Lexical Organisation in Chinese Spoken Word Production:

Evidence from Studies of Homophones

Yue Shen

A thesis submitted for the degree of Doctor of Philosophy

Department of Psychology

University of Essex

April 2019

Abstract

The research reported in this thesis was designed to shed light on the nature of the functional lexical organisation in Mandarin Chinese spoken word production. Two main theoretical issues were investigated: The lexical representations of homophones in Chinese, and how activation is transmitted from lexical to phonological levels in Chinese spoken word production.

The results of two experimental studies of word reading responses to homophones and non-homophonic words (matched for word frequency, and a range of other variables) found no effect of frequency inheritance, contrary to the hypothesis that homophones have a shared phonological representation. The results of six experimental studies that examined the priming of object naming times by the prior presentation of prime words of various relationships showed that there are clear direct effects of repetition, homophone, and phonological (atonal syllable) priming. However, the experiments found no homophone-to-semantic, phonological-to-semantic, or semantic-to-homophone mediated priming effects. These results offer no support for interactive processing models of Chinese word production. However, two experimental studies found that Chinese-English bilinguals show homophone priming of object naming that is mediated by the Chinese translation-equivalents of English prime words.

The results of the research reported in this thesis support two general conclusions concerning the Chinese speech production system: homophones have independent lexical phonological representations, and the processing underlying spoken word production operates in a mainly serial and discrete manner.

Acknowledgements

The completion of this thesis is almost impossible without the assistance of many important individuals. My deepest gratitude goes to my supervisor, Professor Chris Barry, who always kept a sense of humour when I had lost mine. Had it not been for his patience and encouragement, I would not have been able to make this far. His inspiring comments and constructive feedback made my arguments more logically convincing in the thesis. I truly appreciate his time and effort dedicated during my doctoral study. Thank you so much!

I am also thankful to my viva examiners, Dr Kevin Dent and Professor Markus Damian, for their suggestions that were crucial and beneficial to the development of my thesis.

I would also like to extend my thanks to the technicians and administrative staffs in the Department of Psychology. Special thanks should be given to Alan Brignull, Dr Woakil Ahamed and Lesley Monk, for their help in offering me the resources in running my projects.

I would like to take advantage of this opportunity to thank my friends for their encouragement throughout my study.

Finally, I would like to express my special thanks to my loving parents who encouraged me to pursue my PhD and who never stopped supporting me.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW

1. General Introduction

Words that people produce in speech, and hear others say, may often have more than one meaning. Many of these semantically ambiguous words may become disambiguated by context, if available, and some will become disambiguated when written. Words with identical pronunciations whose different meanings are signalled by different written forms are called heterographic homophones, such as *their* and *there*, *steak* and *stake*, *rain*, *reign*, and *rein*, and, for speakers with southern English accents, *or*, *oar*, *ore*, and *awe*. (In English, most heterographic homophones have only two alternative spellings.) Homographic homophones, such as *palm* (hand or tree) and *cut* (used as a verb and as a noun), are ambiguous in speech and remain ambiguous when written. Chinese has very, very many heterographic homophones; for example, 势力，视力，示例，事例，and 势利 are all pronounced identically and with the same tone (as "shi4 li4", the number indicates the tone information). Further, there can be very many alternative homophones; indeed, there are 48 Chinese monosyllabic words that have identical pronunciation /yi4/.

The experimental work reported in this thesis is intended to illuminate the nature of the representation and processing of homophones in spoken word production in Chinese.

Two main theoretical issues were investigated: The lexical representations of homophones in Mandarin Chinese, and how activation is transmitted from lexical to phonological levels in Chinese spoken word production.

There are two main contrasting theories of how homophones are represented in the speech production system. The shared representation account (Jescheniak & Levelt, 1994; Jescheniak, Meyer, & Levelt, 2003) proposes that there is only one, common phonological word-form of homophones; for example, the words *none* and *nun* would share the same phonological representation. The independent representation account (Caramazza, Bi, Costa, & Miozzo, 2004; Caramazza, Costa, Miozzo, & Bi, 2001) proposes that each word has its own separate phonological representation, even if these are redundantly represented for homophones (and so both *none* and *nun* have their own separate and independent representations).

Theories of speech production differ regarding the proposed flow of activation from lexical to phonological levels. The main contrasting theories of the temporal dynamics of information flow are: (1) Serial and discrete models (Levelt, Roelofs, & Meyer, 1999), which propose that processing at one level must be completed before the next level begins; (2) Cascading models (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Morsella & Miozzo, 2002), which propose that processing at one level continuously feeds-forward information (before processing is fully completed at that level) to activate the next level; and (3) Interactive models (Dell, 1986), which propose the bi-directional flow of activation between different processing levels.

2. Homophones in Chinese

Chinese is a tonal spoken language, where the tone is used to distinguish between different words. In Standard (Mandarin) Chinese there are four tones, along with the neutral tone; Cantonese has nine tones. However, tones are not directly represented orthographically; the tone is given by the recognition of the written word. For example, the words 妈, 麻, 马, and 骂 are all pronounced with the same monosyllable "ma" but have different tones, which distinguish the different meaning of the words. Homophones in Chinese are words with identical pronunciations and so have both the same syllable and the same tone.

An important feature of Chinese is its pervasive homophony. Chinese has very many heterographic homophones (especially so for monosyllabic words) and has many more than in English. There are two main reasons for this. First, Chinese orthography is not constrained by any systematic pressure to maintain (or not to further complicate) spelling-to-sound consistency of languages with alphabetic orthographies, and so allows entirely distinctively different written forms of homophones. Second, and more importantly, the syllabic structure of Chinese might be seen as being more 'simple' than English. Modern-day Mandarin Chinese usage includes 1290 distinct syllables (including tone), among which there are about 900 syllables mapping onto more than one Chinese characters. In contrast, there are over 6,000 syllables in English speech. Therefore, it could be argued that spoken English has more potential scope to distinguish between different words than spoken Chinese. However, Chinese orthography has considerably more scope for distinguishing between homophones than English (where homophones are typically orthographically similar).

The number of different written forms of the same sound can be very large in Chinese. Whereas the majority of homophones in English have only two or three alternative forms (e.g., *sale*, *sail*; *two*, *too*, *to*), homophone variants (or homophone "families") in Chinese can be very large (e.g., the 48 forms for /yi4/), and it is not unusual to have homophone families of between two and fourteen different words. An additional difference between English and Chinese concerning homophony is that English homophones can be morphologically different (e.g., *build* and *billed*). Chinese has almost no inflectional affixes at all to convey grammatical relationship. Given these differences between English and Chinese, the study of how characters or words with the same phonemes are represented in the phonological lexicon is of considerable importance for theories of the lexical processing system in Chinese. Indeed, Jescheniak, Meyer, et al. (2003) have acknowledged this importance in their view that "cross-linguistic comparisons are needed to assess the issue of possible language-specific differences" (p. 437) in how homophones are represented.

2.1 Chinese orthography

The basic writing units in Chinese are characters, which generally map onto single-syllable morphemes (not phonemes), and words are very often composed of more than one character. The majority of Chinese characters (about 85%) are actually compounds with two components: a phonetic component and a semantic (or radical) component. Due to historical changes in Chinese pronunciation, the phonetic component does not always provide a reliable guide to the pronunciation of the word; indeed, Y. G. Zhou (1978) estimated that it predicts pronunciation correctly for only 39% of characters.

Although many homophones in Chinese are monosyllabic words, this thesis will focus on both monosyllabic and disyllabic words. To begin with, monosyllabic words often have various and ambiguous meanings. For example, the Chinese word 聪 means acute hearing or wise, while 失聪 only refers to deaf and 聪明 only refers to intelligent. It would be hard to identify which conceptual representation is activated when naming monosyllabic words. Second, it is well known that there are no markers for word boundary in Chinese, and the distinction between characters and monosyllabic words is unclear in Chinese. The common understanding of a word is that it consists of two or more characters, and theories of Chinese syntax continue to debate whether individual characters may or may not qualify as words. It is also worth noting that in word frequency and corpus research, monosyllabic words are given much attention (Liu, Shu, & Li, 2007; Sze, Rickard Liow, & Yap, 2014). Nevertheless, they are not always discussed in the theory of prosodic morphology (Feng, 2001). According to *the Chinese Homophone Dictionary*, there are about 3200 sets of homophones (including homographic homophones but not monosyllabic words), the majority of which are disyllable words. Hence, this thesis will focus on both disyllabic and monosyllabic heterographic homophones (referred to as “homophones” from now on).

Disyllabic words in Chinese are composed of two physically separated characters; most of these words are compounds comprising two morphemes. Whether compound words are represented at the lexical level as wholes or as individual morphemes remains an unsettled issue. One view assumes that compounds are represented regarding their constituent lexical morphemes (Ji, Gagné, & Spalding, 2011; Taft, 1988; Taft & Forster, 1975, 1976). (This decomposition hypothesis is compatible with models that claim that homophones have shared phonological representations.) In these models, a printed word such as *fireman* would be first decomposed into the two morphemes *fire* and *man*, then the orthographic representation of these constituent morphemes would connect to their corresponding semantic and phonological representations, as well as the semantic representations of compound words containing these morphemes. At the phonological and orthographic levels, since the spoken and written forms of compounds are the concatenations of the forms, there is no additional whole-word representation of the compounds at these levels. However, the meanings of compounds are not merely concatenations of constituent morphemes; there are additional representations at the semantic level. How semantic representations of the compounds and their constituent morphemes are organised at this level is still in debate (Taft, 2003; X. Zhou & Marslen-Wilson, 2000). In terms of the decomposition framework, the production of a compound word such as "fireman" would not only be influenced by the retrieval of its own semantic unit, but also by the retrieval of its individual constituents.

In contrast, the full-form representation hypothesis assumes that compounds are represented as whole-word forms in distinct lexical nodes (Caramazza, 1997; Janssen, Bi, & Caramazza, 2008). This model assumes that the constituent morphemes of compounds do not play a role in lexical access. Evidence showing that compounds are represented in terms of their constituent morphemes would undermine this hypothesis.

In order to distinguish between the decomposition and full-form hypotheses, experiments have tried to manipulate word frequency and morpheme frequency independently in tasks requiring the processing of compounds. Unfortunately, research investigating picture naming and lexical decision has provided inconsistent results. Janssen, Bi, and Caramazza (2008) found that the speed of lexical access of Chinese and English compounds was determined by the frequency of the whole-form and not by the frequencies of their constituent morphemes. Similarly, Cai and Brysbaert (2010) measured the reaction time for two-character Chinese words in a lexical decision task and found that the frequency of the first character was no longer a significant predictor when the frequency of the whole word was taken into account. However, other studies showed that both word-level and character-level frequency variables contributed significantly to the reaction times of lexical decision (Tsang et al., 2018; B. Zhang & Peng, 1992). Using the picture-word interference task, Bi, Xu, and Caramazza (2009) found significant facilitation effects when the distracter word (which was a single character in their experiment) was orthographically similar to or phonologically related to the first character of the picture name. (Most of the picture names in their experiment were disyllable words.) This result suggested that the orthographical and phonological activation of the constituent morphemes (or at least the first one) would facilitate the retrieval of picture name or its phonological segments. It seems that the whole word frequency is a consistent effect on Chinese lexical selection and word production, while the morpheme frequency effect is not always observed. One way to interpret these confusing results would be to incorporate the whole word representation in the orthographic and phonological level and to allow parallel activation between constituent morphemes and whole words. There may be two routes to the production of compounds: via the representation for constituent morphemes (giving an effect of morpheme frequency) and directly from the relevant whole-word units (giving an effect of whole word frequency). Whether the morpheme frequency effect would be observed will depend on which route activates the phonological word representation first, and this may depend on a number of different factors, for example, the semantic relation between morpheme and word (X. Zhou & Marslen-Wilson, 2000), the position of morpheme (W. Wang, Lu, He, Zhang, & Zhang, 2017), and morpheme structure (B. Zhang & Peng, 1992). However, a disadvantage of this model is that there would need to be considerable redundancy in representations with the same characters.

3. Models of Spoken Word Production

In general, theories of speech production (e.g., Levelt, 1989) have distinguished three major levels of speech production: (1) conceptualisation, or the planning of the meaning of the intended message; (2) formulation, or the transformation of the conceptualised message into a structured sentence, and the retrieval of sounds needed to create the sentence, and (3) articulation, or the actual execution of the intended words.

The transition from the conceptualisation level to the formulation level is often referred to as lexicalisation, which is the process whereby conceptual representations activate words to be produced in speech. For example, the concepts <animal>, <pet>, <barks>, <man's best friend>, and <wags tail when happy>, can activate the word "dog" to be produced.

3.1 Levelt, Roelofs, and Meyer (1999)

Levelt, Roelofs, and Meyer's (1999) influential model of speech production is presented in Figure 1.1. This model proposes that lexical access in speech production is processed in a staged, feed-forward and activation-spreading network. After the first stage of conceptual preparation, word generation proceeds through lexical selection, morphological and phonological encoding, phonetic encoding, and articulation itself. Each stage produces its own output representation, and they are, respectively, lexical concept, lemma, morphemes, phonological words and phonetic gestural scores (which are executed during articulation). The lexical concept nodes and conceptual links are in the conceptual stratum. Lemmas and a word's syntactic properties are in the lemma stratum. Morphemes and their phonemic segments are in the form stratum. Since the functions of phonetic coding and articulation are not the focus of this thesis, they will not be considered here.

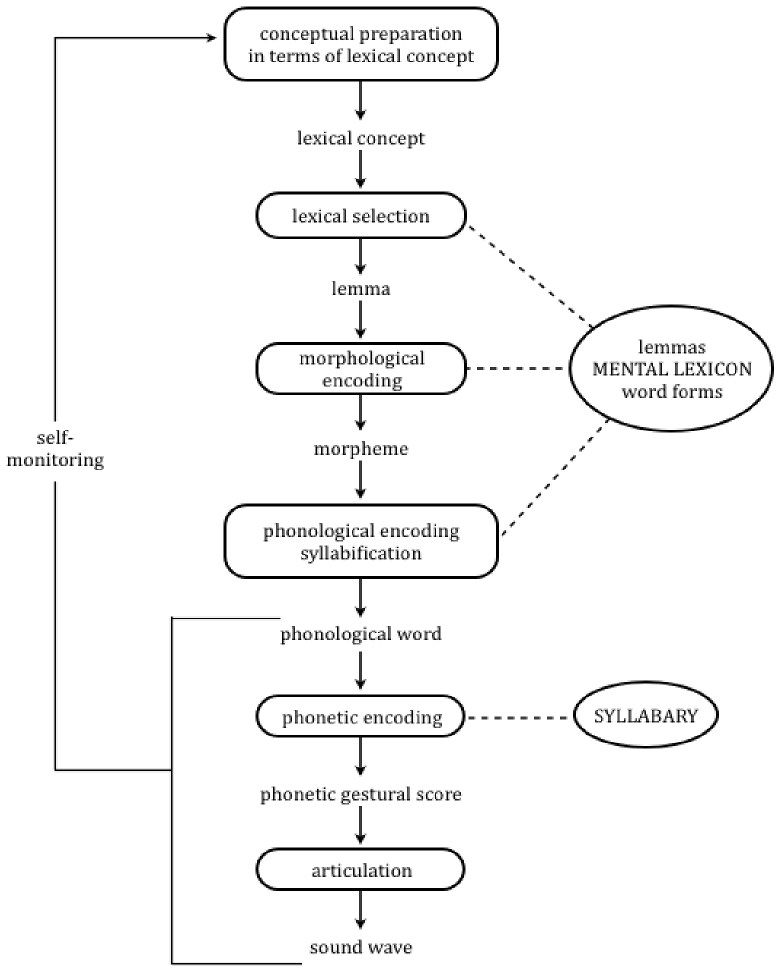


Figure 1.1 Levelt, Roelofs, and Meyer's (1999) model of speech production

Conceptual stratum. Each lexical-semantic concept is represented by an independent concept node, and there are interconnected conceptual links, which specify conceptual relations, for example, between a concept and its super-ordinate. In this conceptual network, nodes spread activation via links to semantically related concepts. A fundamental claim of the model is that lexical concepts are not represented by sets of semantic features, and so they cannot be decomposed (Roelofs, 1997, but for a decomposition view see Bierwisch & Schreuder, 1992). This feature is established because there is no evidence that people do not produce hyponyms or super-ordinates of intended words (Levelt, 1989) nor any clear evidence that words with more complex feature sets are harder to access than words with simpler ones (Levelt, Schreuder, & Hoenkamp, 1978). A word's meaning is represented by the total of the lexical concept's labelled links to other concept nodes. In everyday language use, a lexical concept is often activated as part of a larger message that captures the speaker's intention. For example, if a speaker is asked to name a picture of a hen, she may effectively do so by producing the word *hen*, which involves the activation of the lexical concept "hen"; however, if the speaker does not know the word *hen,* she may use a phrase, such as "female chicken". A major issue here is how speakers get from the notion they intended to express to the corresponding concept node. There is no simple, hard-wired, one-to-one connection between perceptions and lexical concept.

Regarding the time course of conceptual preparation, Indefrey and Levelt (2004) suggested a time window of 150-200ms from picture onset to the selection of the target name. They conducted a comprehensive meta-analysis of word production studies, such as picture naming, and combined results from brain mapping with information on the time course of word production provided by behavioural and electromagnetic studies. (A similar time window has also been suggested by Thorpe, Fize, and Marlot's (1996) study of event-related potentials in a go/no-go categorisation task of deciding whether a picture was an animal or not.) Abdel Rahman and Sommer (2008) found greater P1 (around 120ms) amplitude for rare objects with minimal knowledge relative to both rare objects with in-depth knowledge and familiar objects. This P1 effect was found to be task-independent, as three tasks were employed in this study: a familiarity task (responded by button-presses), a semantic task (with verbal responses) and a naming task. This early ERP component reflects the effect of conceptual knowledge on perceptual analysis and object recognition. Indefrey (2011) argued against a "too rigid interpretation", as naming objects depend on numerous variables and conceptual information may take different times to make names available.

Lemma stratum. An activated lexical concept spreads activation to its lemma node, and lemma selection is a statistical mechanism that favours the selection of the highest activated lemma. Lemma nodes, syntactic property nodes, and labelled links between, them constitute the lemma stratum. Many lemmas have different parameters that have to be set. For instance, in English, verb lemmas have features for number, person, tense and mood. The parameters are valued during the process of grammatical encoding. More generally, the lemma stratum is linked to a set of procedures for grammatical encoding. The evidence of the existence of lemma mainly rose from studies of tip-of-tongue states (e.g., Vigliocco, Antonini, & Garrett, 1997), lateralized readiness potentials (e.g., Coles, 1989), and a short-lived frequency effect in accessing gender (Jescheniak & Levelt, 1994).

Word-form stratum. After a lemma has been selected, its activation spreads to the word-form stratum that contains three types of information: the word's morphemes, its metrical template, and its phonemic segments. Each morpheme node is linked to its relevant segment nodes. Then the phonemic segments with labelled links indicating their correct ordering are available for a set of procedures that generate a phonological word's syllabification, given the syntactic/phonological context. In this theory, syllabification is a late process, because it often depends on the word's phonological environment.

