脾气, /pi2qi4/, temperament	牌子, /pai2zi/, sign	皮球, /pi2qiu2/, ball	鹦鹉, /ying1wu3/, parrot
35	披肩, /pi1jian1/, cape	波浪, /bo1lang4/, wave	劈叉, /pi1cha4/, do the splits	厨房, /chu2fang2/, kitchen
36	评价, /ping2jia4/, evaluate	抨击, /peng1ji1/, attack	瓶子, /ping2zi/ bottle	承诺, /cheng2nuo4/, promise
37	彻底, /che4di3/, thorough	沏茶, /qi4cha2/, make tea	撤退, /che4tui4/, retreat	善良, /shan4liang2/, kind

38	腔调, /qiang1diao4/, tune	控制, /kong4zhi4/, control	枪只, /qiang2zhi1/, gun	核桃, /he2tao2/, walnut
39	羽毛, /yu3mao2/, feather	习惯, /xi2guan4/, habit	宇宙, /yu3zhou4/, universe	玫瑰, /mei2gui/, rose
40	沙漠, /sha1mo4/, desert	炒股, /chao3gu3/, investing in stocks	杀人, /sha1ren2/ murder	钢笔, /gang1bi3/, pen
41	银行, /yin2hang2/, bank	艰苦, /jian1ku3/, arduous	吟唱, /yin2chang4/, sing	声音, /sheng1yin1/, sound
42	绒毛, /rong2mao2/, fine hair	贼船, /zei2chuan2/, pirate ship	荣誉, /rong2yu4/, honor	手机, /shou3ji1/, mobil phone
43	维护, /wei2hu4/, maintenance	难处, /nan2chu4/, difficulty	围绕, /wei2rao4/, around	作品, /zuo4pin3/, work
44	计划, /ji4hua4/, plan	针对, /zhen1dui4/, aim at	寄信, /ji4xin4/, send	桃子, /tao2zi/, peach
45	撑腰, /cheng1yao1/, back up	掌握, /zhang3wo4/, master	称赞, /cheng1zan4/, compliment	答案, /da2an4/, answer
46	限制, /xian4zhi4/, limit	根本, /gen1ben3/, fundamental	线索, /xian4suo3/, clue	闹钟, /nao4zhong1/, alarm clock
47	愉悦, /yu2yue4/, joyful	偷税, /tou1shui4/, evade taxes	鱼翅, /yu2chi4/ fin	考试, /kao3shi4/, examination
48	哽咽, /geng3ye4/, sob	硬盘, /ying4pan2/, hard disk	耿直, /geng3zhi2/, honest	风车, /feng1che1/, windmill
49	椒盐, /jiao1yan2/, spiced salt	淑女, /shu1nv3/, fair maiden	交通, /jiao1tong1/, traffic	纽扣, /niu3kou4/, button
50	隐藏, /yin3cang2/, hide	稳定, /wen3ding4/, stable	饮料, /yin3liao4/, drinks	芝麻, /zhi1ma2/, sesame

51	抑郁, /yi4yu4/, depression	仰望, /yang3wang4/, look up at	艺术, /yi4shu4/ art	彩虹, /cai3hong2/, rainbow
52	唯一, /wei2yi1/, sole	难民, /nan4min2/, refugee	围城, /wei2cheng2/, besieged city	星空, /xing1kong1/, starry sky
53	听见, /ting1jian4/, hear	折断, /zhe2duan4/, break off	厅长, /ting1zhang3/, head of a department	冬眠, /dong1mian2/, winter sleep
54	施肥, /shi1fei2/, fertilize	拖欠, /tuo1qian4/, be behind in	尸体, /shi1ti3/, corpse	幸运, /xing4yun4/, lucky
55	仰慕, /yang3mo4/, admire	抑制, /yi4zhi4/, restrain	氧气, /yang3qi4/, oxygen	音乐, /yin1yue4/, music
56	谅解, /liang4jie3/, forgive	京剧, /jing2ju4/, peking opera	亮丽, /liang4li4/, beautiful	牙齿, /ya2chi3/, teeth
57	怯弱, /qie4ruo4/, cowardice	法律, /fa3lü4/, law	窃贼, /qie4zei2/, thief	冰冷, /leng3bing1/, cold
58	梅花, /mei2hua1/, plum blossom	海滩, /hai3tan1/, beach	媒体, /mei2ti3/, media	屋顶, /wu1ding3/, roof
59	津贴, /jin1tie1/, allowance	律师, /lü4shi1/, lawyer	金色, /jin1se4/, gold	苹果, /ping2guo3/, apple
60	挣钱, /zheng4qian2/, earn	静止, /jing4zhi3/, static	证明, /zheng4ming2/, prove	启发, /qi3fa1/, enlighten
61	错误, /cuo4wu4/, error	蜡烛, /la4zhu2/, candle	挫败, /cuo4bai4/, suffer	蜜蜂, /mi4feng1/, bee
62	陈设, /chen2she4/, display	冻结, /dong4jie2/, freeze	沉默, /chen2mo4/, silence	青蛙, /qing1wa1/, frog
63	佳肴, /jia1yao2/, delicacy	桂林, /gui4lin2/, Guilin, place name	夹子, /jia1zi/, clip	竹林, /zhu2lin2/, bamboo forest