The failure to fully activate a selected lemma's phonological form can produce the tip-of-the-tongue phenomenon: the speaker knows the semantic and conceptual contents of the intended message but fails to articulate the word. Levelt et al. argue that the existence of the tip-of-the-tongue phenomenon suggests that there is a separate lemma stage, because often speakers in a tip-of-the-tongue state can report the grammatical gender of the word (Vigliocco et al., 1997), or even its number of syllables, which are made available before the phonological form and the individual phonemes of the word are retrieved.

3.2 How many stages are involved in lexicalisation?

Levelt et al.'s model contains Levelt's (1989) proposal of a two-stage model of lexicalisation, where the conceptualised message goes through a lemma selection level, followed by lexeme retrieval. A lemma is an abstract representation of a word, which contains semantic and syntactic information, but no phonological information. The lemma mediates between semantic and syntactic information and its phonological representation, which is referred to as the lexeme (or phonological word-form).

The two-stage model argues that appropriate lemmas representing the intended message are chosen at the lemma selection level and that only the lexeme of the selected lemma is activated and encoded before the intended word can be articulated.

An opposing view was put forward by Caramazza (1997) who argued that there was no need for a separate lemma stage and proposed that semantic representations can directly activate the phonological forms of words, stored in a phonological output lexicon. The evidence that Caramazza appealed to in order to support this view came from both neuropsychological patients who made semantic errors in certain modalities, such as oral naming or oral reading, but had no problems with exactly the same concepts in other modalities, such as in written production (e.g., Caramazza & Hillis, 1990), and from the independence of syntactic and phonological information in tip-of-the-tongue states (Caramazza & Miozzo, 1997).

3.3 Information flow between the stages of lexicalisation

An aspect of lexicalisation in speech production that has been widely debated concerns the temporal dynamics of the stages in the process of speech production. How are the different levels of processing organised, and how do they operate in relation to each other? Most models of spoken word production fall into one of three main categories: serial and discrete, cascaded, or interactive. There are also hybrid models that might assume discrete serial processing between some processing levels but cascaded or interactive processing between other processing levels. For example, Levelt et al.'s (1999) model proposes some limited cascading from the conceptual to the lemma level, as semantic information can activate more than one lemma, but then claims that activation is transmitted from selected lemmas to phonological word-forms discretely and serially. Interactive models may be either partial, involving feedback of activation from one level to its immediately previous level, and fully interactive, involving feedback to all activated levels. It is also possible to distinguish between "full-cascading" and "limited-cascading" accounts of speech production (e.g., Kuipers & La Heij, 2009). In full-cascading accounts, the flow of information is not restricted in any way, such that activation at one level will feedback to all subsequent levels. (This is also a claim of interactive models such as Dell's, 1986.) Full-cascading models propose that a semantically activated concept will activate a number of lexical representations, all of which are phonologically encoded. In limited-cascading accounts, there are restrictions on the extent of information flow. For example, Bloem and La Heij (2004) proposed a limited-cascading view of speech production whereby a concept selected to be produced in speech activates a limited semantic cohort of lexical representations which will be phonologically encoded.

Levelt et al.'s (1999) serial-discrete model proposes that only a selected (or target) lemma spreads activation to the phonological level (i.e., semantic processing must be completed before phonological encoding commences). Empirical support for this model was provided with a study reported by Levelt, Schriefers, Vorberg, Meyer, Pechmann, and Havinga (1991). Participants performed lexical decisions to auditory probes presented at different stimulus-onset asynchronies during object naming. Unlike the normal picture-word interference paradigm, the dependent variable in these experiments was the lexical decision latencies for different kinds of probe words. These experiments showed a significant phonological effect such that lexical decision latencies were slower when the probes were phonologically related to the name of the target picture (e.g., the probe *sheet* with the picture of a sheep) compared to unrelated words. However, there was no effect from auditory probe words phonologically related to a semantic associate (e.g., the probe *wood* for sheep, mediated by WOOL) or a same-category member (e.g., the probe *goal* for sheep, mediated by GOAT). These results suggest that there is only phonological encoding of the target name, with no phonological activation of non-target words. Therefore, Levelt et al. (1991) argued that the two-stage discrete model could better explain the data than the activation-spreading cascaded theories.

Evidence in support of the serial-discrete models of lexical access also came from picture-word interference tasks using "mediated" priming. For example, when the target picture is a dog, the mediated distractor could be *can*, which is phonologically similar to the mediator (cat), that is semantically related to the target. Jescheniak, Hahne, and Schriefers (2003) found both a phonological facilitation effect (picture dog plus the distractor *dot*) and a semantic interference effect (e.g., picture dog plus the distractor *cat*) in ERPs, but they observed no mediated effects. In order to increase the saliency of the mediated relation between a picture name and a target word, participants were asked to name a picture of the category associate a few trials before the critical mediated trial. Again, there was a phonological effect but no specific mediated effect.

The issue of whether the phonological activation is restricted to the target lexical node in spoken word production remains controversial, as there is evidence that semantically-activated non-target lexical nodes may also be phonologically encoded (e.g., Jescheniak & Schriefers, 1998; Peterson & Savoy, 1998; Jescheniak, Kurtz, & Schriefers, Günther, Klaus, & Mädebach, 2017); this evidence favours cascaded models of lexicalisation. Morsella and Miozzo (2002) used the picture-picture task in which participants were asked to name a target picture (cued by colour) whilst ignoring a second (distractor) picture superimposed upon the target picture. The name of the distractor picture was either phonologically related or unrelated to the name of the target picture (e.g., *bell* or *hat* for the target name 'bed'). Morsella and Miozzo argued that if, as expected by the serial-discrete model, all processing at the lexical level must be completed before the phonological level, then there should be no difference in naming times between the phonologically related and unrelated distractor conditions. However, contrary to this prediction, they found a facilitation effect from the phonologically related distractor picture; naming times were faster when the distractor object's name was phonologically similar. In a control experiment conducted using the same experimental materials with Italian speakers, where the names of the objects were not phonologically related, there was no difference between the two conditions. Morsella and Miozzo's results support cascaded models, which claim that phonological retrieval of potential words begins before a single lexical representation has been selected and processing at the lexical level has been completed.

Meyer and Damian (2007) investigated three types of phonological relationship in the picture-picture task: homophones, same initial segments and same final segment. All three types of phonological relatedness yielded facilitation and the familiarity phase before the experiment did not affect the phonological facilitation. Further, the phonological facilitation from the context picture was replicated in other languages, for example, in Spanish (Navarrete & Costa, 2005) and Dutch (Roelofs, 2008).

The cascaded view has also received supporting evidence from a Stroop-like colour-naming task. Participants were presented with the object depicted in different colours and were asked to name the picture's colour while ignoring the object. It was found that colour-naming time was shorter when the colour's name was phonologically related to the object's name compared to when it was unrelated (Kuipers & Heij, 2009; Navarrete & Costa, 2005).

The cascaded view has also been investigated in bilingual speakers by examining the effects of a word's translation. For example, Costa, Caramazza, and Sebastián-Gallés (2000) compared picture-naming times for objects with cognate names (i.e., phonologically similar names in both languages) and non-cognate names. For Catalan-Spanish bilinguals, naming latencies were shorter for pictures with cognate rather than non-cognate names. The difference between naming cognate and non-cognate pictures was not present when participants were monolingual speakers. The results indicate that both Spanish and Catalan names were activated in Catalan-Spanish bilinguals during lexical selection. When picture names were cognate words, the Spanish name and Catalan name were phonologically similar. The cognate facilitation in bilinguals indicates that non-selected lexical nodes are also phonologically encoded.

However, cascading of activation from the semantic to the phonological level seems to be affected by the particular tasks used. Kuipers and Heij (2009) found colour-naming times were facilitated when the name of the picture was phonologically similar to its colour, but when the task was object naming, there was no phonological effect from the non-target colour name. Similarly, Roelofs (2008) found that in a picture-word inference task that also recorded eye tracking, phonologically related distracter words produced a facilitation effect on both naming latency and gaze shift, but when the task was word reading, context pictures had no phonological effect on word naming latency. A similar asymmetry was also obtained between naming colour and naming colour-words (Roelofs, 2005). The asymmetry between naming pictures and colours, naming words and colours, and naming pictures and words suggests that the amount of activation that spreads from concepts to phonological forms is limited and task-dependent (Roelofs, 2008). This asymmetry suggests that cascading is limited to the "primary" dimension (i.e., the object name) but that non-target properties such as colour (or size, as shown by Roux, Bonin, & Kandel, 2014) do not seem to be processed in a cascaded fashion. Hence, the evidence suggests a "limited" or "weakly" cascaded view of spoken word production.

A difference in selective attention involved across tasks might explain the asymmetry described above. For example, context picture distractors could yield a phonological effect in picture naming, whereas they do not have any effect on word reading. In the picture–picture task, target and distractor pictures were presented as line drawings in different colours. In order for the participants to separate the target and distractors, it is necessary to pay attention to the non-target. Only when participants attentionally enhance the activation of the picture name do pictures affect word reading, as observed by Peterson and Savoy (1998), and as observed by Roelofs (2003) for the colour-word Stroop task.

The dependence on attention for the flow of activation may also explain why there is a phonological facilitation effect in the picture-picture task, whereas there is no easily detectable activation of the phonological form of semantic competitors of a picture name, given that Levelt et al. (1991) detected no mediated priming effects. A significant difference between the picture-picture task and the mediated priming task situation is that the distractor (and competitor) that yields phonological activation is explicitly presented as a picture in the picture-picture task but not in the mediated priming task situation. Presented pictures may attract more attention than the internally activated words in the mediated priming task. These differences among studies suggest that the amount of activation that cascades through the system are limited.

There is accumulating evidence from both the speech error literature and the many studies of the time course of word production to support cascading activation from semantics to the lexical level, where typically more than one word will be activated by a conceptual representation (Caramazza, 1997; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt et al., 1999; Rapp & Goldrick, 2000). There is also evidence to suggest that lexical activation cascades to phonological word-forms, and/or that there is feedback from phonological encoding to the process of lexical selection (for a review of studies showing facilitation of words phonologically related to semantically activated items see Table 1 in Goldrick, 2006).

The interactive model proposed by Dell (1986) was based primarily on data from the analysis of speech errors. The model proposes that there is feedback from the lexeme nodes to the lemma nodes. For instance, when the lemma CAT is activated, the morpheme CAT and its phonemes /k/, /æ/, and /t/ also become active, and these will feedback activation to all lemmas sharing these phonemes. This proposed interactive phonological-to-lexical feedback is used to explain the production of *mixed* semantic and phonological speech errors (such as saying "rat" when cat is intended), which occur more often than expected from the rates of purely semantic errors (e.g., cat –> "dog") or purely phonological errors (e.g., cat –> "cab" or "hat"). For purely semantic errors, there will be no phonological feedback, as the segments of CAT are not shared by the lemma of DOG or other phonologically unrelated words (Gagnon, Schwartz, Martin, Dell, & Saffran, 1997).

Cascaded activation flow and feedback from phonological word-forms to the lexical level can also explain the *lexical bias* effect in speech errors. This is the tendency for phonological errors to result in words being produced more often than nonwords, which has been reported in studies of spontaneous speech errors (e.g., Dell & Reich, 1981), experimentally induced errors (e.g., Hartsuiker, Corley, & Martensen, 2005), and errors produced by aphasic speakers (e.g., Gagnon et al., 1997). For example, while naming the word *cat*, the activation of its phonemes /k/, /a/ and /t/ fed-back to the lexical level would increase activation of both the target words and phonologically similar words (e.g., *hat* and *cab* etc). If the phoneme /h/ was activated (by hearing it, as in phonemic cueing, or from an earlier production of a word containing this phoneme), then this could, via phoneme-to-lexical feedback, increase activation back to the lexical level, both increase the activation of the similar word *hat*, whereas if the phoneme /z/ was activated, the nonword form *zat* would not be feedback to the lexical level. Therefore, errors at the phonological level are more likely to be a word than a nonword. The feedback of activation could not only enhance the representation of a form-related nontarget word, but it could also enhance the phonological representation of the target as well. For example, words with many phonological neighbours were produced more quickly and more accurately than words with few neighbours.

In Dell's model, CAT would feedback activation from its phonemes to increase the activation of RAT whose activation level would also be increased by semantic-level representations. However, as Levelt et al. have argued, it appears strange to propose interactive feedback solely to account for instances of malfunctioning (i.e., rarely produced speech errors) without such feedback also having functionality for correct production. Indeed, Levelt (1999) argued that such bi-directional feedback lacks independent empirical motivation, as "its functionality can hardly be to induce speech errors" (p. 225). Levelt et al. (1999) and Roelofs (2004) suggest that both mixed errors and the lexical bias effect may be interpreted in terms of the operation of a speech monitoring system whereby, prior to articulation, the phonological form of the speaker's utterance is fed back through the language comprehension system. This "speech monitor" would not require interactive feedback within the speech production system, although Goldrick (2006) has argued that, in order for the monitor to be engaged, cascading activation is still required. More generally, Levelt et al. (1999) argue that theories of speech production must always be corroborated by evidence from the time taken for correct production (for example from studies using the picture-word interference task).

Using the picture-word task, Schriefers, Meyer, and Levelt (1990) varied the stimulus-onset asynchrony (SOA) of target pictures and auditory distractor words to examine the time-course of the effects of semantically and phonologically related distractors (e.g., RADIO+ *television* or *radish*). When the words were presented just before (SOA = -150ms) and simultaneously with the target object (SOA = 0), Schriefers et al. found a reliable semantic interference effect; this interference was absent when the distractor word was presented just after the object (SOA = +150ms). In contrast, a phonological facilitation effect was absent when the SOA = -150ms, but was reliable when SOA = 0 and +150ms. The fact that there was an "early" semantic effect and a "late" phonological effect is consistent with independent stage models where semantic processing precedes phonological processing in discrete stages (rather than reflecting interactive processing). Damian and Martin (1999) extended this work to also include distractor words that were semantically *and* phonologically related to the target objects (e.g., DUCK+*dove*), in addition to only semantically related (e.g., DUCK+*raven*) and only phonologically related words (e.g., DUCK+*dab*). They found that semantically and phonologically related distractors produced the same pattern of results as purely phonologically related words (i.e., a late phonological facilitation effect). This result is similar to that reported by Starreveld and La Heij (1995), who also found that the semantic interference effect was reduced when the distractor words were both semantically and orthographically/phonologically related to the target. These results appear to contradict the discrete two-step account of Levelt et al., as they show that semantic interference is eliminated when semantically related distractor words are also phonologically related; it would appear that the "later" stage of phonological encoding influenced the “earlier” stage of semantically activated lexical retrieval, which is more consistent with an interactive than a discrete model.

To conclude, this review of the evidence from speakers of European languages has not produced a consistent pattern of results. However, there appears to be increasing acceptance of the idea that there is cascading activation from the lexical level to phonological representations (at least in a “limited” or “weakly” cascaded form), and studies of spoken naming have begun to address the circumstances under which cascaded processing operate (e.g., Oppermann, Jescheniak, Schriefers, & Görges, 2010; Roelofs, 2008). However, the evidence for interactive feedback from the phonological level to the lexical level is more limited (and, in the case of speech errors, open to alternative interpretations), and the idea that there is fully interactive processing remains controversial.

3.4 Models of word production in Chinese

Most of our understanding of spoken word production is based on evidence from Indo-European languages such as English, German, Spanish and Dutch. However, there has been increasing acknowledgement that the architecture of lexical retrieval might not be universal across languages. For instance, Levelt et al.'s (1999) model postulates that word form encoding has three levels: (1) the appropriate morpheme form corresponding to the target; (2) segments, and the metrical frame of the target word were retrieved independently; and (3) phonological word, which was produced by attaching or inserting the segments to the structure, was then ready to be articulated. For Chinese speakers, however, syllables rather than segments (or phonemes) have particular prominence. Chinese phonological has relatively few syllables of comparatively simple phonological complexity (compared to English) and has clear syllable boundaries and no re-syllabification. Therefore, O'Seaghdha, Chen, and Chen (2010) proposed that languages might differ in the "proximate unit" of phonological encoding (i.e., the primary selectable unit below the word level, carrying particular salience as a speech planning unit). They found that differential proximate unit between Indo-European languages and Chinese using the "implicit priming" task. In this task, participants first learn small sets of word pairs (e.g., single–loner, place–local, fruit–lotus). Then, in the experiment, the first word in the pair is visually presented (e.g., *single*) and participants were asked to respond with the second word of the learned pair (*loner*). Two types of blocks of trials are compared: homogeneous blocks, where the words to be produced all share the same phonological feature (e.g., the initial segment in *loner, local, lotus*), and heterogeneous blocks, where the words to be produced share no segments. In homogeneous blocks, participants were expected to name words faster, as the shared initial segment has contributed to advance preparation. Studies have documented significant implicit priming (or facilitation) effects based on the initial phonemic segment in Dutch (Meyer, 1991; Roelofs, 2006b), in English (Damian & Bowers, 2003), and in French (Alario, Perre, Castel, & Ziegler, 2007). However, in Mandarin speakers, the facilitation effect was found only when responses to a series of disyllable words shared the same first syllable, but not when they shared only the first segments (J.-Y. Chen, Chen, & Dell, 2002). Furthermore, O'Seaghdha et al. (2010) found that the absence of implicit priming from shared onsets applied even to monosyllable word production. The language difference arises because syllables are proximate units in

Mandarin Chinese, whereas segments are proximate in English and other Indo-European languages (O'Seaghdha et al., 2010).

Additional evidence has come from the phonological facilitation effect in the widely used picture-word interference task with Chinese speakers. When the distractor word is phonologically related to the target object's name, the object naming times are faster than when the distractor is unrelated. In English, phonologically related distractors are often confounded with orthographic similarity, due to the constraints of the orthography. Studies using Chinese words can separate distractors into phonologically related only and orthographically related only. Q. Zhang, Chen, Stuart Weekes, and Yang (2009) observed independent effects of pure phonological facilitation and pure orthographic facilitation. By varying the stimulus-onset asynchrony (SOA) of the words and pictures, they examined if there were interactive effect across SOA, which is assumed to reflect independent processing stages in naming (Damian & Martin, 1999; Starreveld & La Heij, 1996). First, the phonological facilitation effect occurred between SOAs of 0ms and 150ms, whereas the orthographic facilitation effect spanned a range of SOAs from –150ms to 150ms. Second, the magnitude of the phonological effect was smaller than the magnitude of the orthographic effect. The independence of the effects of orthographically and phonologically related distractors were replicated by various studies in Chinese (Bi et al., 2009; Q. Zhang & Weekes, 2009). However, Zhao, La Heij, and Schiller (2012) failed to observe the difference in the time courses of orthographic and phonological facilitation, and they also found that the orthographic facilitation effect was not larger than the phonological facilitation effect at an SOA of 0ms.

Roelofs (2015) recently postulated for Mandarin Chinese phonological encoding the following four levels: (1) a morpheme corresponding to the target is activated; (2) atonal syllable nodes are activated, and, simultaneously, a tonal frame is activated; (3) segments are activated, and finally (4) segments and tonal frames are merged into syllable motor programs. In terms of this framework, phonological encoding for Mandarin speakers involves an additional processing layer compared to Western languages. Figure 1.2 shows the theoretical framework for word phonological encoding in European and Chinese languages summarised by Q. Zhang, Zhu, and Damian (2018). This shows a segmental level for Chinese, despite the findings that segmental overlap has only small and often unreliable effects in Chinese (e.g., Chen et al., 2002).

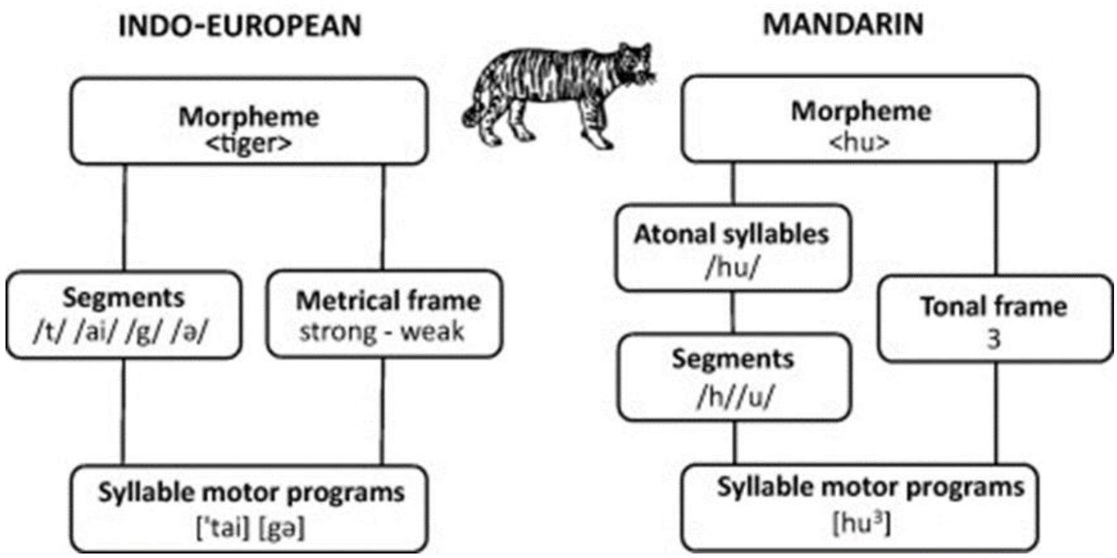


Figure 1.2 Zhang, Zhu, and Damian's (2018) model for phonological encoding in Mandarin Chinese

Q. Zhang et al. (2018) used a semantic blocking (or "cyclic") naming procedure. Semantic blocking is the effect where participants name objects slower when they presented in semantically homogenous blocks of trials (e.g., all vehicles, or all items of furniture) rather than in semantically heterogenous blocks (Belke, Meyer, & Damian, 2005; Damian & Als, 2005; Damian, Vigliocco, & Levelt, 2001). This is explained by the idea that a block of objects from the same semantic category will cause increased activation of their related concepts –and their corresponding lexical representations– that enhances the competition involved in lexical selection. Zhang et al. found that phonologically related distractors facilitated naming times in both English and Chinese (and equally for both semantically homogeneous and heterogeneous blocks of trials). In English, mediated distractors that were phonologically related to a semantic competitor of the target object names (e.g., the word *note* superimposed on the picture of ARM, which is similar to the semantic competitor 'nose') slowed object naming, but only for semantically homogeneous blocks. This is indicative of cascaded processing but limited to when the targets are already pre-activated by semantic context. In Chinese, mediated distractors had no effect for either homogeneous or heterogeneous blocks. Zhang et al. concluded that semantic-to-phonological encoding reflects "weak cascadedness in English, but strictly serial transmission in Mandarin" (p. 840).

4. The Lexical Representation of Homophones

It is clear that heterographic homophones, which have different meanings, must have different semantic representations. It is also clear that they must also have different and separate orthographic lexical representations for both word recognition and orthographic word production. For example, *awe* and *or* are distinct and separate word-forms, and even more similarly spelt homophones (such as *or*, *ore*, and *oar*) are not invariably more similar than non-homophonous words (e.g., *at* and *ate*), and so both the visual word recognition system and the spelling production system must distinguish between them.

Given that homophones have identical pronunciations, it would appear an easy matter to think that they share the same lexical, phonological representation in the speech production system. In the Levelt et al. (1999) model homophones have different lemma representations, which is necessary to indicate their different meanings (in the case of heterographic homophones, such as *none–nun*) and different syntactic classes (in the case of homographic homophones, such as *the nurse–to nurse*). However, Levelt et al. proposed that homophones have the same, *shared* lexeme. This proposal clearly has the merit of an economy of storage. According to this model, homophones (all the words in one homophone family) share one, common lexeme representation, and so the model predicts that factors that affect the ease of accessing, retrieving and producing the shared lexeme from one homophone should also extend to its homophone mate.

An alternative view is that homophones have separate or *independent* phonological representations, which attributes no special status to homophones. This view proposes that lexemes of homophones are represented redundantly (e.g., that there is one representation for *none* and another, separate representation for *nun*. This model predicts that factors that affect the ease of accessing, retrieving and producing the lexeme of one word should not necessarily affect the activation of that word's homophone.

These two theoretical accounts are illustrated in Figure 1.3, taken from Caramazza et al. (2001). The Levelt et al. (1999) serial-discrete model incorporates the view that homophones have a shared lexeme-level representation, a claim also made by Dell's (1990) interactive model. In contrast, Caramazza and colleagues (Caramazza, 1997; Caramazza et al., 2004; Caramazza et al., 2001), Harley (1999), and Jacobs, Singer, and Miozzo (2004) have proposed that all words, including homophones, have separate, independent lexeme-level representations, even though, for homophones, the contents of these representations will be identical.

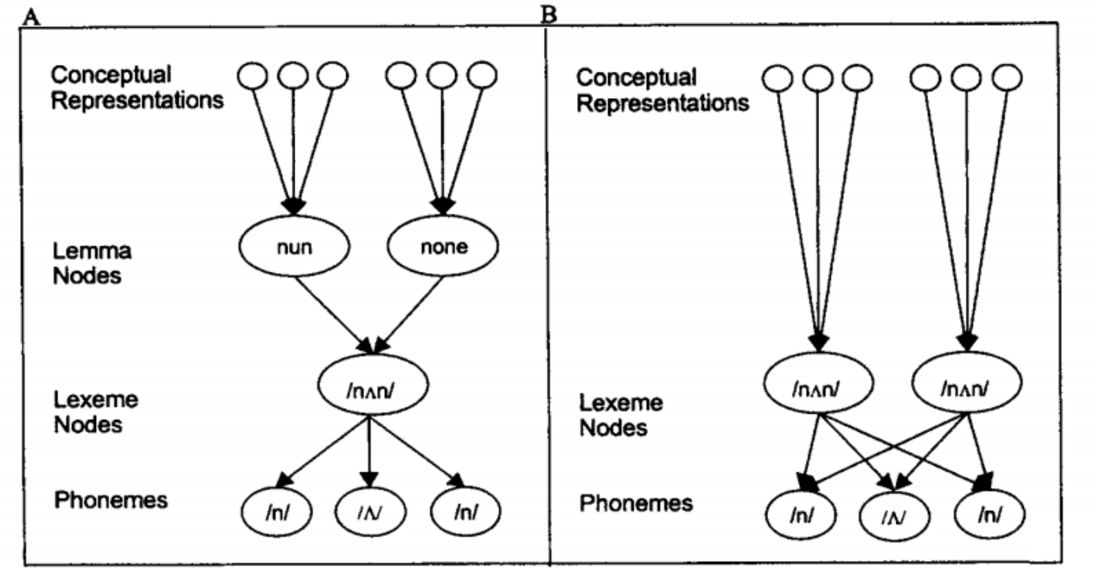


Figure 1.3 Alternative Models of the Phonological Representations of Homophones (Caramazza et al., 2001).
Note: Panel A shows the shared representation hypothesis and Panel B shows the independent representation hypothesis.

4.1 Frequency "inheritance" effects in the production of homophones

The effect of word frequency has played a prominent role in many theories of lexical access (e.g., Levelt et al., 1999). The advantage of high-frequency over low-frequency words has been documented in a multitude of production tasks in normal speakers (e.g., Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Dell, 1990; Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965) and brain damaged speakers (e.g., Kittredge, Dell, Verkuilen, & Schwartz, 2008). However, the precise locus (or loci) of the frequency effect in speech production is still debated. Some researchers argue that word frequency only affects the retrieval of phonological word-forms (e.g., Jescheniak, Hahne, et al., 2003; Jescheniak & Levelt, 1994), whereas others argue that frequency only affects the early stage of lexical access (e.g., Dahan, Magnuson, & Tanenhaus, 2001; Dell, 1990). Behavioural (e.g., Navarrete, Basagni, Alario, & Costa, 2006), neuroimaging (e.g., Graves, Grabowski, Mehta, & Gordon, 2007) and patient studies (e.g., Kittredge et al., 2008; Knobel, Finkbeiner, & Caramazza, 2008) all provide clear evidence for a lexical process sensitive to frequency at all stages. Word frequency is, therefore, an important variable to use in the study of how homophones are accessed and produced in speech.

The shared representation view predicts that the time to produce a lower-frequency member of a homophone pair (e.g., *nun*) would be faster than a non-homophone of the same specific word frequency (e.g., *zip*) and, indeed, should be the same as a word with the same cumulative frequency, that is, the summed frequency of *nun* plus *none* (e.g., *king*). In other words, a lower-frequency homophone should “inherit” the frequency advantage from its higher-frequency homophone mates. This is because the same, common lexeme would be activated both by saying *nun* and by saying *none*.

In contrast, the independent representation view predicts that the time to produce a word will be affected by its own word-specific frequency and so will not necessarily be affected by the activation of its homophone mate (which will have its own, separate, and independent representation). Thus, this view expects no homophone cumulative frequency inheritance, and so saying *none* would not impact upon the frequency of saying *nun.*

The shared representation view has been supported by the results of a study by Jescheniak and Levelt (1994). Dutch-English bilingual speakers were visually presented with English words and were required to produce the Dutch translation as quickly as possible. There were three types of Dutch translation words: (1) Low-frequency homographic homophones that had a high-frequency homophone mate; (2) Non-homophonic words matched to the homophones on specific-word frequency; and (3) Non-homophonic words matched to the homophones on cumulative homophone frequency. As Dutch has a transparent orthography, all the homophones were also homographs. The results showed that mean naming times for the homophones (in condition 1) were about 100ms faster than those for the specific-word frequency controls (in condition 2), and roughly equal to those for the cumulative homophone frequency controls (in condition 3). Jescheniak, Meyer, et al. (2003) replicated this pattern of results with English-German bilinguals. These results support the shared representation hypothesis and suggest that low-frequent homophones inherit the accessing speed of their higher-frequency homophone twins. (1990) finding that a word's susceptibility to phonological errors was determined by the cumulative frequency of the word and its higher-frequency homophone, rather than only by its own specific frequency, also supports this hypothesis.

Caramazza et al. (2001) failed to replicate Jescheniak and colleagues' findings concerning the word translation latencies. They tested a group of English-Spanish bilinguals, who were shown printed Spanish words and were required to produce their English translations. They found that the times to produce lower-frequency homophones (e.g., translating *liebre* into *hare*) were essentially the same as specific frequency-matched control words (e.g., translating *ciruela* into *plum*), 1058 vs. 1060ms, and were slower than the (overall more frequent) cumulative homophone frequency control words (e.g., translating *drool* into *tree),* 852ms.

Caramazza et al. also conducted two picture-naming experiments with objects whose names were: (1) the lower-frequency members of a homophone pair, some of which were heterographic homophones (e.g., *nun*, *pear*) and some were homographic homophones (e.g., *well*, *safe*); (2) specific word frequency matched controls; and (3) cumulative-homophone frequency matched controls. The difference in naming times between the homophones and the specific-word controls, 764 vs. 752ms, was not consistently reliable; it was significant in the analysis by participants but not by items. The difference in naming times between the homophones and the overall more frequent cumulative frequency controls, 764 vs. 714ms, was significant. (Each picture was presented three times in the experiment, and the same pattern of results was found for each presentation.) Caramazza et al. did not report an analysis of any possible difference between the heterographic and homographic homophones. The possibility that the homophone pictures were harder to recognise was excluded by a control experiment where the same stimulus pictures where named by Italian participants, for whom the object names were not homophones, and no significant difference was found between the homophones and their specific-word controls, 767 vs. 784ms. Further, to exclude the possibility that the homophone pictures were somehow easier to articulate, a delayed naming control task showed no significant difference between any of the three conditions.

These results from both translation times and object naming times support the independent representation hypothesis. They also undermine arguments that take the assumed shared representations for homophones to support the lemma-lexeme distinction.

Further, the same object naming results have also been found for Spanish homographic homophones (Cuetos, Bonin, Alameda, & Caramazza, 2010) and for French heterographic homophones (Bonin & Fayol, 2002; Cuetos et al., 2010).

Caramazza et al. (2001) also report an object naming experiment in Chinese, testing 28 Mandarin speakers in Beijing. The homophone names were all heterographic homophones, and all had mono-morphemic names (as represented by one character if written). They used the same design and found that the difference in naming times between the homophones and the specific-word controls of 34ms (783 vs. 749ms) was significant in the analysis by participants but not by items. The difference in naming times between the homophones and the overall more frequent cumulative frequency controls of 66ms (783 vs. 717ms) was significant. (Each picture was presented three times in the experiment, and there was a consistent difference, between 32 and 36ms, between the homophones and their specific-word controls for each presentation.) As in the results for the object naming experiment in English, the homophone picture names were not easier or harder to articulate as a delayed naming control task showed no significant difference between any of the three conditions. These results are very similar to those found in English, but the numerical difference between the means times to names homophones and specific-word frequency-matched controls was somewhat larger in Chinese (34ms) than in English (12ms). However, overall, they support the independent representation hypothesis for Chinese words, at least for the mono-morphemic homophones tested, and as discussed above mono-morphemic words constitute a minority of words in standard Chinese.

In the picture-word interference task, the distractor frequency effect is the finding that semantically unrelated low-frequency words produce longer picture naming than high-frequency words (e.g., Dhooge & Hartsuiker, 2010, 2011). Miozzo and Caramazza (2005) found that this effect was also found for homophones: naming times were slower when the distractor words were low-frequency homophones compared to high-frequency homophones. Although this result may result from processing operating at the stage of the visual word recognition, if it is attributed to phonological processes, then it offers no support for the shared representation hypothesis. Homophone processing appears to vary as a function of the frequency of the individual homophonic forms.

Jacobs et al. (2004) report a cognitive neuropsychological study of how homophones are represented. They found that an anomia patient's accuracy for naming low-frequency homophones was significantly lower than that for naming high-frequency homophone words, which supports the independent representation hypothesis. Caramazza and Hillis (1991) looked at response accuracy for verb/ noun homographic homophones in written and spoken modalities. They found a double dissociation between two aphasic speakers, SJD and HW, in written and spoken naming performance for word category and modality. The authors interpreted this result as evidence for independent representations of homophones. Biedermann, Blanken, and Nickels (2002) presented a treatment study with a man who had global aphasia and severe anomia using homographic homophones in German. They tested the patient's naming of homographic homophones (e.g., *ball*, a large formal gathering for social dancing, and *ball*, a spherical or ovoid body used in a game or sport). Treatment was intensive picture-naming training for only one word of the homophone pair with exclusively phonological cues. The outcome of this phonological treatment showed significant improvement of treated and untreated homophones, which supports the shared representation hypothesis. Additionally, they replicated and expanded these findings by showing evidence that heterographic homophones benefit to the same extent as homographic homophones in English (Biedermann & Nickels, 2008a, 2008b).

4.2 Word-frequency effects on the articulation duration of spoken homophones

The evidence reviewed so far mainly comes from the analysis of word production latencies, naming time taken to initiate a spoken word. Spoken word latencies, in tasks such as object naming and word translation, have been centrally important in the study of the major process of lexical selection in the production of spoken language. However, another source of evidence that is relevant to the study of the lexical representation of homophones comes from the smaller number of studies of the word production (or articulatory) durations.

A reduction effect on word articulation (i.e., shorter spoken durations of words) has been observed for words that are of higher frequency (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Fidelholtz, 1975; Hooper, 1976), repeated within a discourse (Fowler, 1988), or contextually predictable (Cohen Priva, 2015; Seyfarth, 2014). The phonetic reduction includes not only durational shortening but also consonant deletion (Cohen Priva, 2015) and vowel centralisation (Fidelholtz, 1975). These phenomena are usually described as the probabilistic hypothesis: words with a generally higher probability of being produced are shortened or reduced, and, conversely, lower frequency and lower probability words tend to be lengthened. The intelligibility-based explanation suggests that speakers adjust their speech so as to maximise the intelligibility of words that might otherwise be difficult to recognise. In contrast, the production-based explanation suggests that these spoken word durations reflect the speed of lexical access, retrieval, and encoding in word production. It is hard to differentiate these two approaches as they often make identical predictions, as high-frequency words are easier to produce and easier to recognise than low-frequency words.

Phonological neighbourhood density has been shown to be a lexical variable that affects recognition and production differently. A word's phonological neighbours are other words whose pronunciations differ by the deletion, insertion or substitution of one segment (e.g., Luce, Pisoni, & Goldinger, 1990). Words with many neighbours are recognised more slowly and less accurately than words with only a few neighbours (Paul A Luce & Pisoni, 1998). However, words with more neighbours are produced in speech more quickly (e.g., Vitevitch, 2002). According to intelligibility-based accounts, words from dense phonological neighbourhoods should have longer durations, as they are easily misunderstood. The opposite effect would be expected according to the production-based accounts. Gahl, Yao, and Johnson (2012) examined words duration in conversational speech and found that words from the dense phonological neighbourhood were shorter and contained more centralised vowels than words from sparse phonological neighbourhoods when other phonological variables were controlled. These findings support production-based accounts that claim that differences in word durations reflect the time course of word retrieval and encoding in speech production. Gahl et al. (2012) also suggested that such a reduction might actually be stored in lexical phonological representations. This idea finds some support in the study by Seyfarth (2014) who found that usually predictable words are reduced even when they occur in unpredictable contexts.

Gahl et al. (2012) claim that the shorter durations of high-frequency words reflects the ease of retrieval for speech production, and the reduced form is actually stored in lexical phonological representations. Therefore, the study of articulation duration may shed light on the representational status of homophones. Under the shared representation hypothesis, lower frequency homophones should inherit any phonological reduction from their higher-frequency homophone mates; their common lexeme should enjoy the same phonetic reduction benefit. The independent representation hypothesis would not expect any such inheritance and would expect that each homophone would have its own independent phonological length, as determined by its own word frequency.

Gahl (2008) analysed the articulation durations of 223 homophone pairs, represented by roughly 80,000 tokens, from a large corpus of tape-recorded telephone conversations. She found that the average durations of higher-frequency members of homophone pairs were significantly shorter than their lower-frequency counterparts (368 vs. 396ms on average). The frequency effect remained significant when local speaking rate, predictability from neighbouring words, position relative to pauses, syntactic category, and orthographic regularity were statistically controlled. These results suggest that the specific-word frequency, and not homophone cumulative frequency, affects the duration of spoken word production.