64	西瓜, /xi1gua1/, watermelon	晒斑, /shai4ban1/, sunburn	吸管, /xi1guan3/, straw	邮编, /you2bian1/, postcode
65	慰问, /wei4wen4/, condolence	熨斗, /yun4dou3/, iron	味道, /wei4dao4/, flavor	翻阅, /fan1yue4/, browse
66	贬值, /bian3zhi2/, devalue	眨眼, /zha3yan3/, blink	扁担, /bian3dan4/, shoulder pole	裙子, /qun2zi/, dress
67	偶尔, /ou2er3/, occasionally	遇见, /yu4jian4/, meet	呕吐, /ou3tu4/, vomit	局面, /ju2mian4/, situation
68	聆听, /ling2ting1/, listen	冷静, /leng3jing4/, calm down	零食, /ling2shi2/, snacks	洗澡, /xi3zao3/, bath
69	凉爽, /liang2shuang3/, cool	惊险, /jing1xian3/, breathtaking	良好, /liang2hao3/, good	婚姻, /hun1yin1/, marriage
70	聒噪, /guo1zao4/, noisy	活力, /huo2li4/, energy	锅盖, /guo1gai4/, pot cover	办法, /ban4fa3/, method
71	猎人, /lie4ren2/, hunter	措施, /cuo4shi1/, measures	烈火, /lie4huo3/, fire	篮球, /lan2qiu2/, basketball
72	鬼魂, /gui3hun2/, ghost	魄力, /po4li4/, courage	轨道, /gui3dao4/, pathway	杯子, /bei1zi/, cup
73	爆炸, /bao4zha4/, explode	瀑布, /pu4bu4/, waterfall	报纸, /bao4zhi3/, newspaper	蚂蚁, /ma3yi3/, ant
74	乘客, /cheng2ke4/, passenger	乖巧, /guai1qiao3/, cute	承认, /cheng2ren4/, admit	柜台, /gui4tai2/, counter
75	纱布, /sha1bu4/, gauze	吵闹, /chao3nao4/, din	杀人, /sha1ren2/ murder	汽车, /qi4che1/, car
76	滴水, /di1shui3/, drip	摘要, /zhai1yao4/, abstract	低头, /di1tou2/, lower one’s head	慌忙, /huang1mang2/, hurry