If we assume that the effect of word duration arises from the storage of lexical phonological word-forms, then the fact that the time to actually articulate higher-frequency homophones (e.g., *time*) is reliably shorter (by 22ms on average) than lower-frequency homophones (e.g., *thyme*) appears to support the independent representation hypothesis. Further support has been provided by a recent study by Lohmann (2017) who analysed the phonological durations of homographic homophones, such as *cut* as a noun and *cut* as a verb. He found that homophones durations were related to the frequency of each word's separate sense and did not show an inheritance from the other sense.

In summary, the evidence for the claim that homophones have shared lexical phonological representations comes from: (a) the frequency inheritance effect in translation times (Jescheniak & Levelt, 1994; Jescheniak, Meyer, et al., 2003); (b) the finding that remediation of specific words in anomic patients generalised to homophones of the treated words (e.g., Biedermann et al., 2002).

However, Caramazza et al. (2001) did not replicate the frequency inheritance effect on translation times, and they (and others) have also found that there was no frequency inheritance effect on object naming times. These results support the claim that homophones have independent lexical phonological representations. There has been relatively little research on the remediation of naming accuracy in anomic patients and on the articulatory durations of homophones. Gahl's (2008) finding that higher-frequency homophones have shorter articulatory durations than lower-frequency homophones also supported the independent representation hypothesis. The experiments reported in this thesis are intended to further explore how homophones are represented in Chinese, a language that has very many more homophones than in English. The experiments reported in Chapter 2 will record the times adult readers take to read aloud Chinese words and will also record the articulatory durations of these reading responses. This combination of word reading times and durations for reading Chinese words has not been studied before and so represents a novel program of research.

5. Word Reading

There exists a large body of research into the processes involved in oral reading (i.e., reading aloud written words), much of which has been conducted in English and in other European languages. Unlike Chinese, these languages have alphabetic orthographies and also vary in the regularity or consistency of their spelling-to-sound relationships. English orthography has rather inconsistent spelling-to-sound relationships, and there exists many "irregular" (or "exception") words; for example, *pint* (compared to hint, mint, flint, etc.), *have* (compared to cave, save, pave, etc.), and unambiguous exceptions such as *yacht* and *colonel*. Both the experimental and the neuropsychological study of reading regular and irregular words, and non-words, in alphabetic orthographies has led to the development of the "dual-route" model of reading, which contains separate routes for semantic-lexical word reading and sub-lexical assembled phonological recoding.

5.1 Dual-route models reading in English

The "dual route cascaded" (or DRC) model (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001) shown in Figure 1.4 is a computational realization of the dual-route framework, developed to account for both experimental studies of skilled adult reading and cognitive neuropsychological studies of patients with acquired disorders of reading. The lexical semantic route involves a written word being recognised by the visual recognition system, and its phonological form being retrieved from the phonological lexicon. The assembled phonological recoding route operates by applying a set of grapheme-to-phoneme correspondence rules, and this enables the reading aloud of non-words (although it can also assist word reading, but only for regular words). There also exists a direct lexical but non-semantic route that connects visual word recognition to the phonological output lexicon.

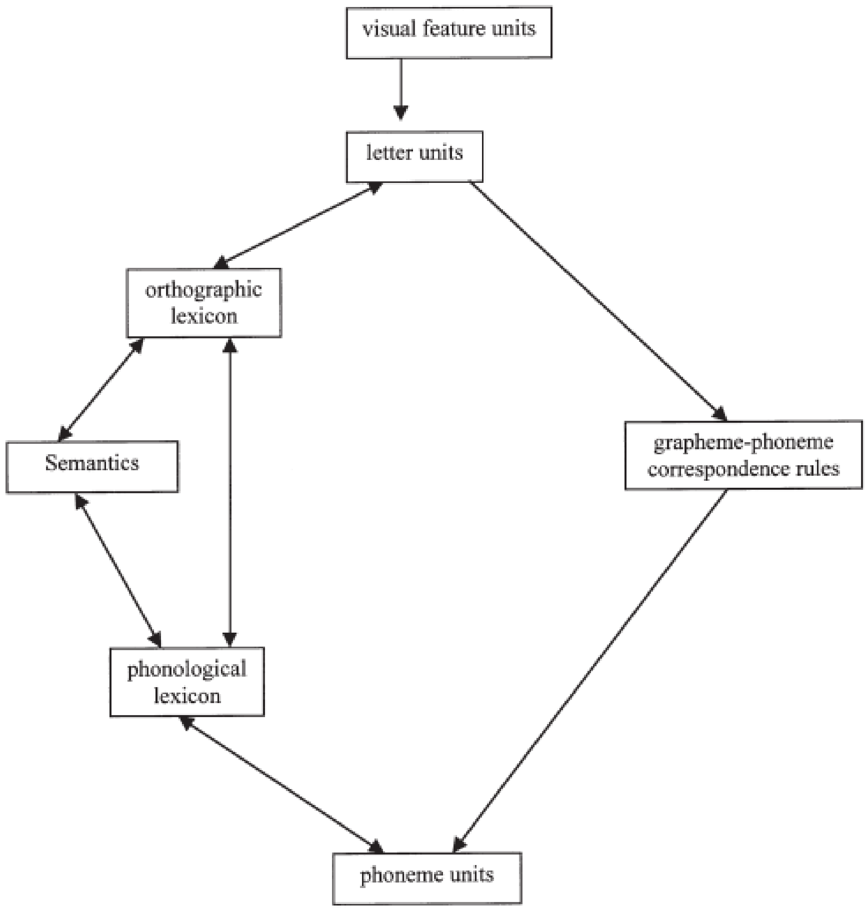


Figure 1.4 The Dual-Route Cascaded model of word reading (from Coltheart et al., 1993)

Impairment of the assembled phonological recoding route would result in phonological dyslexia, where patients show impaired nonword reading but intact word reading (e.g., Beauvois & Dérouesné, 1979). Impairment of visual word recognition would result in an over-reliance upon assembled phonological recoding producing surface dyslexia (Patterson, Marshall, & Coltheart, 1985; Zevin & Balota, 2000)), where patients make regularisation errors to irregular words (e.g., reading *pint* as "pin-t"). Impairment within the lexical-semantic route would result in deep dyslexia (Patterson, Coltheart, & Marshall, 1980), where patients produce semantic errors in reading aloud, e.g., reading *yacht* as "boat"). An impairment of the direct lexical non-semantic route would result in patients being able to read irregular words correctly but without comprehension(Blazely, Coltheart, & Casey, 2005; Funnell, 1983).

"Direct" acquired dyslexia (i.e., reading aloud without semantics) has also been reported in Japanese patients (Sasanuma, 1980). Patients with Alzheimer's disease were able to read aloud words presented in Kanji (a script based on Chinese logographs) despite being unable to comprehend their meaning (as indicated by semantic categorisation tasks). As the reading of Kanji words is assumed to be a purely lexical process (that cannot be supported by sub-lexical phonological recoding), the fact that Sasanuma's patients were able to read them aloud without accessing their semantics suggests that there exists a direct and non-semantic reading route. However, it is important to note that Kanji differ from the characters used in Chinese. In particular, Chinese characters have, with only rare exceptions, only one pronunciation, whereas characters in Japanese Kanji usually have two pronunciations: An On-reading, which is based on Chinese from which these characters were historically derived, and a Japanese Kun-reading (although sometimes there may be several On-readings). Japanese readers must decide which reading should be employed using the context of the particular character. Regarding the phonetic component in Kanji (i.e., the structural element which may provide the reader with a clue to how the character is to be pronounced), this is not as reliable as that in Chinese characters.

According to the DRC model, processing along the lexical and non-lexical routes occur simultaneously and converges on the phonological system. This model can explain reading times of skilled readers to particular sets of stimuli (e.g., that reading times for regular words are faster than for irregular words, as the two routes will generate same pronunciation). Newcombe and Marshall (1980) speculated that in normal readers the non-lexical route acts as a check on the output of the lexical route (e.g., if the lexical-semantic route generated both ‘small’ and ‘little’ as possible responses to *small*, the phonological recoding route would determine the correct response).

5.2 Word reading in Chinese

The evidence reviewed so far mainly comes from the analysis of word reading in alphabetical orthographies, and especially English (which has highly inconsistent spelling-to-sound correspondences). Chinese orthography does not use an alphabet, and its “logographic” characters typically have unambiguous phonological correspondences.

There are some studies of both acquired and developmental disorders of reading in Chinese that have been interpreted within dual-route models of Chinese reading. Such models propose a distinction between a lexical route that converts a whole written word with its complete pronunciation, and a sub-lexical route that utilises the phonological components of a written character. W. Yin and Butterworth (1992) reported the study of 11 Chinese brain-damaged patients whose selective acquired reading disorders, they argued, show Chinese "analogues" of deep dyslexia (due to an impaired sub-lexical route) and surface dyslexia (due to an impaired lexical route). However, the role of any sub-lexical route in Chinese is open to some debate. It is widely claimed that reading Chinese words is an essentially print-to-meaning process, with the assumption that characters contain no consistently reliable non-lexical information as to their pronunciation. About 81% of modern characters consist of a phonetic component which can provide some clue as to how the character is to be pronounced and a semantic radical which can be quite informative in suggesting the character's meaning (Y. Zhou, 2003). A phonetic component, in isolation, is usually a normal character that stands for a word or morpheme, and so has an associated pronunciation. However, the phonetic component often does not have the same pronunciation as a compound character that contains it. Due to historical reasons, only 36% of phonetic components have exactly the same pronunciation as the characters that contain it; 48% of phonetic components have a similar pronunciation; and 16% of phonetic components have no relationship with the pronunciation of the characters at all. The classification of Chinese characters into "regular" and "irregular" (Stone, Vanhoy, & Orden, 1997) on the basis of the phonological congruence of the pronunciation of the phonetic component and the whole character is therefore problematic. For "irregular" characters, the pronunciation can be very different from what may be predicted by its phonetic component, and this leads to a tendency for inexperienced speakers to read the character according to a legitimate but wrong pronunciation. Weekes and Chen (1999) called errors of this type "legitimate alternative reading of components" (or LARC errors), rather than regularisation" errors as referred to by Yin and Butterworth. LARC errors involve the incorrect pronunciation of character that is nevertheless approximate to other characters containing the same component. For example, 清 (qing4) is a low-frequency character, and people who do not recognise it may pronounce it as 请 (qing3) or 青 (qing1) or 情 (qing2) or even 猜 (cai1), as all these high-frequency characters share the same phonetic component. Moreover, phonetic components typically take the same position in all characters. For example, 里 (li3), always appears on the right (理, 鲤, 俚, all pronounced li3). This rule also enables the invention of pseudo-characters by putting the phonetic component in its legal position. By using pseudo-characters and comparing "regular" and "irregular" characters, Butterworth and Yin (1992) identified the Chinese analogues of deep dyslexia (who made semantic errors in reading) and surface dyslexia (who made LARC errors when reading "irregular" characters). The idea that reading Chinese depends on two distinct routes, a lexical-semantic route and a sub-lexical route, has been supported by other studies of acquired dyslexia patients (Weekes & Chen, 1999; Yin & Weekes, 2003) and developmental dyslexic children (Shu, Meng, Chen, Luan, & Cao, 2005; L.-C. Wang & Yang, 2014).

In conclusion, cognitive neuropsychology data suggest that reading Chinese, like reading alphabetic scripts, involves two distinct routes: one that associates a whole written word with its complete pronunciation, and one that utilises the phonological correspondences of the phonetic components of characters. Each route may be selectively impaired by brain damage, resulting in different patterns of reading disability. Yet, there is no convincing evidence for a direct lexical non-semantic reading route, although discovering patients with dementia (and so impaired semantics) with intact reading of "irregular" characters would suggest that this route would exist. Further, it is yet to be established convincingly that sub-lexical phonological correspondences play a substantive role in the time to read aloud or comprehend words by skilled, adult readers of Chinese.

6. Plan of the experimental work reported in this thesis

This thesis will examine Chinese word production to address two major theoretical issues: (1) The nature of the lexical phonological representations of the very many heterographic homophones in Chinese; and (2) How activation is transmitted from lexical to phonological levels in spoken Chinese.

Chapter 2 reports two experiments that examine both reading times and articulation durations of homophones and matched non-homophone Chinese words. Two classes of homophones will be compared: homophone twins, where there are only two words sharing the same pronunciation (an English example would be *pair* and *pear*); and homophone families, where there are three or more identically pronounced words (English examples would be *rain*, *reign*, and *rein*, and *oar*, *or*, *ore*, and *awe*). These studies were designed to arbitrate between the shared representation and the independent representation hypotheses of the lexical representation of homophones hypotheses.

Chapter 3 reports two experiments that examine the priming of object naming times by the prior reading (either reading aloud or reading silently) of a homophone of the target object name or a phonologically related word. (English examples would be reading *awe/or*, and *paw/poor*, on naming a picture of an OAR.) The study of priming effects should help to shed light on the temporal dynamics of activation flow in the lexical selection and phonological encoding in Chinese spoken word production by investigating the persistence of activation at processing stages common to both word reading and object naming.

Chapter 4 reports four experiments that examine object naming times by the prior oral or silent reading of a word related to the target name through some intermediary word. The investigation of such *mediated* priming will further the study of the flow of activation in Chinese word production. The experiments will examine both semantic-to-homophone priming (an English example might be *either* –> OAR, via 'or') and homophone-to-semantic priming (e.g., *hair* –> RABBIT, via 'hare').

Chapter 5 will report four experiments that examine homophones in primed object naming in Chinese-English bilinguals. Participants will explicitly translate (or simply read) printed words whose translations are homophones of target object names or are the target names themselves. In proficient bilinguals, the act of translating a word into another language is semantically driven, and so involves the same lexicalisation process as in picture naming and spontaneous speech. It is possible that translation also involves activating possible direct lexical level interconnections (Dylman & Barry, 2018). Would producing a word as a translation prime the naming of an object whose name is a homophone? (An English example would be producing the word "boy" as a translation, and then naming a picture of a BUOY.)

CHAPTER 2 READING ALOUD CHINESE HOMOPHONES

Homophones have been investigated in reading English in a number of studies. Most have examined homophones in visual word recognition tasks, especially the lexical decision task. Rubenstein, Lewis, and Rubenstein (1971) were the first to show that low-frequency homophone words were recognised slower that non-homophones, but there was no difference between high-frequency homophones and controls, although Coltheart, Davelaar, Jonasson, and Besner (1977) did not replicate the disadvantage for low-frequency homophones. Pexman and colleagues have reported slower lexical decision times to homophones (Pexman, Lupker, & Jared, 2001; Pexman, Lupker, & Reggin, 2002).

Despite the findings of a homophone disadvantage in the lexical decision task in English, several studies using Chinese stimuli have shown that homophones are responded to more rapidly than non-homophonic controls. Ziegler, Tan, Perry, and Montant (2000) found characters with a high phonological frequency were processed faster than characters with a low phonological frequency in both lexical decision and naming tasks. A phonological frequency effect (of 31ms) was observed, with reading latencies to characters of high phonological frequency being faster than to characters of low phonological frequency when the number of homophone mates was controlled. H.-C. Chen, Vaid, and Wu (2009) conducted both lexical decision and reading tasks using homophonic characters with matched radical variables. Even though they failed to replicate the phonological frequency effect found by Ziegler et al. (2000), they found a clear facilitative effect of homophone density. Moreover, in an auditory lexical decision experiment in Chinese using an almost exhaustive set of more than 1,200 monosyllable morphemes, Yao and Sharma (2017) found facilitation effects of homophone density and phonological frequency for response accuracy, and that phonological syllable frequency was predictive of the speed of lexical decision.

Homophone effects in word recognition in English appear to be quite different from those in Chinese. A processing disadvantage has been observed for English homophone recognition, but a processing advantage is seen for Chinese monosyllables. Most of the studies of Chinese have focused on monosyllabic homophones. An interesting question is whether the processing of Chinese disyllable homophones produce similar effects, and, more generally, whether homophones are processed differently in English and Chinese.

In studies of reading aloud English words, Pexman et al. (2002) found that low-frequency homophones were named slower (and less accurately) than non-homophone control words, though the difference was not significant. Edwards, Pexman, and Hudson (2004) found a small but significant processing disadvantage for homophones, which were named slower than non-homophone controls both in a standard word reading task and in one in which many of the stimuli were non-words (presumed to selectively engage sub-lexical phonological recoding). In their study, homophones and control words were matched for word frequency, initial letter, length, and orthographic neighbourhood size, but not for morphological complexity (e.g., *aide* was matched with *adds*), number of syllables, or phonological onsets (e.g., *knight* was matched with *killer*). Biedermann, Coltheart, Nickels, and Saunders (2009) pointed out that Edwards et al.'s small homophone effect could have been due to artefacts of phonological onset, as the homophones and controls differed in the number of multi-consonant onsets. Biedermann et al. (2009) compared the oral reading of twenty-five irregular low-frequency homophones, whose homophone mates were higher in frequency (to maximise any possible homophone effect), and twenty-five irregular non-homophonic controls. All words had irregular spelling-to-sound correspondences, in order to minimize any contribution from non-lexical, grapheme-to-phoneme conversion. The two sets of words were also matched on frequency, the within-word position of the spelling-to-sound irregularity, number of letters, bigram frequency, age of acquisition, and number of phonological neighbours. Further, all words were both monosyllabic and mono-morphemic. Biedermann et al. found that naming latencies to homophones were significantly slower (and less accurate) than controls in immediate naming, but the effect disappeared in delayed naming. The homophone disadvantage effect has been confirmed by analysis of the data from the English Lexicon Project (Balota et al., 2007). Ziegler, Montant, and Jacobs (1997) found that words with phonological bodies that map onto more than one spelling (e.g., / i:f / as in *beef, leaf* or *thief*) were named slower than words whose phonological bodies are spelt in only one way (e.g., / uk / as in *duck*). This result is relevant because sound-to-spelling variability provides the means to distinguish many heterographic homophones in English (e.g., *tea*, *tee*; *beat*, *beet*).

The homophone disadvantage effect found in the lexical decision task could be explained with a fully interactive model in which homophones share the same phonological representation. However, the homophone inhibition effect observed in reading aloud could be explained by neither the shared representation hypothesis, which would predict a frequency inheritance effect, nor by the independent representation hypothesis, where no difference between homophone and non-homophones were assumed. Additionally, attempts to simulate the homophone disadvantage using the computational dual route cascaded (DRC) model, with either independent or shared representations of homophones, have been unsuccessful (Biedermann et al., 2009).

Currently, there exits very little research on the lexical representation of homophones in Chinese spoken word production. Caramazza et al. (2001) studied the naming of objects whose names were monosyllabic Chinese heterographic homophones; English examples would be BUOY, HARE and OAR. They found that naming latencies to objects with homophone names (mean = 783 ms) were not any faster than to specific-word frequency controls (mean = 749 ms) —in fact, it was slower— but were reliably slower than to cumulative frequency controls (mean = 717). These results replicate the pattern found in their study of object naming in English. As both naming in English and Chinese show no frequency inheritance effect from higher-frequency homophones, these results support the independent representation theory of the lexical organisation of homophones. However, there were some shortcomings of Caramazza et al.'s Chinese experiment. Inspection of their stimulus materials showed that the words used in their two control groups actually were also homophones (although not of the object names). This is not particularly surprising given that monosyllable words in Chinese tend to have many homophones. The pictures used all had monosyllabic names, but these names may not have been the most common name. For example, Caramazza et al. used the word 鼠 (meaning rat) as a monosyllabic homophone, but when naming a picture of a RAT, people often prefer to use the name 老鼠, which is a derivative of 鼠. This is because the everyday understanding of a "word" is that it consists of two or more syllables. Also, from a prosodic perspective, a Chinese word must contain at least one foot, and every foot must be bimoraic or disyllabic. By transitivity, then, a prosodic word must contain at least two moras or syllables (Feng, 1995, 2001). Moreover, the word 老鼠 is not a homophone and the specific-word frequency of 老鼠 and 鼠 is different (0.00030% and 0.00023%, respectively).

Wong and Chen (2008) report additional evidence that sheds light on the phonological representation of Chinese homophones. They conducted a series of picture-word interference tasks, in which a target picture and a distractor word were presented simultaneously or sequentially. The target picture names and the distractors shared the same lexical tone or sub-syllabic components (onset and rhyme), or the same syllable, in various combinations. They found a facilitation effect when the targets and distracters shared both the same rhyme and tone, and when they shared the same syllable. The strongest facilitation effect was observed in the homophone condition. These results were explained within an interactive model, where activation is fed-back from phonemic units to all lexical nodes that share the similar segmental units. The homophone effect found by Wong and Chen (2008), if attributed to activation feedback, could be explained by both the shared and the independent representation hypotheses. However, the target names were monosyllabic words in Cantonese, and there was no non-homophone control condition in their study.

The empirical evidence relating to the lexical phonological representations of Chinese homophones has been mixed. The experiments reported in this chapter will examine the reading aloud of homophones and specific-word frequency matched non-homophones in Chinese. The purpose of these experiments was to arbitrate between the shared representation and the independent representation hypotheses of the lexical representation of homophones in the speech production system.

If homophones have a common, shared lexical phonological representation, then it would be expected that homophones should inherit a processing advantage from their higher-frequency homophones mates. This leads to the specific predictions that (1) homophones should be read aloud faster than non-homophones matched on specific word frequency, and (2) that the difference between homophones and frequency matched non-homophones should be larger for low-frequency than for high-frequency homophones. These predictions were tested in two experiments examining word reading-times, recording both reading latencies and the articulation durations of responses.

Reading Chinese words by skilled, adult readers is very likely to be a lexically and semantically-mediated process, as there is no clear evidence to suggest that sub-character, phonetic component to pronunciation correspondences play a major role. There were also practical reasons for studying word reading rather than picture naming (as was tested by Caramazza et al., 2001). Unlike the lexical decision task, which is primarily based on the activation of the orthographic units, reading aloud involves the full retrieval of representation from the phonological lexicon. Although picture naming is a good way to investigate language production, it is restricted to a relatively small set of potential stimuli. The number of homophones that can be depicted ambiguously is low. Moreover, variables that affect pictured objects (and also naming latencies), such as name agreement (Vitkovitch & Tyrrell, 1995), viewpoint (Gomez, Shutter, & Rouder, 2008) colour of the pictures (Uttl, Graf, & Santacruz, 2006), and so on, are often difficult to control. In many cases, when one homophone can be depicted with high name agreement etc., its mate is often “unpicturable”. Therefore reading aloud was chosen as a practical but reliable means to investigate word production in Chinese.

Chinese has a large number of homophones, which can also have large phonological neighbourhood sizes. On average the same pronunciation can be generated by 11 different characters in Chinese. The existence of a large number of heterographic homophonic characters in Chinese compared to English may have implications for how they are represented in the speech production system. Two classes of homophones will be compared: homophone *twins*, where there are only two words sharing the same pronunciation (an English example would be *pair* and *pear*); and homophone *families*, where there are three or more identically pronounced words (English examples would be *rain*, *reign*, and *rein*, and *oar*, *or*, *ore*, and *awe*). The purpose of this comparison is to both assess the generality of the results obtained and to investigate that possibility that these two classes of homophones differ in their lexical representations.

The experiments will record both word naming latency, articulation duration and average intensity. Gahl (2008) found that, in her analysis of a corpus of tape-recorded telephone conversations, the articulation durations of higher-frequency members of homophone pairs were significantly shorter than their lower-frequency counterparts (by a mean difference of 28 ms). These results suggest that it is specific-word, and not homophone cumulative frequency that affects the duration of spoken word production.

The experiments will extend this work in two respects. First, it will examine high and low frequency homophone and matched non-homophone control words in a factorial experimental design. The possibility that other lexical factors, for instance, visual complicity and character frequency, may be responsible for the differences between homophones and non-homophones was explored in additional linear mixed effects models. Second, it will extend possible differences between homophones and non-homophones in duration to another aspect of articulatory features, namely intensity (or acoustic intensity) that is perceived as the loudness of sound. The difference in acoustic intensity between homophones has not been investigated so far.

"Intensity" may be understood in terms of the level of "loudness" of a spoken form. The difference in acoustic intensity between homophones has not been investigated so far. Damian (2003) argued that articulation duration, as a purely quantitative measure, is possibly too insensitive to reveal articulation processing. The phonetic cues related to prominence include duration, fundamental frequency, and intensity (Fry, 1954). Kochanski, Grabe, Coleman, and Rosner (2005) found intensity to be a good predictor of prominence in an English corpus, and duration and intensity changes may compensate for each other. Therefore, intensity may add more information to study the process of articulation. If homophones have independent representations, then it is possible that they will differ from specific-word frequency matched non-homophones in both duration and intensity.

In summary, the experiments reported in chapter 2 were designed to further explore how homophones are lexically represented in the Chinese speech production system, and builds upon the work of Caramazza et al. (2004) and Gahl (2008). Experiments 1 and 2 will build upon existing work by extending research in three respects: (1) by the study of word production in Chinese, a language that has very many more homophones than in English; (2) by the study of word reading latencies rather than only picture naming (as Chinese word reading is a semantically-driven process); and (3) by the study of both the duration and intensity of the articulatory reading responses. It is the combination of these three features that represents the novel contribution of this research.

Experiment 1 Reading Aloud Homophones from Homophone Families

This experiment will assess the role of word-frequency on the latency, duration and intensity of oral reading responses to disyllabic Chinese heterographic homophones and matched non-homophone control words in a laboratory study. There were three levels of word frequency (high, medium, and low) and two levels of word type (homophones and controls), making a 3x2 factorial design.

1. Method

1.1 Participants

A group of 24 undergraduates (13 women; mean ages, 20.5 years; aged between 18 and 25 years) were recruited from the University of Essex. All were native Chinese speakers and were paid for their participation. Prior to this experiment, an online survey was used to collect subjective judgment about the words' familiarity and concreteness.

1.2 Stimulus words

Given that bisyllabic words are the most common word type in contemporary Chinese (Mandarin), this experiment examined only two-character words. Chinese dictionaries include about 3,200 pairs/cohorts of homophones, which consisted of 6,900 words. The majority of these words have two-characters (and so two syllables), although there are a few three-character words. Since the difference of word frequency is the main focus of this study, it is important to employ a representative estimate of the daily language exposure and capture the variation within word production. In this study, the measure of word frequency was taken from the SUBTLEX-CH corpus (Cai and Brysbaert (2010), which sampled 46.8 million characters (from 33.5 million words) in film and television subtitles. It has been found that, in English, French and Dutch, word frequency based on film and television subtitles is more valid than traditional samples taken from books and printed text (Cai & Brysbaert, 2010). Cai and Brysbaert also found that the log10 of the total number of times the character has been observed in the corpus was the most significant frequency predictor ($p<0.001$) among other available sources of Chinese word frequencies; for example, it explained 25.2% of the variance of reaction times in a lexical decision task. Therefore, we first selected homophone cohorts that (a) have distinct word-frequencies, (b) have neither semantic relation nor association, and (c) are included in the SUBTLEX-CH corpus.

A total of 29 sets of homophone cohorts were selected, which made up to 87 homophones. They all have two or more than two homophone mates. Each homophone word was paired with a non-homophonic control word, which has same or similar syllable to the homophones (e.g., homophone 负责, meaning responsible, /fu4ze2/, was paired with 否则, meaning otherwise, /fou3ze2/). The homophonic control words were also matched on word frequency.

For each homophone cohort, homophones were divided into higher frequency homophone (HHF), medium frequency homophone (MHF) and lower frequency homophone (LHF); and their control words were also divided as higher-frequency non-homophones (HNF), medium frequency homophone (MHF) group and lower-frequency non-homophones (LNF). These six sets of words were used in the homophone family study of Experiment 1. Based on previous studies, the following lexical characteristics of the words were collected to investigate whether there were differences across word conditions. These variables were grouped into character level and word level, as it has been suggested that word processing could be affected by the properties of the component of the characters (Taft, Zhu, & Peng, 1999; Yan, Tian, Bai, & Rayner, 2006). If the variables were not matched within groups, they would be entered into a mixed effect model to explore the contribution of their effect on word production.

Lexical variables.

Word-form frequency. As mentioned earlier, measures of word frequency were taken from Cai and Brysbaert (2010) SUBTLEX-CH corpus. Rather than using the raw word count measures in the corpus, the logarithms (base 10) of the word counts were calculated, since the distribution of word frequencies trend to follow Zipf's law (Zipf, 1949), and are generally skewed. In addition, word frequency was Laplace-transformed to deal with words not observed in the corpus (Brysbaert & Diependaele, 2013); Laplace transformation involves correction of the corpus size and assumes that the theoretical corpus size equals the number of word tokens plus the number of word types. The total word count of the corpus is 33.5 million with 99,121 different words. A constant of 7.5 was added to make all values positive. The following equation is needed to calculate the Zipf values by the frequency counts of the total corpus:

$$\text{Zipf} = \log 10\left(\frac{frequency\_count + 1}{total\ word\ count + total\ word\ types}\right) + 7.5$$

Homophone density and phonological frequency. The control words selected had a homophone density of one, and their phonological frequencies were equal to their word-form frequencies. For words in homophone pairs, the homophone density was two, while, for words in homophone families, the density is three or more. The phonological frequencies of homophones were the logarithms (base 10) of the cumulative word counts of the homophones when added in the mixed-effect models.

Familiarity and concreteness/imageability. Familiarity and concreteness/imageability have been examined in previous studies in Chinese. These two variables were obtained by subjective ratings using two separate online surveys. Thirty-eight participants (8 males) with a mean age of 23.1 took part in the familiarity rating, and 24 Chinese speakers (8 males) with a mean age of 23.3 provided the concreteness/imageability ratings. In these online surveys, participants were asked to rate the words on familiarity or concreteness/imageability on 7-point scales, and the presentation order of words was randomised for each participant.

Grammatical category and number of meaning. It has been demonstrated that the semantic precision of Chinese characters affects word recognition and activation of meaning (Tan, Hoosain, & Siok, 1996). Usually, single Chinese characters have vague meanings, and two-character words have precise meanings. However, some words could be used in more than one grammatically category. The characters were divided into verbs, nouns, and others based on their main grammatical category in SUBTLEX-CH-WH_POS.

The Number of strokes. Chinese characters are constructed from radicals, which are composed of single or multiple strokes. The number of strokes in a Chinese character may be seen as an index of visual complexity. For example, 青 has eight strokes, while 精 has 14 strokes. The total number of strokes was considered as a word-level variable, while numbers of strokes of the individual characters was a character-level variable.

Character variables.

Character frequency. The Zipf scale of character specific frequency was based on the word counts on the SUBTLEX-CH-CHR corpus with a Laplace transformation. The total character count of the corpus is 46.84 million with 5,936 different characters.

Phonological frequency and homophone density. In Chinese, most characters do not have a unique pronunciation, that is to say, they are homophones (same syllable and same tone). Homophone density refers to the number of homophonic characters. Phonological frequency is defined as the total word counts of the characters that have the same pronunciation. Some Chinese characters may have more than one pronunciation, and for these, the calculation of phonological frequency and homophone density were based on their most common pronunciations. The phonological frequency was log-transformed when adding in the mixed effect models.

Regularity and consistency. The majority of Chinese characters are compounds, consisting of a phonetic element, which gives a cue to the pronunciation of the character, and a semantic radical, which provides a generally more useful cue to the meaning of the whole character. Based on whether the pronunciation of the phonetic compound is consistent with that of the character, phonetic compounds can be divided into two categories: regular and irregular. For example, the character 蜻 means “dragonfly” and is pronounced as /qing1/ in Pinyin (Chinese phonetic system). It consists of a semantic radical (虫) on the left, which means "insect," and a phonetic radical (青) on the right, which is pronounced the same as the character itself/qing1/. However, sometimes phonetic radicals may have same segments but a different tone from the characters; for example, 情 is pronounced as /qing2/, while its phonetic radical is /qing1/, which are referred as semi-regular characters. Based on which segments are different from the phonetic radicals, the irregular characters can be divided into alliterating (e.g., 演/yan3/, sharing an onset with its phonetic part 寅/yin3/), rhyming (e.g., 靖/jing4/, sharing a rime with its phonetic part 青/qing1/), and radically irregular (e.g., 猜 /cai1/ having no apparent relationship with its phonetic part 青/qing1/) (Hsiao & Shillcock, 2006). The compound characters constitute about 80% of modern Chinese; the remaining are simple characters that have only one element. For example, the phonetic elements and radical mentioned above (虫 and 青) are, in addition to the component of compound characters, themselves simple characters. Each character of the selected words was categorised into one of three types: regular (including the semi-regular), irregular, and other (which included simple characters and compounds with no phonetic element).

Another way to categorise phonetic compounds concerns the consistency of their phonetic elements. This concept is very similar to the spelling-sound consistency in English (Jared, McRae, & Seidenberg, 1990). In order to treat it as a continuous variable, Lee, Tsai, Su, Tzeng, and Hung (2005) defined the consistency value as the relative size of the characters within its orthographic neighbours that share the same phonetic element and have the same pronunciation. For example, the character 精/ jing1 / has a phonetic element 青 /qing1/ and according to *Wieger's Chinese Characters* there are 23 characters share the same phonetic element, while only 6 of them are pronounced as / jing /(tone differences were disregarded here). Therefore, 精 is an irregular character with a consistency of 0.26 (i.e., 6/23). In order to take the size of orthographic neighbour into consideration, this method was adjusted one step further by adding one to the size of the orthographic neighbour.

The tone. Chinese is a tonal language. The phonological elements of a character include not only the vowel and consonants but also the tone that is applied to the syllable. The tone information is as important as vowels in distinguishing words from each other (Surendran & Levow, 2004). Mandarin Chinese has four tones: high-level tone (tone 1), rising tone (tone 2), dipping tone (tone 3), and the falling tone (tone 4). (Some people say that there are five tones in Mandarin, but the neutral tone 0 is used only on weak syllables and is relatively uncommon.) Even though the tone is not a standard lexical characteristic in psycholinguistic research using Chinese, word production data have shown that tones affect the overall duration. Tones 2 and 3 tend to be the longest, and Tone 4 to be the shortest (Jongman, Wang, Moore, & Sereno, 2006). Since the duration of word articulation and the mean intensity can be affected by tones, the tone of the initial and second character in each group was recorded. Table 2.1 to Table 2.4 provides a summary of the lexical level and character level characteristics of the homophones and their control words.

Table 2.1 Descriptive Statistics of Word Properties of Homophones

Word-level Variables	**Homophone Groups**		
	Higher Frequency	**Medium Frequency**	**Lower Frequency**
Example	冲击	充饥	冲积
Pinyin	chong1 ji1	chong1 ji1	chongji1
Word Variables			
Mean Word Count per Million in Subtlex-CH	32.98	8.81	0.85
(SD)	(30.97)	(3.80)	(0.62)
Phonological Frequency		42.64	
(SD)		(35.39)	
Familiarity	4.52	3.57	3.30
(SD)	(0.70)	(0.93)	(1.10)
Concreteness/imageability	2.60	3.11	3.28
(SD)	(1.25)	(1.58)	(1.74)
Grammatical category			
Nouns	14	17	14
Verbs	13	8	12
Others	2	4	3
Mean Number of Meanings	1.30	1.17	1.07

Table 2.2 Descriptive Statistics of Character Properties of Homophone

Character-level variables	Homophone Groups		
	Higher Frequency	Medium Frequency	Lower Frequency
Number of Tones			
The First Character			
Tone1		8	
Tone2		10	
Tone3		1	
Tone4		10	
The Second Character			
Tone1		4	
Tone2		3	
Tone3		4	
Tone4		18	
Character Frequency (per Million)			
The First Character			
Word-form Frequency	758.87	489.76	430.06
(SD)	(804.93)	(780.15)	(722.35)
Phonological Frequency (SD)		4384.79 (8221.21)	
Homophone Density (SD)		8.07 (5.15)	
The Second Character			
Word-form Frequency	1515.54	679.32	483.07
(SD)	(4532.36)	(1156.85)	(798.14)
Phonological Frequency (SD)		6525.48 (10630.52)	
Homophone Density (SD)		12.03 (6.47)	
Regularity			
First Character			
Regular	7	12	9
Irregular	8	5	6
Others	14	12	14
Second Character			
Regular	4	8	4
Irregular	2	4	6
Others	23	17	19
Consistency			
First Character	0.23	0.3	0.27
Second Character	0.09	0.2	0.13
Number of Strokes			
First Character	8.38	9.45	9.24
(SD)	(3.14)	(2.95)	(3.03)
Second Character	6.38	7.55	7.82
(SD)	(2.73)	(3.13)	(3.19)

Table 2.3 Descriptive Statistics of Word Properties of Non-homophones

Word-level Variables	**Non-homophone Groups**		
	Higher Frequency	**Medium Frequency**	**Lower Frequency**
Example	出击	冲力	冲量
Pinyin	chu1 ji1	chong1 li4	chong1 liang4
Word Variables			
Word Count per Million in Subtlex-CH	31.22	8.70	0.81
(SD)	(3.45)	(3.77)	(0.63)
Phonological Frequency	31.22	8.70	0.81
(SD)	(3.45)	(3.77)	(0.63)
Familiarity	4.61	3.97	3.28
(SD)	(0.60)	(0.96)	(0.85)
Concreteness/imageability	2.97	3.82	3.59
(SD)	(1.60)	(1.93)	(1.88)
Grammatical Category			
Nouns	10	13	12
Verbs	13	11	11
Others	6	5	6
Mean Number of Meaning	1.67	1.27	1.23

Table 2.4 Descriptive Statistics of Character Properties of Non-homophone

Character-level variables	Non-homophone Groups		
	Higher Frequency	Medium Frequency	Lower Frequency
The First Character			
Tone1	8	5	9
Tone2	10	10	9
Tone3	3	3	2
Tone4	8	11	9
The Second Character			
Tone1	4	5	8
Tone2	11	7	4
Tone3	7	6	5
Tone4	7	11	12
Character Frequency (per Million)			
The First Character			
Word-form Frequency	493.98	704.43	462.75
(SD)	(543.21)	(2046.90)	(707.71)
Phonological Frequency	1340.55	4373.47	3828.53
(SD)	(1043.78)	(8251.42)	(8148.40)
Homophone Density	6.37	7.00	6.77
(SD)	(3.69)	(5.15)	(4.94)
The Second Character			
Word-form Frequency	1085.78	559.33	400.81
(SD)	(1858.81)	(806.21)	(500.77)
Phonological Frequency	1669.67	1527.12	997.63
(SD)	(1983.13)	(1580.18)	(1035.01)
Homophone Density	4.40	4.63	4.23
(SD)	(3.60)	(3.00)	(3.60)
Regularity			
First Character			
Regular	15	5	5
Irregular	6	6	7
Others	18	18	17
Second Character			
Regular	5	7	6
Irregular	11	5	3
Others	13	17	20
Consistency			
First Character	0.16	0.16	0.20
Second Character	0.25	0.19	0.12
Number of Strokes			
First Character	7.97	8.59	9.17
(SD)	(2.51)	(2.76)	(3.13)
Second Character	9.17	7.59	8.45
(SD)	(3.52)	(2.78)	(2.98)

In order to compare the homophones and the control words, one-way ANOVAs were performed on numerical variables (e.g., concreteness), and chi-square values were calculated for non-numerical variables, and the likelihood ratio was used if there were more than 20% cells with an expected count less than 5. The summary of these results is displayed in Table 2.5.