77	课堂, /ke4tang2/, classroom	裸露, /luo3lou4/, bare	克制, /ke4zhi4/, restrain	赌博, /du3bo2/, gambling
78	话题, /hua4ti2/, topic	活动, /huo2dong4/, activity	化学, /hua4xue2/, chemistry	压缩, /ya1shuo1/, compress
79	借口, /jie4kou3/, excuse	错过, /cuo4guo4/, miss	介绍, /jie4shao4/, introduce	笑容, /xiao4rong2/, smile
80	锦旗, /jin3qi2/, silk banner	棉花, /mian2hua1/, cotton	紧张, /jin3zhang1/, nervous	短袖, /duan3xiu4/, short sleeve
81	精神, /jing1shen2/, spirit	倩影, /qian4ying3/, shadow	鲸鱼, /jing1yu2/, whale	转身, /zhuan3shen1/, turn around
82	惜别, /xi1bie2/, farewell	借用, /jie4yong4/, borrow	希望, /xi1wang4/, hope	跑步, /pao3bu4/, run
83	酒精, /jiu3jing1/, alcohol	洒脱, /sa3tuo1/, magnanimous	久远, /jiu3yuan3/, distant	教堂, /jiao4tang2/, church
84	割裂, /ge1lie4/, split	豁免, /huo4mian3/, exempt	歌唱, /ge1chang4/, sing	收拾, /shou1shi2/, tidy up
85	排球, /pai2qiu2/, volleyball	诽谤, /fei3bang4/, slander	牌照, /pai2zhao4/, license	宿舍, /su4she4/, dorm
86	清洁, /qing1jie2/, clear	猜测, /cai1ce4/, guess	倾诉, /qing1su4/, pour out	镇定, /zhen4ding4/, calm
87	兑换, /dui4huan4/, exchange	说话, /shuo1hua4/, speak	队长, /dui4zhang3/, captain	宽恕, /kuan1shu4/, forgive
88	蛙泳, /wa1di4/, breaststroke	挂号, /gua4hao4/, register	挖苦, /wa1ku3/, persiflage	诊断, /zhen3duan4/, diagnose
89	牲口, /sheng1kou3/, beast	性别, /xing4bie2/, gender	升华, /sheng1hua2/, sublimation	存款, /cun2kuan3/, deposit

90	渗透, /shen4tou4/, permeate	惨重, /can3zhong4/, grievous	慎重, /shen4zhong4/, careful	琢磨, /zhuo2mo2/, consider

Fillers – semantically related

	Target	S+O-P-
1	公车, /gong1che1/, bus	地铁, /di4tie3/, subway
2	鞭炮, /bian1pao4/, firecracker	炸弹, /zha4dan4/, bomb
3	芒果, /mang2guo3/, mango	柿子, /shi4zi/, persimmon
4	炎热, /yan2re4/, hot	中暑, /zhong4shu3/, sunstroke
5	消费, /xiao1fei4/, consume	购物, /gou4wu4/, shopping
6	餐饮, /can1yin3/, dining	饭店, /fan4dian4/, restaurant
7	枕头, /zhen3tou2/, pillow	床垫, /chuang2dian4/, mattress
8	牡丹, /mu3dan2/, peony	荷花, /he2hua1/, Lotus flower
9	敬佩, /jing4pei4/, esteem	崇拜, /chong2bai4/, adore
10	焦虑, /jiao1lü4/, anxious	放松, /fang4song1/, relax
11	暴露, /bao4lu4/, expose	出现, /chu1xian4/, appear
12	鸽子, /ge1zi/, pigeon	乌鸦, /wu1ya1/, crow
13	枯萎, /ku1wei3/, withered	盛开, /sheng4kai1/, bloom
14	降价, /jiang4jia4/, reduce price	优惠, /you1hui4/, discount
15	名言, /ming2yan2/, dictum	谚语, /yan4yu3/, proverb
16	薪水, /xin1shui3/, salary	收入, /shou1ru4/, income
17	迷信, /mi2xin4/, superstition	封建, /feng1jian4/, feudal
18	阻挠, /zu3nao2/, obstruct	中断, /zhong1duan4/, break

19	顺从, /shun4cong2/, comply	听话, /ting1hua4/, obedient
20	削弱, /xue1ruo4/, weaken	降低, /jiang4di1/, reduce
21	拼搏, /pin1bo2/ struggle	努力, /nu3li4/, work hard
22	针灸, /zhen1jiu1/, acupuncture	推拿, /tui1na2/, massage
23	离开, /li2kai1/, leave	告别, /gao4bie2/, farewell
24	围巾, /wei2jin1/, scarf	手套, /shou3tao4/, gloves
25	判断, /pan4duan4/, judge	鉴别, /jian4bie2/, identify
26	星座, /xing1zuo4/, horoscope	生肖, /sheng1xiao1/, zodiac
27	语气, /yu3qi4/, tone	态度, /tai4du4/, attitude
28	麻雀, /ma2que4/, sparrow	大雁, /da4yan4/, wild goose
29	垃圾, /la1ji1/, trash	废物, /fei4wu4/, waste
30	存款, /cun2kuan3/, deposit	账户, /zhang4hu4/, account
31	飞翔, /fei1xiang2/, fly	翅膀, /chi4bang3/, wing
32	破坏, /po4huai4/, break	伤害, /shang1hai4/, hurt
33	和尚, /he2shang4/, monk	寺庙, /si4miao4/, temple
34	支持, /zhi1chi2/, support	鼓励, /gu3li4/, encourage
35	约束, /yue1shu4/, restrain	自由, /zi4you2/, free
36	悲哀, /bei1ai1/, grief	喜悦, /xi3yue3/, joy
37	哭泣, /ku1qi4/, cry	伤心, /shang1xin1/, sad
38	味精, /wei4jing1/, glutamate	胡椒, /hu2jiao1/, pepper
39	倒霉, /dao3mei2/, unfortunate	幸运, /xing4yun4/, lucky
40	曾经, /ceng2jing1/, ever	过去, /guo4qu4/, past

41	兴奋, /xing1fen4/, excite	平静, /ping2jing4/, calm
42	呼吸, /hu1xi1/, breath	氧气, /yang3qi4/, oxygen
43	震撼, /zhen4han4/, shake	惊艳, /jing1yan4/, stunning
44	洋葱, /yang2cong1/, onion	大蒜, /da4suan4/, garlic
45	服装, /fu2zhuang1/, clothes	鞋子, /xie2zi/, shoes
46	宝藏, /bao3zang4/, treasure	财富, /cai2fu4/, wealth
47	欺负, /qi1fu4/, bully	保护, /bao3hu4/, protect
48	安全, /an1quan2/, safe	危险, /wei1xian3/, dangerous
49	风寒, /feng1han2/, cold	感冒, /gan3mao4/, cold
50	街道, /jie1dao4/, street	马路, /ma3lu4/, street
51	邮轮, /you2lun2/, cruise	快艇, /kuai4ting3/, boat
52	地球, /di4qiu2/, earth	世界, /shi4jie4/, world
53	恋爱, /lian4ai4/, love	结婚, /jie2hun1/, marriage
54	做梦, /zuo4meng4/, dream	睡觉, /shui4jiao4/, sleep
55	赞颂, /zan4song4/, extol	表扬, /biao3yang2/, praise
56	圆滑, /yuan2hua2/, tactful	天真, /tian1zhen1/, innocent
57	稚嫩, /zhi4nen4/, immature	成熟, /cheng2shu2/, mature
58	休息, /xiu1xi1/, take a rest	工作, /gong1zuo4/, work
59	居住, /ju1zhu4/, live	搬家, /ban1jia1/, move house
60	画面, /hua4mian4/, image	印象, /yin4xiang4/, image
61	强大, /qiang2da4/, strong	弱小, /ruo4xiao3/, weak
62	娱乐, /yu2le4/, entertainment	游戏, /you2xi4/, game

63	出生, /chu1sheng1/, birth	死亡, /si3wang2/, death
64	灾难, /zai1nan4/, disaster	地震, /di4zhen4/, earthquake
65	青春, /qing1chun1/, youth	年迈, /nian2mai4/, old
66	天空, /tian1kong1/, sky	白云, /bai2yun2/, cloud
67	医生, /yi1sheng1/, doctor	护士, /hu4shi4/, nurse
68	雪花, /xue3hua1/, snowflake	雨水, /yu3shui3/, rain
69	天堂, /tian1tang2/, paradise	地狱, /di4yu4/, hell
70	大象, /da4xiang4/, Elephant	狮子, /shi1zi/, lion
71	机场, /ji1chang3/, airport	车站, /che1zhan4/, station
72	毛笔, /mao2bi3/, writing brush	墨水, /mo4shui3/, ink
73	风扇, /feng2shan4/, fan	空调, /kong1tiao2/, air conditioner
74	电脑, /dian4nao3/, computer	键盘, /jian4pan2/, keyboard
75	梦想, /meng4xiang3/, dream	追求, /zhui1qiu2/, follow
76	新闻, /xin1wen2/, news	消息, /xiao1xi1/, information
77	辩论, /bian4lun4/, debate	探讨, /tan4tao3/, discussion
78	坍塌, /tan1ta1/, collapse	倒下, /dao3xia4/, falling down
79	摧毁, /cui1hui3/, destroy	建立, /jian4li4/, build
80	钢琴, /gang1qin2/, piano	吉他, /ji2ta1/, guitar
81	美术, /mei3shu4/, art	素描, /su4miao2/, sketch
82	蝌蚪, /ke1dou3/, tadpole	青蛙, /qing1wa1/, frog
83	皇帝, /huang2di4/, emperor	大臣, /da4chen2/, minister
84	小偷, /xiao3tou1/, thief	罪犯, /zui4fan4/, criminal