Table 2.5 ANOVA and Chi-square Analyses of Group Difference for Lexical variables

Variables	**Statistic**	
Tone	**$\chi 2$ (15,174)**	***P***
First Character	4.552	.995
Second Character	23.972	.066
Regularity	**$\chi 2$ (10,174)**	***p***
First Character	8.449	.585
Second Character	15.157	.126
Grammatical Category	7.880	.641
	F (5,173)	***p***
Concreteness/imageability	1.833	.109
Number of Meanings	6.067	<.001
Word-form Frequency		
First Character	0.648	.663
Second Character	1.169	.326
Phonological Frequency		
First Character	0.763	.578
Second Character	4.142	.001
Homophone Density		
First Character	1.447	.230
Second Character	20.73	<.001
Consistency		
First Character	1.223	.300
Second Character	1.436	.214
Number of Strokes	1.955	.088

In summary, the character-level variables were well matched on the character level, despite the phonological frequency of the second character of the selected words. For the word-level variables, concreteness/imageability and main grammatical category were matched across all sets. Unmatched variables (i.e., character's phonological frequency and homophone density, word's phonological frequency, number of meanings) were entered into mixed-effect models to explore their effect on naming latency, duration and intensity.

1.3 Procedure

Participants were individually tested in a sound booth. They sat in front of a PC screen at a distance of approximately 60 cm. The experiment was presented using SuperLab5, and the whole experiment was recorded using Audacity. There were practice trials before the main experiment so that the participants could familiarise themselves with the procedure, and the experimenter could adjust the voice key parameter in SuperLab5.

The homophones were divided into three sets assigned to blocks A, B and C. For the homophones in Block A, their control words were assigned to Block B, using a Latin-square procedure. The allocation of the experiment materials was shown in Table2.6. Participants were randomly presented with order ABC, BCA or CAB.

Table 2.6 The Allocation of homophones and their control words in Experiments

		Block A	Block B	Block C
Homophone				
	Higher frequency	Item 1-10	Item 11-20	Item 21-29
	Medium frequency	Item 11-20	Item 21-29	Item 1-10
	Lower frequency	Item 21-29	Item 1-10	Item 11-20
Non-homophone				
	Higher frequency	Item 21-29	Item 1-10	Item 11-20
	Medium frequency	Item 1-10	Item 11-20	Item 21-29
	Lower frequency	Item 11-20	Item 21-29	Item 1-10

In the main experiment, there were 87 homophone trials and 87 non-homophone trials. On each trial, participants were first presented with a beep signal for 360ms, and then a word was presented in the centre of the screen for the participants to name. They were instructed to read aloud the words as fast as they could in their normal speaking rate. The presented words disappeared from the screen when the voice key was considered released, and the naming latencies were measured from the onset of the words to the beginning of the naming response. The next trial began 500ms after the voice key was released. Naming latencies were recorded using the built-in voice key in SuperLab5. Any pronunciation errors and voice-key triggering errors were noted by the experimenter and excluded from the analysis. Naming durations and intensities were measured offline using Praat (Boersma, 2002).

1.4 Analyses

In order to measure the duration and intensity of the naming responses, all responses produced during the experiment were recorded as individual sound files. The sound files were then imported into Praat (Boersma, 2002). A script was used to annotate the boundaries of sounds' onsets and ends. The boundaries were then manually checked for accuracy (and to avoid any noise). Duration and average intensity measurements regarding the labelled words were then automatically extracted.

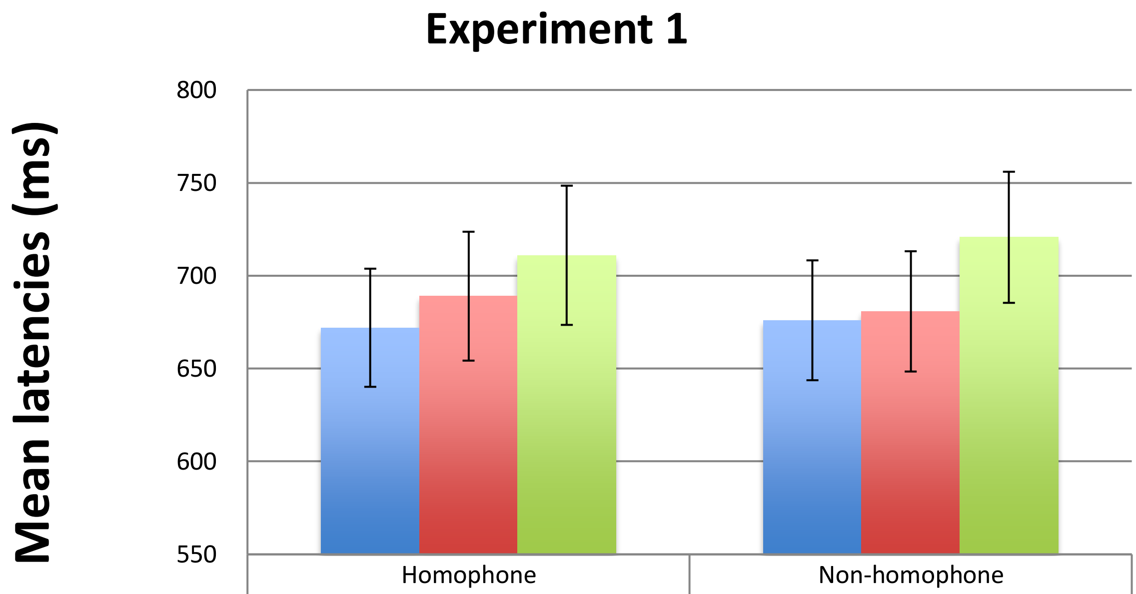
Only response latencies, durations and intensities from correctly named trials were analysed. Prior to statistical analysis, latencies over 2000ms or less than 300ms were removed, followed by those that were more than three standard deviations from each individual's means. (The durations of these words were also excluded.) The latencies of two participants were excluded from the data analysis due to a large number of voice key triggering errors, but their duration and intensity data were not excluded. The eliminated data represented about 5% of the total number of responses. The overall word reading errors, which were not separated from the voice key triggering errors, were considered to be too low for statistical analysis.

2. Results

2.1 Latencies

Figure 2.1 shows the mean word naming latencies in each condition of the experiment. Separate 3x2 related (repeated measurement) analyses of variance (ANOVAs), with Greenhouse-Geisser correction when the assumption of sphericity been violated, were performed both by participants (F_1) and by items (F_2) on the harmonic mean of naming latencies. The two factors in the analyses were word frequency (higher, medium, lower) and word type (homophone vs. non-homophone).

The main effect of frequency was significant; $F_1(1.32, 27.66) = 18.44$, $MSE = 1619.79$, $p < .001$, $\eta^2 =. 47$; $F_2(2, 56) = 17.27$, $MSE = 1101.432$, $p < .001$, $\eta^2 =. 38$. Pairwise comparisons using the Bonferroni correction revealed that lower-frequency words were named slower than both higher-frequency words by an average of 43ms ($p < .001$ in both F_1 and F_2) and medium-frequency words by an average of 27ms ($p = .003$ in F_1, and $p = .013$ in F_2). Medium-frequency words were named slower than higher-frequency words by an average of 16ms, however this difference was significant only in by participant analysis ($p = .005$ in F_1, and $p = .061$ in F_2). The main effect of word type was not significant; F_1 (1, 21) = 2.18, $MSE = 826.63$, $p = .154$; F_2 (1, 28) = 2.02, $MSE = 808.12$, $p =. 166$. Overall, homophones were named only 6ms faster than control words. Importantly, the interaction between frequency and word type was also nonsignificant: F_1 (1.45, 30.46) = 0.76, $MSE = 319.92$, $p =. 44$; F_2 (2, 56) = 0.11, $MSE = 861.27$, $p =. 89$. There was no evidence that the difference between homophones and control words (which was only small overall) was larger for the lower-frequency than for the higher-frequency words.



	Homophone	Non-homophone
Higher frequency	672	676
Medium frequency	689	681
Lower frequency	711	721

Figure 2.1 Mean Word-Reading Latencies (ms), with Standard Errors, in Each Condition of Experiment 1.

2.2 Durations

Figure 2.2 shows the mean naming durations in each condition. The harmonic means of articulation durations were analysed by 3x2 related ANOVAs by participants and by items. The main effect of frequency was significant in the analysis by participants, F_1 (1.60, 36.79) = 9.71, *MSE* = 212.74, $p < .001$, $\eta^2 =. 30$, but not in the analysis by items, *F2* (2, 56) = 1.58, *MSE* = 1151.40, $p = .22$, $\eta^2 =. 05$. Higher-frequency words had durations that were only 10ms shorter than low-frequency words. The main effect of word type (of only 7ms) was just significant in the analysis by participants, F_1 (1, 23) = 4.91, MSE = 1759.40, p= .037, η^2 =. 18, but was absent in the analysis by items, F_2 (1, 28) = 0.42, MSE = 2877.51, p = .520, η^2 =. 02. The critical interaction between word type and frequency was absent, F_1 (2, 46) = .022, MSE = 139.37, p =. 978, and F_2 (2, 56) = 0.007, MSE =986.45, p =. 993.

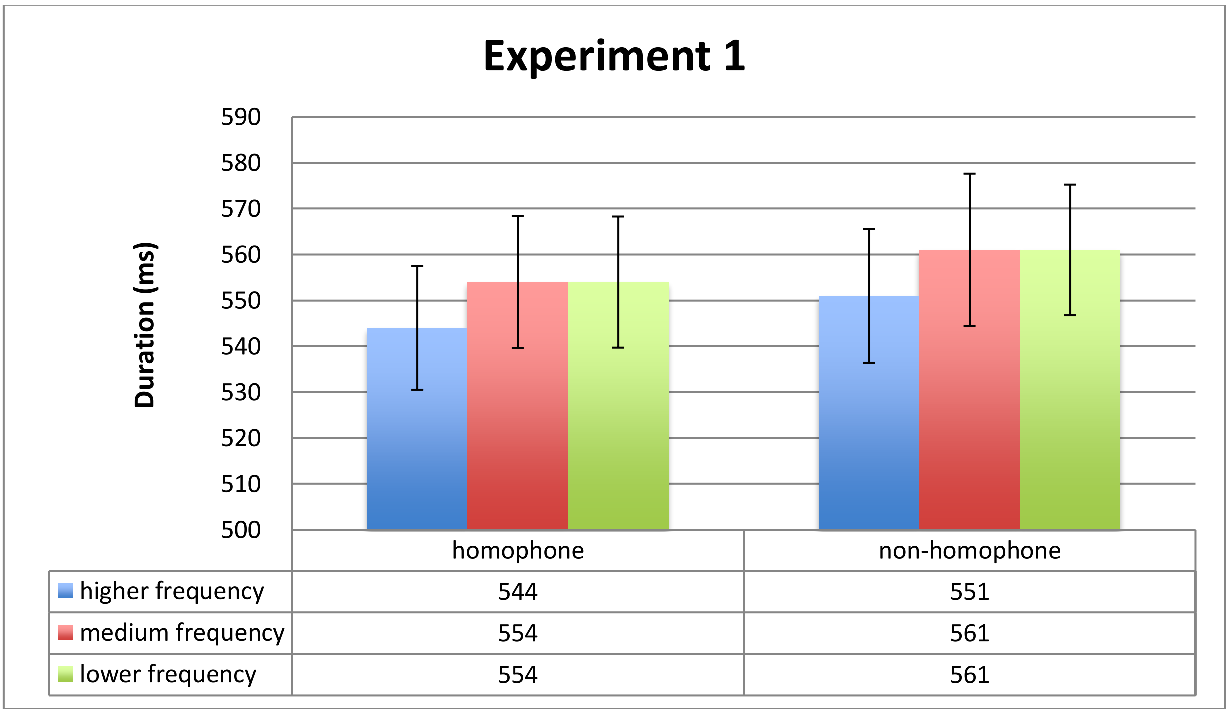


Figure 2.2 Mean Response Durations (ms), with Standard Errors, in Each Condition of Experiment 1

2.3 Intensities

Figure 2.3 shows the mean intensities of the reading responses in each condition. The mean intensities were analysed by 3x2 related ANOVAs by participants and by items. The main effect of frequency was significant in the analysis by participants, F_1 (2, 46) = 4.19, MSE = 0.322, p = .021, η^2 = .15, but not in the analysis by items, F_2 (2, 56) = 1.42, MSE = 1.46, p= .25, η2 = .05. Overall, low frequency words were named slightly louder than higher frequency words. The main effect of word type was also significant in the analysis by participants, F_1 (1, 23) = 24.06, *MSE* = 0.26, $p < .001$, η^2 =. 51, but not in the analysis by items, F_2 (1, 28) = 1.40, *MSE* = 4.24, p= .24, η^2 = .05. Non-homophones were named slightly louder than homophones. There was no interaction between frequency and word type, F_1 (2, 46) = 0.781, *MSE* =0.10, p= .464, and F_2 (2, 56) = 0.21, *MSE* =0.87, p = .81.

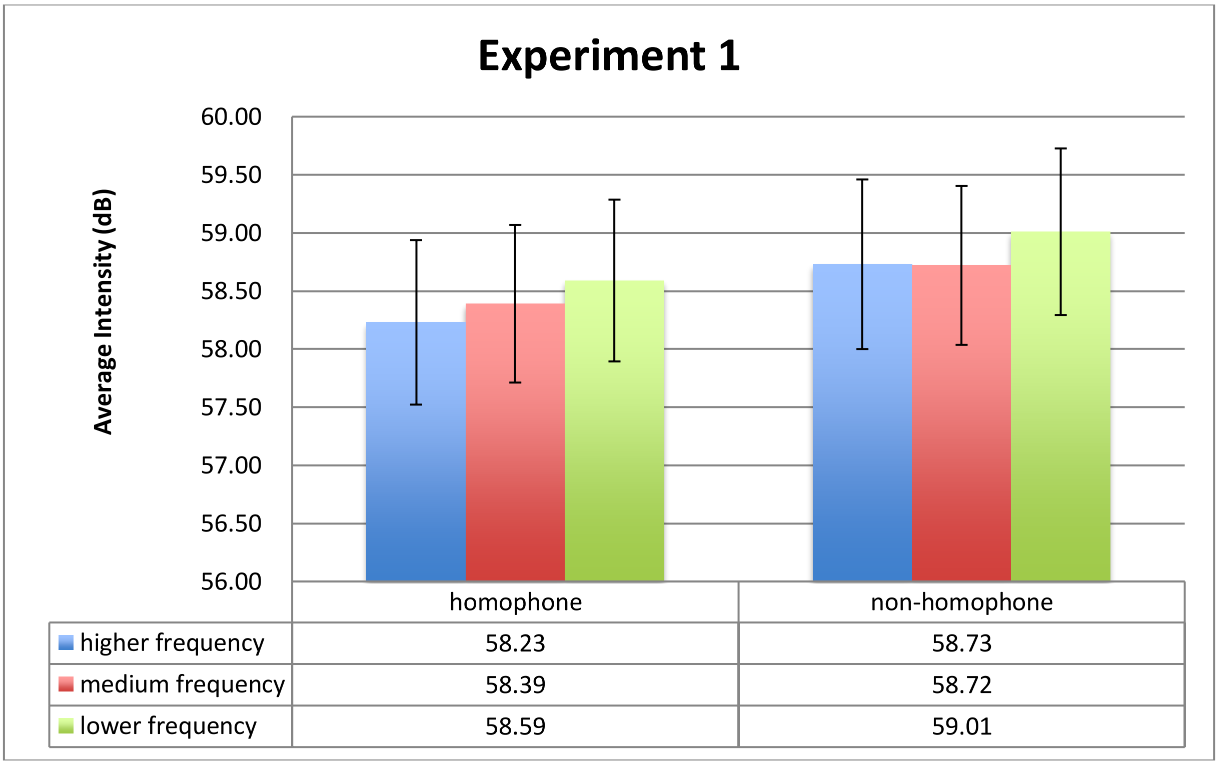


Figure 2.3 Mean Intensities (dB), with standard errors, in each condition of Experiment 1

To examine whether there was an effect from the presentation order of blocks of trials, and three-way ANOVAs was performed with the factors of block sequence (first, second, third), word type, and frequency, for latencies, durations, and intensities. There were no main effects of block order and no interactions including block order for any analysis. Thus, the patterns of results reported do not vary as a function of block order.

3. Discussion

Experiment 1 examined the oral reading responses to two-character homophone and control Chinese words of three levels of frequency. There was a reliable effect of frequency on word naming latencies (and a smaller and inconsistently reliable effect on articulation durations). However, there was no reliable difference between homophones and their word-specific frequency matched controls on any of the dependent measures of spoken word production. This pattern of results does not support the shared representation hypothesis of the lexical organisation of homophones in Chinese reading. Experiment 2 will examine reading homophones and control words for homophone twins in order to see if these results generalize to other types of homophones.

Experiment 2 Reading aloud homophone twins

1. Method

1.1 Participants

The same participants from Experiment 1 also took part in this experiment.

1.2 Stimulus words

A total of 54 pairs of homophones were selected that were homophone twins, that there were only two words that shared the same pronunciation. Each homophone word was paired with a non-homophonic control word, which had same or similar syllable as the homophones, and was matched on word-specific frequency. The homophone pairs and their control non-homophones were divided into higher frequency and lower frequency words, in a two-by-two design.

The characteristics of the word sets are shown in Table 2.7 for homophones and Table 2.8 for controls. One-way ANOVAs were performed on numerical variables, and chi-square values were calculated for non-numerical variables, and Table 2.9 shows these results, indicating that the words were well matched.