85	迅速, /xun4su4/, rapid	缓慢, /huan3man4/, slow
86	假日, /jia4ri4/, holiday	工作, /gong1zuo4/, work
87	沉默, /chen2mo4/, silence	安静, /an1jing4/, quiet
88	邮件, /you2jian4/, mail	快递, /kuai4di4/, delivery
89	大海, /da4hai3/, sea	小溪, /xiao3xi1/, stream
90	绵羊, /mian2yang2/, sheep	奶牛, /nai3niu2/, dairy cow
91	时尚, /shi2shang4/, fashion	潮流, /chao2liu2/, trend
92	成功, /cheng2gong1/, success	失败, /shi1bai4/, failure
93	化学, /hua4xue2/, chemistry	物理, /wu4li3/, physics
94	忠诚, /zhong1cheng2/, loyalty	背叛, /bei4pan4/, betrayal
95	虚幻, /xu1huan4/, illusory	现实, /xian4shi2/, reality
96	勇敢, /yong3gan3/, brave	胆小, /dan3xiao3/, timid
97	诚实, /cheng2shi2/, honest	欺骗, /qi1pian4/, lie
98	考试, /kao3shi4/, examination	测验, /ce4yan4/, test
99	勤劳, /qin2lao2/, hardworking	懒惰, /lan3duo4/, lazy
100	潮湿, /chao2shi1/, moist	干燥, /gan1zao4/, dry
101	教育, /jiao4yu4/, education	大学, /da4xue2/, university
102	原谅, /yuan2liang4/, forgive	宽恕, /kuan1shu4/, forgive
103	关怀, /guan1huai2/, care	照顾, /zhao4gu4/, care
104	高尚, /gang1shang4/, noble	卑鄙, /bei1bi4/, mean
105	阿姨, /a1yi2/, aunt	叔叔, /shu1shu/, uncle
106	中秋, /zhong1qiu1/, mid-autumn	团聚, /tuan2ju4/, reunion

107	冬天, /dong1tian1/, winter	暖气, /nuan3qi4/, heating
108	洪水, /hong2shui4/, flood	干旱, /gan1han4/, drought
109	螃蟹, /pang2xie4/, crab	龙虾, /long2xia1/, lobster
110	珊瑚, /shan1hu2/, coral	海洋, /hai3yang2/, ocean
111	植物, /zhi2wu4/, plant	鲜花, /xian1hua1/, flower
112	钱包, /qian2bao1/, wallet	财物, /cai2wu4/, belongings
113	萝卜, /luo2bo/, radish	白菜, /bai2cai4/, cabbage
114	玉米, /yu4mi3/, corn	高粱, /gao1liang2/, sorghum
115	红酒, /hong3jiu4/, wine	饮料, /yin3liao4/, drinks
116	田野, /tian2ye3/, field	水稻, /shui3dao4/, rice
117	眉毛, /mei2mao2/, eyebrow	眼睛, /yan3jing1/, eye
118	凤凰, /feng4huang2/, phoenix	麻雀, /ma2que4/, sparrow
119	车辆, /che1liang4/, car	马路, /ma3lu4/, street
120	圣诞, /sheng4dan4/, Christmas	礼物, /li3wu4/, gift

Fillers – semantically unrelated

	Target	**S+O-P-**
1	铅笔, /qian1bi3/, pencil	饭盒, /fan4he2/, lunch box
2	鼠标, /shu3biao1/, mouse	手机, /shou3ji1/, mobile phone
3	背包, /bei1bao1/, backpack	轮椅, /lun2yi3/, wheelchair
4	夜晚, /ye4wan3/, evening	草原, /cao3yuan2/, grassland
5	风筝, /feng1zheng1/, kite	音乐, /yin1yue4/, music

6	心脏, /xin1zang4/, heart	车祸, /che1huo4/, accident
7	震动, /zhen4dong4/, shake	咖啡, /ka1fei1/, coffee
8	堵车, /du3che1/, traffic jam	优惠, /you1hui4/, discounts
9	书店, /shu1dian4/, bookstore	雪灾, /xue3zai1/, snow disaster
10	相信, /xiang1xin4/, believe	海报, /hai3bao4/, poster
11	相框, /xiang4kuang1/, photo frame	复印, /fu4yin4/, copy
12	完全, /wan2quan2/, completely	风暴, /feng1bao4/, windstorm
13	头发, /tou2fa/, hair	厨房, /chu2fang2/, kitchen
14	准备, /zhun3bei4/, prepare	失眠, /shi1mian2/, sleeplessness
15	康复, /kang1fu4/, recovery	香水, /xiang1shui3/, perfume
16	丝袜, /si1wa4/, silk stockings	中药, /zhong1yao4/, traditional
17	闲谈, /xian2tan2/, chitchat	风格, /feng1ge2/, style
18	卓越, /zhuo2yue4/, excellent	消失, /xiao1shi1/, disappear
19	写字, /xie3zi4/, write	扫把, /sao4ba3/, broom
20	皇帝, /huang2di4/, emperor	行李, /xing2li3/, luggage
21	睫毛, /jie2mao2/, eyelash	运动, /yun4dong4/, sport
22	泉水, /quan2shui3/, spring water	广告, /guang3gao4/, advertising
23	叛逆, /pan4ni4/, rebel against	邀请, /yao1qing3/, invite
24	森林, /sen1lin2/, forest	电脑, /dian4nao3/, computer
25	兔子, /tu4zi/, rabbit	交通, /jiao1tong1/, traffic
26	傍晚, /bang4wan3/, sunset	抽屉, /chou1ti/, drawer
27	作业, /zuo4ye4/, homework	停车, /ting2che1/, parking

28	汇合, /hui4he3/, converge	举杯, /ju3bei1/, toast
29	明星, /ming2xing1/, star	老鼠, /lao3shu3/, mouse
30	美容, /mei3rong2/, beauty	雕塑, /diao1su4/, sculpture

Appendix C (Questionnaire for proficiency assessment)

Screeningbogen für Verhaltensexperimente	Philipps Universität Marburg
Vom Versuchsleiter auszufüllen:	
Experiment:	
Versuchsleiter (Durchführung):	
Verantwortlicher Wissenschaftler:	
VP-Nr.:	Version:
Datum:	Uhrzeit:
CODE:	
Vom Probanden auszufüllen:	
Alter:	Geschlecht:
Beruf: ☐ Sprachkursteilnehmer/in, Stufe___________ ☐ Student/in oder Doktorand/in, Studienfach und Semester_________________________ ☐ Arbeitnehmer/in, wie lange haben Sie in Deutschland gearbeitet_________ ☐ andere _________________	
Muttersprache:	Dialekt:
Sind Sie einsprachig aufgewachsen?	☐ JA ☐ NEIN, zweite Sprache:
In welchem Alter haben Sie begonnen, Deutsch zu lernen?	

Wo und wie lange haben Sie Deutsch gelernt?

Wie lange sind Sie in Deutschland geblieben?

Wie häufig benutzen Sie Deutsch aus folgenden Aspekten?

Schreiben Sie bitte die Nummer auf: 1: nie (0%) 2: selten (0%-20%) 3: gelegentlich (20%-50%) 4: oft (50%-70%) 5:sehr oft (70%-90%) 6: immer (>90%)

Hören: Sprechen: Lesen: Schreiben:

Tragen Sie bitte ein, wie Sie Ihre Deutschkenntnisse einschätzen (A1. A2. B1. B2. C1. C2.):

Hören: Sprechen: Lesen: Schreiben: Insgesamt:

Haben Sie TestDaF, DSH oder ähnliche Sprachprüfungen erledigt?

Wenn Ja, geben Sie bitte den Namen der Prüfung ______________________________

Ergebnisse der Prüfung ______________________________

Prüfungsdatum am letzten Mal __________________________

Welche anderen Fremdsprachen beherrschen Sie noch?