Table 2.7 Descriptive Statistics of Word Properties of the materials

Word-level Variables	Homophone Groups		Non-homophone Groups	
	Higher Frequency	Lower Frequency	Higher Frequency	Lower Frequency
Example	负责	否则	附则	抚弄
(pinyin)	fu4 ze2	fou3 ze2	fu4 ze2	fu3 nong3
Word count per million in Subtlex-CH	22.70	22.71	1.03	1.02
(SD)	24.26	24.09	1.05	1.03
Phonological frequency	23.73	22.71	23.73	1.02
(SD)	24.05	24.09	24.05	1.03
Familiarity	4.58	4.69	3.55	3.58
(SD)	0.71	0.66	0.95	1.00
Concreteness/ imageability	3.90	3.89	3.86	3.77
(SD)	1.88	1.93	1.85	1.85
Grammatical category				
Nouns	25	23	31	24
Verbs	18	20	15	23
Others	8	8	5	4
Mean Number of meanings	1.49	1.29	1.14	1.29

Table 2.8 Descriptive Statistics of Properties of the Characters

Character-level variables	Homophone Groups		Non-homophone Groups	
	Higher Frequency	Lower Frequency	Higher Frequency	Lower Frequency
Tone				
First Character				
Tone1	20	17	20	20
Tone2	16	14	16	10
Tone3	3	9	3	9
Tone4	12	11	12	12
Second Character				
Tone1	10	13	10	12
Tone2	9	15	9	8
Tone3	3	8	3	8
Tone4	29	15	29	23
Character Frequency (per Million)				
The First Character				
Word-form Frequency	845.84	1096.01	736.53	677.81
(SD)	5056.07	5198.90	5056.07	7611.48
Phonological Frequency	2984.90	2985.39	2984.90	3541.44
(SD)	1060.04	1295.35	1028.69	1028.61
Homophone Density	6.51	4.94	6.51	6.31
(SD)	4.27	4.31	4.27	5.08
The Second Character	593.40	1273.63	427.45	712.85
Word-form Frequency	642.72	3645.94	482.90	1856.48
(SD)	2809.68	2013.03	2763.36	1619.59
Phonological Frequency	5329.76	4621.87	5347.51	2668.93
(SD)	8.35	4.45	8.31	4.27
Homophone Density	5.53	3.48	5.62	3.24
(SD)				
Regularity				
First Character				
Regular	9	7	8	9
Irregular	14	10	19	21
Others	28	34	24	21
Second Character				
Regular	9	14	13	14
Irregular	12	7	11	12
Others	30	30	27	25
Consistency				
First Character	0.15	0.13	0.17	0.17
Second Character	0.16	0.20	0.17	0.21
Number of Strokes				
First Character	7.02	7.43	7.11	8.11
(SD)	2.70	3.03	2.76	3.11
Second Character	7.46	8.00	8.57	8.33
(SD)	3.39	2.94	3.25	3.18

Table 2.9 ANOVA and Chi-square Analyses of Group Difference for Lexical variables

Variables	Statistic	
Tone	**χ2 (9,204)**	***P***
First Character	8.129	0.521
Second Character	13.645	0.136
Regularity	**χ2 (6, 204)**	***p***
First Character	8.5	0.204
Second Character	3.22	0.728
Grammatical Category	7.88	0.641
	F (3,203)	***p***
Concreteness/imageability	0.259	0.855
Number of Meanings	3.554	0.015
Word-form Frequency		
First Character	1.418	0.239
Second Character	1.581	0.195
Phonological Frequency		
First Character	0.763	0.578
Second Character	9.766	<.001
Homophone Density		
First Character	1.447	0.23
Second Character	12.57	<.001
Consistency		
First Character	0.259	0.855
Second Character	0.433	0.73
Number of Strokes	2.123	0.098

1.3 Procedure

For each pair of homophones, and their corresponding control words, half were assigned to block A, and their twins were assigned to block B, and half had the reverse assignment. Words in each block were randomised. Half of the participants were presented with block A first and then block B, and the other half received the reverse order. The Experimental procedure was the same as that in Experiment 1.

2. Results

The response latencies, durations and intensities of incorrectly named trials excluded, as were trials on which there were voice-key failures. Latencies over 2000ms or less than 300ms were removed, as were those more than three standard deviations from each individual's means. Two low-frequency words were not pronounced correctly by half of the participants, and so the data from these two words, along with matched homophones or control words, were excluded from the analysis.

2.1 Latencies

Figure 2.4 shows, in the left side panel, the latency results from the homophone twins from Experiment 2. It also shows, for comparison, on the right-side panel, the results from higher- and lower-frequency homophone families (which are very similar). Separate 2x2 repeated measurement ANOVAs, with the factors of frequency (higher vs. lower) and word type (homophones vs. controls), were carried out by participants and by items on the harmonic mean of non-excluded naming latencies.

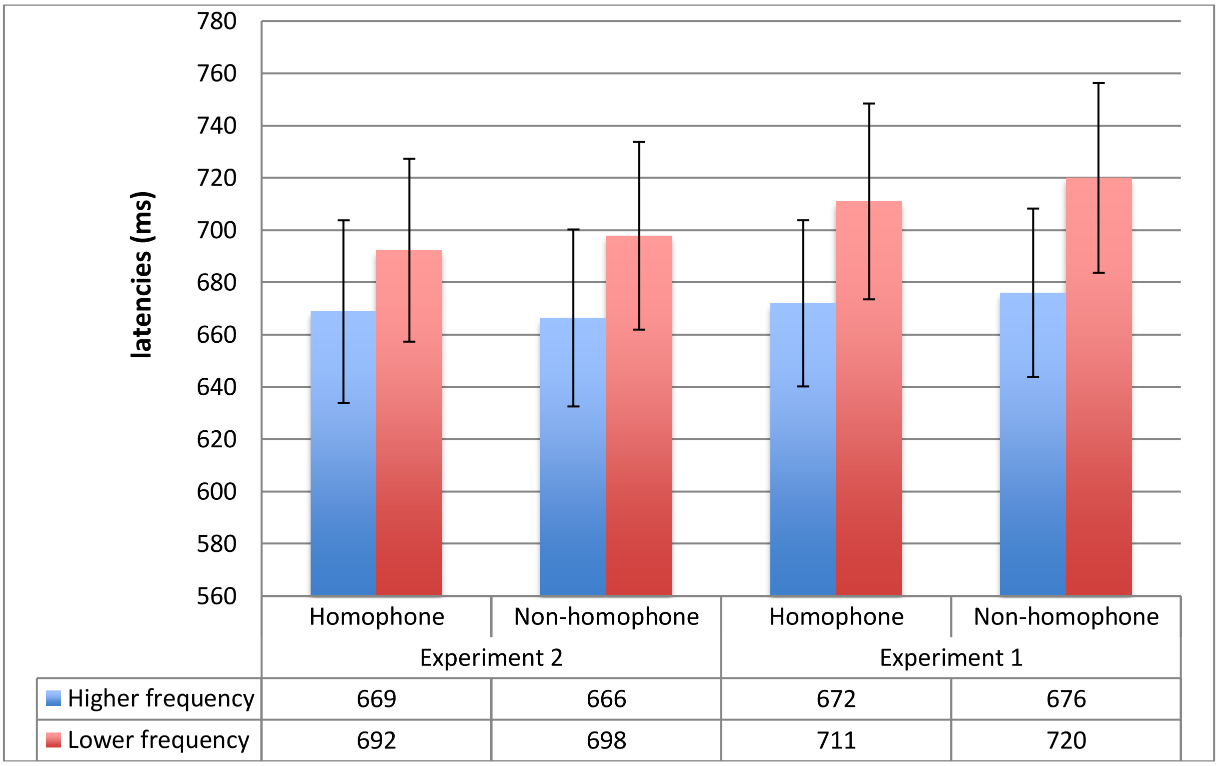
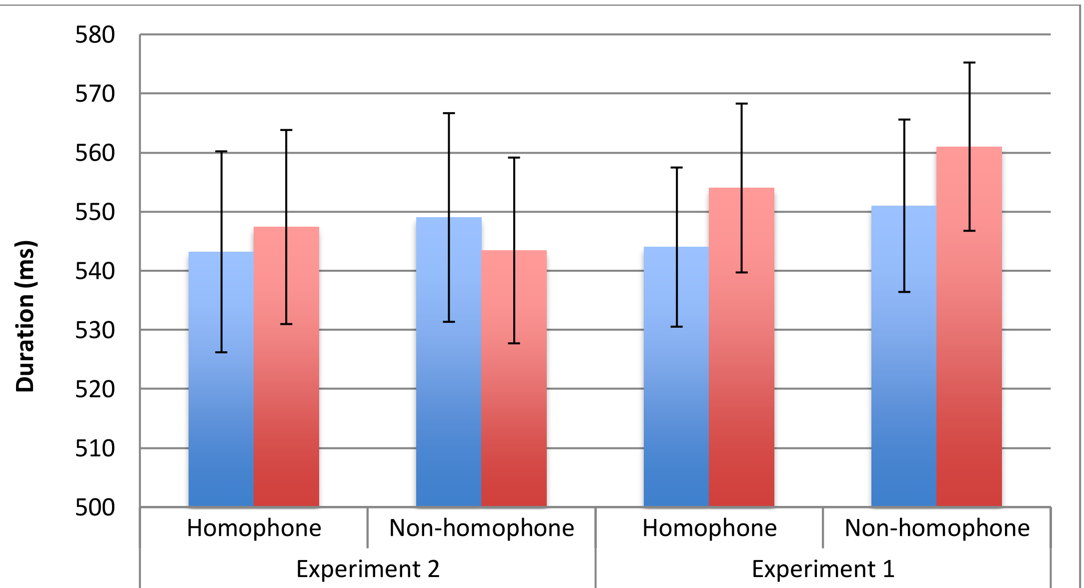


Figure 2.4 Mean Reading Latencies, with Standard Deviations, in the Four Conditions for Both Homophone Pairs (from Experiment 2) and Homophone Families (from Experiment 1)

The main effect of frequency was significant; $F_1(1, 22) = 52.18$, $MSE = 332.04$, p < .001, η2 = .70; $F_2(1, 51) = 26.83$, $MSE = 1568.50$, $p < .001$, η2 = .35. Lower frequency words were named significant slower than higher frequency words by an average of 27ms. The main effect of word type was not significant, $F_1(1, 22) = 0.93$, $MSE = 58.10$, $p = .35$; $F_2(1, 22) = 4.09$, $MSE = 89.62$, $p = .06$, and neither was the critical interaction between frequency and word type, $F_1(1, 22) = 4.09$, $MSE = 89.62$, $p = .06$; $F_2(1, 51) = 0.83$, $MSE = 1078.32$, $p = .367$.

2.2 Durations.

Figure 2.5 shows, in the left side panel, the articulation durations of the homophone twins from Experiment 2, and the similar results from higher- and lower-frequency homophone families. Mean articulation durations were analysed by 2x2 related ANOVAs and showed no significant main effects of either frequency, $F_1(1, 22) = 0.084$, MSE = 130.36, $p = .77$, $F_2(1, 51) = 0.192$, MSE = 913.84, $p = .66$, or word type, $F_1(1, 22) = 0.158$, MSE = 123.667, $p = .69$, $F_2(1, 51) = 0.145$, MSE = 2426.08, $p = .70$. The interaction between frequency and word type achieved significance in the analysis by participants, $F_1(1, 22) =$ 8.81, MSE = 62.31, $p < 0.01$, $\eta 2 = .29$, but not in the analysis by items, $F_2(1, 51) = 0.62$, MSE = 996.69, $p = .436$. The small trend in the results was for slightly larger difference between homophones and controls for the high-frequency words (i.e., the opposite of what was predicted by the shared representation hypothesis).



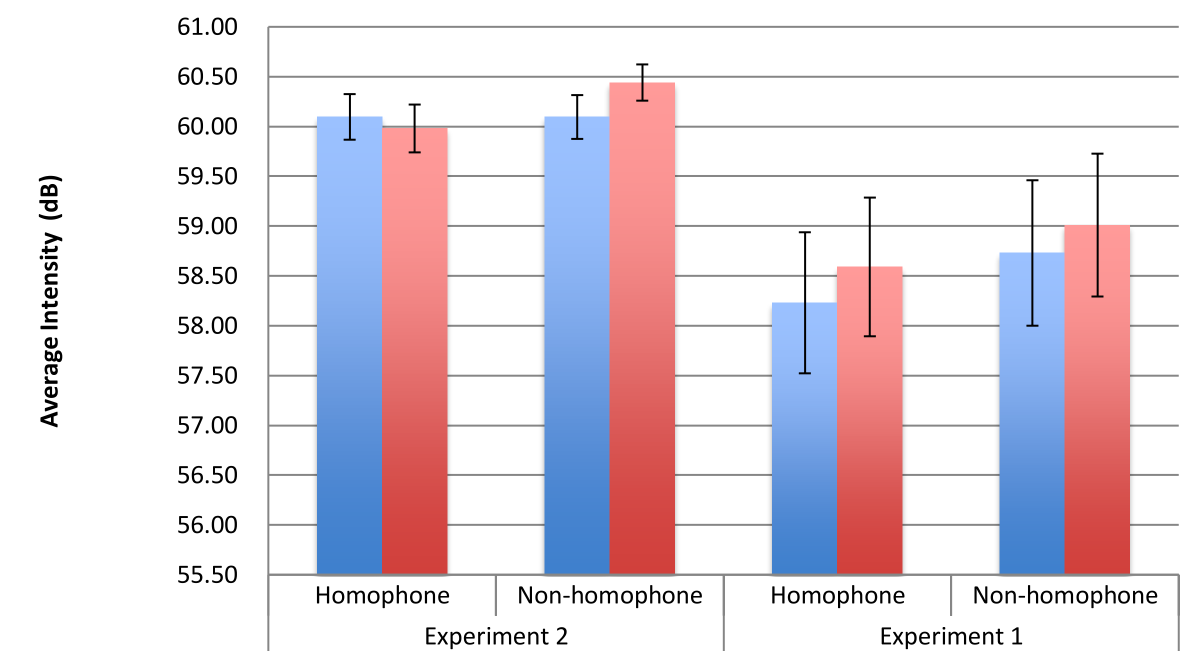
	Experiment 2 Homophone	Experiment 2 Non-homophone	Experiment 1 Homophone	Experiment 1 Non-homophone
Higher frequency	543	549	544	551
Lower frequency	547	543	554	561

Figure 2.5 Mean articulation durations, with standard deviations, in the four conditions for homophone pairs (from Experiment 2) and homophone families (from Experiment 1).

2.3 Intensities

Figure 2.6 shows the mean articulation intensities from the homophone twins from Experiment 2, along with the results from higher- and lower-frequency homophone families in Experiment 1, which had generally lower intensity. The analysis of mean intensities showed no main effect of frequency, $F_1(1, 22) = 2.95$, $MSE = 0.12$, $p = .10$; $F_2(1, 51) = 0.69$, $MSE = 0.99$, $p = .410$. The main effect of word type was significant by participants, $F_1(1, 22) = 10.77$, $MSE = 0.10$, $p < .01$, $\eta 2 = .33$, but not by items, $F_2(1, 51) = 0.82$, $MSE = 3.36$, $p = .370$. The interaction between frequency and word type was significant by participants, $F1\ (1, 22) = 12.71$, $MSE = 0.09$, $p < .01$, $\eta 2 = .37$, but not by items, $F_2(1, 51) = 2.386$, $MSE = 1.16$, $p = .129$. Non-homophonic words were read louder than the homophone words by an average of 0.23dB.

As in Experiment 1, a three-way related ANOVA was performed to assess any effect of the order of the presentation of the blocks. Neither the main effect of block sequence nor any interaction effect involving block sequence was found.



	Experiment 2 Homophone	Experiment 2 Non-homophone	Experiment 1 Homophone	Experiment 1 Non-homophone
Higher frequency	60.10	60.10	58.23	58.73
Lower frequency	59.98	60.44	58.59	59.01

Figure 2.6 Mean average intensities, with standard deviations, in the four conditions for homophone pairs (from Experiment 2) and homophone families (from Experiment 1).

3. Discussion

The results of Experiment 2 essentially replicated those of Experiment 1. For both homophone twins and homophone families, there were similar effects of word-specific frequency on reading latencies, articulation durations and mean intensities in homophone reading. The theoretically critical interaction between word type and frequency, which is predicted by the shared representation hypothesis, was observed to be either absent or unreliable in analyses both by participants and items. The results from both experiments support the independent representation hypothesis of homophones in the phonological output lexicon, and that the speed of lexical retrieval is strongly affected by the word-specific frequency of a homophone specific word-form (and appears to be unaffected by the higher frequency of either its homophone twin or larger family).

Mixed-effects Modelling Analysis

The results from both Experiments 1 and 2 provided no evidence for the hypothesis that homophones access a common or shared lexical phonological representation when reading aloud words. Rather, homophones appear to have separate or independent phonological representations, each affected by its own word-specific frequency of usage.

One potential problem for the interpretation of the results of the factorial studies in Experiments 1 and 2 might be that the phonological cumulative frequency of the words was not taken into consideration. The shared representation hypothesis predicts that the ease of word production should reflect the cumulative frequency of all the members of a homophone cohort, rather than the specific word-form frequency. Even though the experimental results show an effect of word-form frequency, they may not provide direct evidence against the shared representation hypothesis. Therefore, a mixed-effect statistical analysis using MLwiN (Charlton, Rasbash, Browne, Healy, & Cameron, 2017) was performed that included both word-specific and cumulative homophone frequency, along with a range of other variables (as described in the Stimulus materials sub-section).

The data used in the mixed-effect models was a subset of that reported that excluded non-homophones and pooled data from the family stimuli and twin stimuli. Previous analysis suggested that both word-level and character-level variables were predictors of word production times. First, a baseline model will be constructed containing all lexical variables excluding frequency (and familiarity was not included as it was highly correlated with word-form frequency). Then, it will be tested whether word-form frequency or phonological cumulative frequency was a better measurement. This was achieved by including Zipf and LogPho and their quadratic effects separately and evaluating all these candidate models by means of their likelihood values.

1.1 Latencies

The summary of the baseline model and the best fit models is shown in table 2.10. The results showed that word-form frequency was a significant predictor (-2loglikehood_baseline = 49590, -2loglikehood_baseline+zipf = 49553, $\chi 2(1) = 47$, $p < 0.001$), but cumulative phonological frequency was not (-2loglikehood_baseline+Log_Pho = 49589, $\chi 2(1) = 1$, $p = 0.26$). Increases of word-form frequency were associated with faster naming latencies. Including the quadratic effect of word-form frequency resulted in an improvement, but this did not reach significance. Model comparisons also revealed that including the random slopes at both participants level and item level for the word-form frequency variable improved the models (with random slop at the participant level vs. without, $\chi^2(2) = 26.52$, $p < .0001$; with random slope at the item level vs. without, $\chi^2(2) = 23.26$, $p < .0001$).

The null effect of the features of the first character of the bi-syllabic words might be a result of material selection. Even though the homophones employed were heterographic, some shared the first or the second character. In the 51 pairs of homophone twins, 21 of shared the first character and only 11 of shared the second one. Among the 28 sets of homophone families, 12 of the homophone cohorts (at least two of them) shared the first character and 11 shared the second character.