_________________________________ ☐ sehr gut ☐ fortgeschritten ☐ Anfänger

_________________________________ ☐ sehr gut ☐ fortgeschritten ☐ Anfänger

_________________________________ ☐ sehr gut ☐ fortgeschritten ☐ Anfänger

Die folgenden Fragen betreffen Sachverhalte, die das Experiment beeinflussen können

Sehschwäche 视弱: ☐ JA ☐ NEIN	Stärke 视力度数 (in Dioptrien):
Kontaktlinsen 隐形眼镜： ☐ JA	Brille 配戴眼镜: ☐ JA ☐ NEIN

☐ NEIN	
Fühlen Sie sich momentan müde oder erschöpft? 当下您是否感到劳累或者疲惫？ ☐ NEIN ☐ ETWAS ☐ SEHR	
Nehmen Sie momentan Medikamente ein, die Ihre Aufmerksamkeit oder das Nervensystem beeinflussen? 当下是否有服用影响注意力和神经系统的药物？ ☐ JA ☐ NEIN ggf. Erläuterung:	
Ist bei Ihnen eine Hörschädigung, Hörstörung bzw. Schwerhörigkeit bekannt? 您是否有听觉损伤或者听觉障碍？ ☐ JA ☐ NEIN ggf. Erläuterung	
Haben Sie gestern oder heute in größeren Mengen Alkohol getrunken? 昨天或者今天是否有大量饮用酒精饮料？ □☐ JA ☐ □NEIN Wenn ja, in welcher Form und wieviel?	
Haben Sie gerade körperliche oder psychische Beschwerden, die Sie beeinträchtigen (z. B. Kopfschmerzen, Menstruationsbeschwerden, Konzentrationsstörungen)? 当下您是否有身体上或者心理上的不适，会引起头疼、经期不适、注意力集中障碍？ ☐ JA ☐ NEIN	
Bei den folgenden Fragen geht es darum, Ihre Händigkeit festzustellen	
Mit welcher Hand führen Sie die folgenden Tätigkeiten aus? 您用哪只手做以下这些事 Ball werfen 扔球 ☐ Linke Hand ☐ Rechte Hand Zähne putzen 刷牙 ☐ Linke Hand ☐ Rechte Hand Kämmen 梳头 ☐ Linke Hand ☐ Rechte Hand Brot schneiden 切面包 ☐ Linke Hand ☐ Rechte Hand	

Schreiben 写字 ☐ Linke Hand ☐ Rechte Hand

Können Sie mit der anderen Hand annähernd so gut schreiben? 您用另外一只手也可以差不多这样书写吗？ ☐ Ja ☐ Nein

Sind Sie „umgelernter“ Rechtshänder? 您是后天才改为惯用右手的吗？ ☐ Ja ☐ Nein

Tragen Sie bitte auf folgender Skala ein, wie Sie Ihre Händigkeit einschätzen:

请评估您使用左右手倾向性程度并勾选

Links ☐ ----- ☐ ----- ☐ ----- ☐ ----- ☐ Rechts

Für den Versuchsleiter:

Bemerkungen:

Appendix D (Models of group comparisons in Experiment 2)

R script

```
#Mixed Effects Model
library(lme4)
library(lmerTest)
library(effectsize)
library(MuMIn)

#accuracy of German native Group
d  = read.csv('/Users/lu/Documents/ de_daten_de_acc.csv')
d$condition <- factor(d$condition)
d$condition <- relevel(d$condition, ref = "s-o-p+")

model1 = glmer(acc ~ condition + (1|subject) + (1|target), data=d,
family="binomial")
model1a = glmer(acc ~ 1 + (1|subject) + (1|target), data=d,
family="binomial")
anova(model1,model1a)
summary(model1)

library(emmeans)
emmeans(model1, list(pairwise ~ condition),adjust = 'bonferroni')

r.squaredGLMM(model1)

#rt of German native Group
d  = read.csv('/Users/lu/Documents/ de_daten_de_rt.csv')
model1 = lmer(rt ~ condition + (1|subject) + (1|target), data=d, REML =
FALSE)
```

```
model1b = lmer(rt ~ 1 + (1|subject) + (1|target), data=d, REML = FALSE)
anova(model1, model1b)
summary(model1)

library(emmeans)
emmeans(model1, list(pairwise ~ condition),adjust = 'bonferroni')

r.squaredGLMM(model1)

##Chinese group
#accuracy
d  = read.csv('/Users/lu/Documents/ de_daten_ch_acc.csv')
d$condition <- factor(d$condition)
d$condition <- relevel(d$condition, ref = "s-o-p+")
d1 = d[which(d$proficiency == "1"),]
d2 = d[which(d$proficiency == "2"),]

model1a = glmer(acc ~ condition + (1|subject) + (1|target), data=d,
family="binomial")
model1b = glmer(acc ~ 1 + (1|subject) + (1|target), data=d,
family="binomial")
model1 = glmer(acc ~ condition + proficiency + (1|subject) + (1|target),
data=d, family="binomial")
anova(model1a,model1b)
anova(model1,model1a)
summary(model1)

emmeans(model1, list(pairwise ~ condition),adjust = 'bonferroni',
pbkrtest.limit = 4928)
r.squaredGLMM(model1)
```

```
#the influence of proficiency

model0 = glm(acc ~ condition*proficiency, d, family="binomial")

contrast(emmeans(model0, specs = c("condition","proficiency")), by =
"condition", method = "pairwise")


#group a

model1 = glmer(acc ~ condition + (1|subject) + (1|target), data=d1,
family="binomial")

summary(model1)

emmeans(model1, list(pairwise ~ condition),adjust = 'bonferroni')

r.squaredGLMM(model1)


#group b

model1 = glmer(acc ~ condition + (1|subject) + (1|target), data=d2,
family="binomial")

summary(model1)

emmeans(model1, list(pairwise ~ condition),adjust = 'bonferroni')

r.squaredGLMM(model1)


#interaction

model2 = glmer(acc ~  condition  + proficiency + condition*proficiency +
(1|subject) + (1|target), data=d, family="binomial")

anova(model2, model1, test = 'Chisq')




#rt

#the influence of proficiency

model0 = lm(acc ~ condition*proficiency, d)

contrast(emmeans(model0, specs = c("condition","proficiency")), by =
"condition", method = "pairwise")
```

```r
model1 = lmer(rt ~ condition + proficiency + (1|subject) + (1|target),
data=d, REML = FALSE)

model1a = lmer(rt ~ 1 + (1|subject) + (1|target), data=d, REML = FALSE)

model1b = lmer(rt ~ condition + (1|subject) + (1|target), data=d, REML =
FALSE)

anova(model1b, model1a)

anova(model1, model1b)

summary(model1)


#emmeans for Chinese group

emmeans(model1, list(pairwise ~ condition),adjust = 'bonferroni',
pbkrtest.limit = 4928)


library(MuMIn)

r.squaredGLMM(model1)


#group a

model2 = lmer(rt ~ condition + (1|subject) + (1 |target), data=d1, REML =
FALSE)

emmeans(model2, list(pairwise ~ condition),adjust = 'bonferroni')

r.squaredGLMM(model2)


#group b

model2 = lmer(rt ~ condition + (1|subject) + (1 |target), data=d2, REML =
FALSE)

emmeans(model2, list(pairwise ~ condition),adjust = 'bonferroni')

r.squaredGLMM(model2)


##interaction of group DE * group CH

d = read.csv('/Users/lu/Documents/ de_daten_combi_acc.csv')

model0 = glm(acc ~ condition*nation, data=d,family="binomial")
```

```
contrast(emmeans(model0, specs = c("condition","nation")), by = "condition",
method = "pairwise",adjust = 'bonferroni')

model3 = glmer(acc ~ condition + nation + condition*nation + (1|subject) +
(1|target), data=d, family="binomial")

model3a = glmer(acc ~ condition + nation + (1|subject) + (1|target), data=d,
family="binomial")

anova(model3,model3a)


d = read.csv('/Users/lu/Documents/ de_daten_combi_rt.csv')

model0 = lm(rt ~ condition*nation, data=d)

contrast(emmeans(model0, specs = c("condition","nation")), by = "condition",
method = "pairwise",adjust = 'bonferroni')

model3 = lmer(acc ~ condition + nation + condition*nation + (1|subject) +
(1|target), data=d, REML = FALSE)

model3 = lmer(acc ~ condition + nation + (1|subject) + (1|target), data=d,
REML = FALSE)

anova(model3,model3a)
```

For the complete dataset and R script, kindly refer to the OSF platform using the following website link:
https://osf.io/3hk4c/?view_only=a5fcf50ccf664ff486801e892808ecdf