Table 2.10 A Summary of the Baseline Model and the Best Fit Model for Reading Latencies

	Baseline		Baseline + Zipf	
	B	***p*-value**	**B**	***p*-value**
Fixed Part				
Constant	679.233	**	677.658	**
Number of Stroke	2.734	**	2.192	**
Concreteness	1.91	.14	2.023	.12
Number of Meanings	-6.391	.15	1.376	.76
Grammatical Category=N	3.361	.64	5.61	.43
Grammatical Category=V	11.746	.09	13.433	.05*
Zipf_C1*	-10.176	.01*	-5.981	.11
Zipf_C2*	-15.772	**	-11.615	**
Log_Pho_C1	-8.219	.17	-10.463	.08
Log_Pho_C2	8.392	.17	9.164	.14
Pho_density_C1	4.141	**	4.086	**
Pho_density_C2	-.293	.56	-.289	.56
Consistence_C1	23.617	.13	18.032	.24
Consis_C2-gm	-68.422	**	-64.034	**
Regularity_C1: Regular	-23.717	.02*	-18.066	.08
Regularity_C1: Irregular	2.439	.73	2.659	.70
Regularity_C2: Regular	25.049	.02*	26.882	.01**
Regularity_C2: Irregular	11.948	.10	13.112	.07
Tone_C2_tone1	11.323	.05*	13.353	.02**
Tone_C2_tone2	4.382	.49	6.927	.27
Tone_C2_tone3	4.534	.61	7.039	.43
Tone_C1_tone1	13.473	.02*	6.864	.24
Tone_C1_tone2	0.906	.88	-2.015	.73
Tone_C1_tone3	2.557	.83	4.311	.71
Zipf_WF			-14.362	**
Random Part				
Level: SS	23570.5		23559	
Level: Word	13767.7		13655.8	

Note: C1 refers to the first character and C2 refers to the second character.
**= p < .05, **= p < .01*

1.2 Durations

A summary of the models is shown in Table 2.11. Entering the word-form frequency (Zipf) variable into the model lead to a significant improvement (-2loglikehood_baseline = 49337, -2loglikehood_baseline+zipf = 49312, $\chi2(1)$= 25, p < .001), as did entering cumulative phonological frequency (Log_Pho), although to a slightly smaller extent (-2loglikehood_baseline+Log_Pho = 49325, $\chi2(1)$= 12, p = .026). These results show that both of these predictors improve the model significantly, while Zipf was a better one. It worth noting that the coefficient of Log_Pho is positive, while that of Zipf is negative, suggesting that they are actually having the opposite effect on words' articulation duration. Entering the interaction between phonological frequency and word-form frequency resulted in significant improvement in model fit ($\chi2(3)$ = 43.72, p < .001). Based on exploration, the optimal model contains the effect of word-form frequency both in the fixed part and in the random part at the participant level and the fixed effect of phonological frequency and their interaction.

The processing disadvantage for low-frequency disyllabic homophones may be explained by competition created by feedback from the phonological level to the lexical level, as suggested by Edwards et al. (2004). In their account, phonological activation arises automatically when a printed word is presented, and activation at the phonological level then feeds-back to the lexical level. Disyllabic homophones often have one or two homophonic mates. When a low-frequency homophone is presented, the lexical representations are activated not only for the presented homophone but also for its homophonic mates due to phonological-to-lexical feedback. Thus, competition is created by its higher frequency mate lexical level, reflected longer word duration.

Table 2.11 A Summary of the Fixed Effects in the Models for Durations

	Baseline		Baseline+Zipf		+Log_Pho		+Interaction	
	B	*p*	B	*p*	B	*p*	B	*p*
Fixed Part								
Constant	549.31	**	548.03	**	544.72	**	544.99	**
Concreteness	1.32	.07	1.49	.04*	1.62	.02*	1.46	.04*
Grammatical Category=N	-9.13	.01*	-8.49	.02*	-7.60	.03*	-7.52	.03*
Grammatical Category=V	-13.94	**	-12.87	**	-12.98	**	-13.18	**
Number of Meanings	-6.63	.01*	-2.55	.31	-3.35	.18	-4.03	.11
Zipf_C1	1.60	.43	4.13	.05*	4.82	.02*	4.64	.03*
Zipf_C2	0.18	.92	1.70	.35	1.25	.49	1.29	.48
Log_Pho_C1	2.22	.55	.52	.89	-3.19	.40	-3.02	.43
Log_Pho_C2	4.46	.25	5.00	.19	3.15	.42	3.19	.41
Pho_density_C1	0.90	.02*	0.90	.02*	1.11	**	1.11	**
Pho_density_C2	-0.08	.81*	-0.09	.78	-0.10	.74	-0.08	.80
Number of Stroke_C1	-0.70	.12	-0.80	.08	-0.65	.15	-0.72	.11
Number of Stroke_C2	0.27	.53	-0.18	.69	-0.01	.98	0.11	.81
Consistence_C1	19.76	.03*	15.88	.07	16.68	.06	16.72	.06
Consis_C2-gm	-17.53	.06	-15.77	.09	-16.29	.08	-17.40	.06
Regularity_C1: Regular	-17.50	**	-14.72	.01*	-15.70	.01*	-15.77	.01*
Regularity_C1: Irregular	6.56	.09	7.30	.06	8.54	.03*	8.39	.03*
Regularity_C2: Regular	22.17	**	23.98	**	23.26	**	24.58	**
Regularity_C2: Irregular	-0.60	.88	0.42	.92	-0.55	.89	-1.14	.77
Tone_C2_tone1	3.33	.37	0.35	.93	5.77	.15	6.11	.13
Tone_C2_tone2	24.48	**	23.44	**	28.83	**	28.58	**
Tone_C2_tone3	-17.35	.02*	-16.71	.03*	-14.80	.05	-14.60	.05*
Tone_C1_tone1	57.09	**	58.20	**	57.23	**	57.95	**
Tone_C1_tone2	30.21	**	31.19	**	29.44	**	29.76	**
Tone_C1_tone3	-28.50	**	-27.05	**	-26.13	**	-25.99	**
Zipf_WF-gm			-6.15	**	-6.56	**	-5.92	**
Log_Pho					17.07	**	17.24	**
Zipf_WF *Log_pho							-5.81	.04

Note: C1 refers to the first character and C2 refers to the second character.
** $p < 0.05$, ** $p < 0.01$*

In contrast to the model for reading latency, character frequency and the phonological frequency did not significantly predict word durations, while tone information was a significant predictor. Moreover, the position of the tone interacted with the one effect on duration. For the first character, tone 2 has the longest duration, and tone 3 has the shorter one. However, for the second character, tone 1 showed the longest duration, and tone 3 showed the shortest.

1.3 Intensities

A summary of the models is shown in Table 2.12. The results showed that word- form frequency rather than cumulative phonological frequency was a significant predictor. While including the Zipf into the random part at the level of participants would significantly improve the modelling, the main effect of word-form frequency was no longer significant. In another word, the variance was due to subjects rather the frequency difference.

Table 2.12 A summary of the fixed effects in the models for intensities

	Baseline		**Baseline+Zipf**		**+Random Slope**	
	B	p-value	B	p-value	B	p-value
Fixed Part						
Constant	60.262	**	60.281	**	60.269	**
Concreteness	0.093	**	0.091	**	0.088	**
Grammatical Category=N	0.007	.96	-0.003	.98	0.004	.98
Grammatical Category=V	-0.243	.08	-0.258	.07	-0.247	.08
Number of Meanings	-0.347	**	-0.408	**	-0.411	**
Zipf_C1*	0.045	.55	0.007	.93	0.004	.96
Zipf_C2*	0.008	.91	-0.017	.81	-0.016	.82
Log_Pho_C1	0.179	.18	0.204	.13	0.209	.12
Log_Pho_C2	0.206	.14	0.199	.15	0.208	.14
Pho_density_C1	-0.177	**	-0.177	**	-0.177	**
Pho_density_C2	-0.083	**	-0.083	**	-0.083	**
Number of stroke _C1	0.054	**	0.055	**	0.053	**
Number of Stroke _C2	-0.007	.65	-0.001	.97	0.001	.96
Consistence_C1	0.583	.07	0.641	.05 *	0.656	.04*
Consis_C2-gm	0.268	.44	0.239	.49	0.233	.50
Regularity_C1: Regular	-0.482	.03	-0.525	.02 *	-0.533	.02 *
Regularity_C1: Irregular	-0.278	.05 *	-0.289	.05 *	-0.285	.05 *
Regularity_C2: Regular	-0.464	.03 *	-0.491	.02 *	-0.488	.02 *
Regularity_C2: Irregular	-0.478	**	-0.494	**	-0.5	**
Tone_C2_tone1	0.46	**	0.504	**	0.509	**
Tone_C2_tone2	-0.106	.44	-0.09	.51	-0.088	.52
Tone_C2_tone3	-3.178	**	-3.188	**	-3.194	**
Tone_C1_tone1	0.049	.71	0.032	.81	0.033	.80
Tone_C1_tone2	0.304	.03 *	0.288	.04 *	0.284	.05 *
Tone_C1_tone3	-2	**	-2.022	**	-2.005	**
Zipf_WF						
Log_pho			0.094	.05 *	0.095	.09
Zipf_WF *Log_pho						
Random Part						
Level: SS						
Var(cons)	18.589		18.589		18.584	
Level: Item					0.331	
Var(cons)					0.044	
Level: Word						
Var(cons)	1.684		1.676		1.695	

General discussion

These experiments examined the latencies, articulation durations, and average intensities of reading responses to Chinese two-character (and so bisyllabic) homophones and non-homophonic control words matched on word-specific frequency. For word reading latencies, there was a reliable word-form frequency effect, with higher frequency words being named faster than lower frequency ones. There is no evidence that reading homophones of lower frequency inherited any processing advantage from their higher-frequency homophones, as the difference between homophone and control words was the same for both high- and low-frequency words. Further, the mixed-effects analyses showed that lower frequency homophones did not receive any benefit from their higher frequency counterpart. This pattern of results supports the notion that Chinese homophones have independent representations in the speech production system.

For the articulation durations of reading responses, there was an effect of word-specific frequency, with higher frequency words having shorter durations than lower frequency words. Similar to the results of latencies, there appeared to be no effect of the cumulative frequency of homophones. This does not support the results reported by Gahl (2008). This discrepancy may be due either to the possibility that there are differences in how English and Chinese homophones are articulated or to differences in the tasks in which homophones are actually produced (in conversation compared to oral reading responses). Future work will be required to explore these possibilities further.

The mixed-effect analyses of durations that used only the data from homophones revealed a positive effect of cumulative phonological frequency when the word-specific frequency was also included in the modelling. It is possible that this reflects an interactive process that feeds-back activation from the phonological level to the lexical level Edwards et al. (2004).

For the intensity of reading responses, the results were slightly different for the data from homophone families and twins. Significant frequency and cumulative homophone frequency effects were shown in the homophone family data, suggesting that lower frequency words and homophones were read louder than higher frequency words and non-homophones. For the data from homophone twins, an interaction was observed between specific-word frequency and cumulative homophone frequency, showing that the higher the word frequency of a homophone, the lower was the intensity of its spoken response. For non-homophones, the result showed the opposite pattern. Further analysis for the homophone data showed that the frequency effect was no longer significant when it was also included in the random part of the model.

The most simple and straightforward explanation of these results is that homophones do not share a common phonological representation, and so support the independent representation hypothesis of the lexical organisation of homophones in the phonological output lexicon.

CHAPTER 3 HOMOPHONE AND PHONOLOGICAL PRIMING OF OBJECT NAMING IN CHINESE

The task of object naming has been used extensively in the study of spoken word production (for a review see Glaser, 1992). Object naming involves the same semantically driven processes of word activation (and access), word selection, and phonological encoding (and articulation control) that are central aspects of spoken word production in spontaneous speech. There is considerable neuropsychological (and converging experimental) evidence to show that object naming necessarily requires semantic mediation (e.g., Humphreys & Bruce, 1989; Riddoch & Humphreys, 2001; Kroll & Stewart, 1994). Further, object naming is a practically straightforward task that provides experimenters with a clear idea of what target word the participant should produce, which is not always possible in spontaneous speech, although it is limited to the production of single, concrete nouns.

Naming a pictured object involves a sequentially organised (and feed-forward) set of underlying processing stages (e.g., Humphreys & Bruce, 1989; Snodgrass & McCullough, 1986). Perceptual analysis constructs a structural representation of the object that is used to access its stored structural description that permits object recognition. The recognized object activates its corresponding semantic representations (i.e., stored functional, associative and sensory knowledge about the object), which permits comprehension. Semantic information is then used to activate stored lexical phonological forms (in a phonological output lexicon); this is the process of lexicalisation. Finally, names must be phonologically encoded to enable articulation. The processes of perceptual analysis and object recognition (and accessing the semantics of recognised objects) are specific to the task of naming. However, naming also involves the activation of semantic representations, lexicalisation, and phonological encoding, which are also shared by spoken word production in normal (i.e., self-generated) speech.

Figure 3.1 shows a theoretical framework for understanding the processing involved in both object naming and Chinese word reading. This framework shows that object naming is necessarily semantically mediated; there is no "direct" connection between object recognition and lexical phonological representations. Normal speech is similarly semantically mediated. Word reading can be semantically mediated (which is likely to be the major route, as we mainly read for understanding) but may also be performed by a direct lexical route (which has been established for reading in both English and the Japanese logographic script Kanji) and is theoretically possible in Chinese. The model does not show any sub-lexical or assembled phonological recoding route (i.e., the grapheme-to-phoneme conversion route that exists for reading English, as presented by the dual-route cascaded model). This is because the basic writing units of Chinese are characters, which have highly arbitrary symbol-to-sound correspondences (X. Zhou, Shu, Bi, & Shi, 1999), and there is little support for the idea that fluent Chinese readers rely on sub-lexical processes in reading (Law, Weekes, Wong, & Chiu, 2009; Perfetti, Liu, & Tan, 2005). Indeed, the mixed-effect analyses of the homophone reading latency data presented in Chapter 2 showed no major effect of whether the first characters of the bi-syllabic homophones were "regular" or "irregular" (defined in terms of the consistency of the pronunciation of the character and its component phonetic radical).

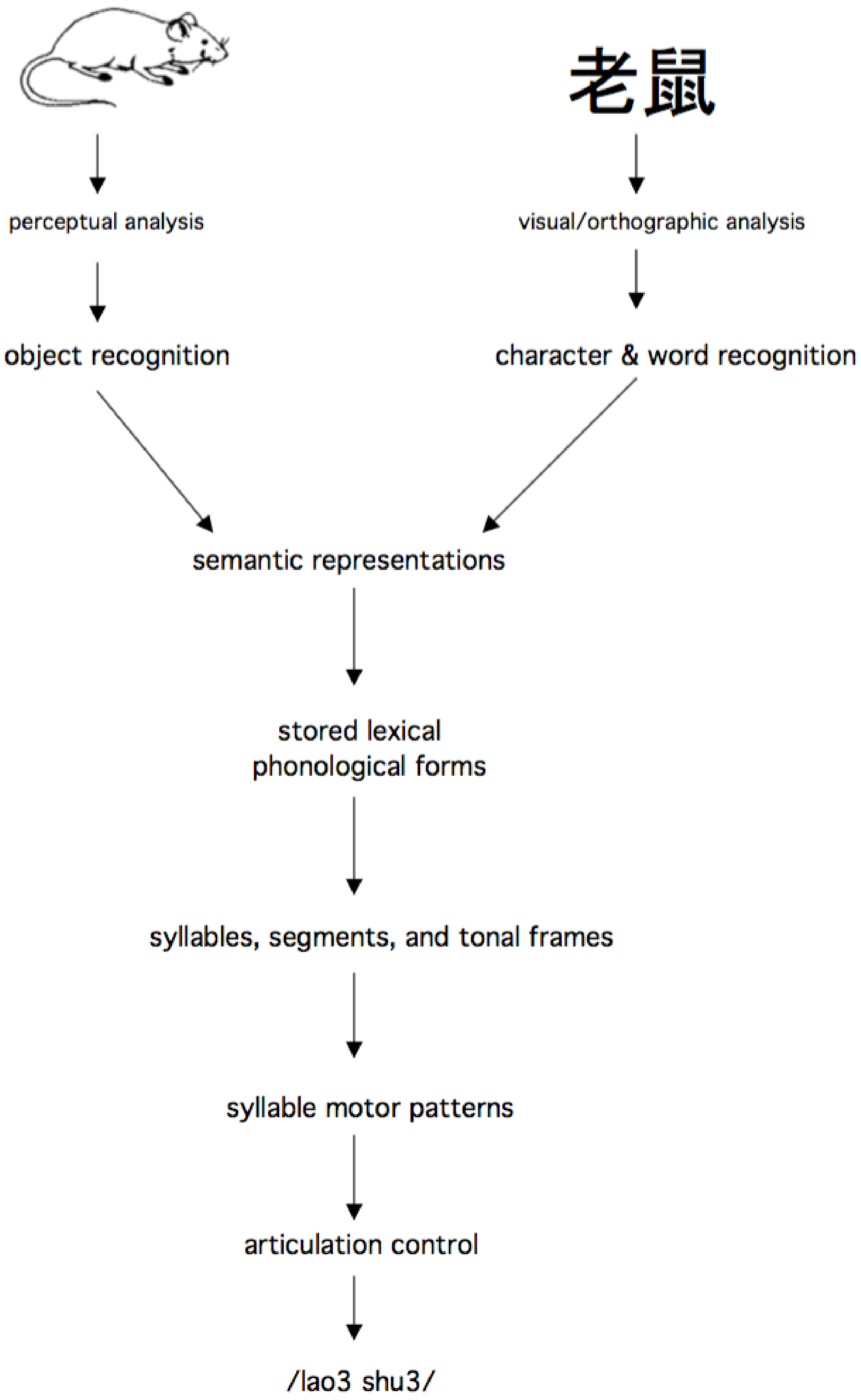


Figure 3.1 A framework of the processes involved in object naming and Chinese word reading.

The model of object naming shows the feed-forward flow of activation (i.e., recognition –> semantics –> retrieval of lexical phonology –> phonological encoding). However, the use of single lines arrows between processing stages in the diagram in Figure 3.1 should not be taken to mean that processing is necessarily discrete. Ferrand, Humphreys, and Segui (1998) proposed a cascaded processing model of object naming that involves the continuous flow of activation from one stage to the next. For example, the cascaded flow of information from semantics to lexical phonology, which is also the central aspect of lexicalization in speech, proposes that a number of semantically related words will be activated in the phonological output lexicon, which presents the problem of lexical selection in both naming and speech production. (Further, all activated lexical representations will initiate their phonological recoding.) Although the model does not show feed-back connections, it is possible that these exist either between all levels, as would be proposed by a fully interactive account (e.g., Dell's, 1990, models of speech production), or between particular levels, such as from syllables and phonemes to the lexical level. Such sub-word phonological level to lexical feedback might be required to explain the phonemic cueing of object naming especially in neurological patients with word-finding difficulty (Howard & Orchard-lisle, 1984; Pease & Goodglass, 1978).

Figure 3.1 shows no distinction between lemmas and lexemes, but rather shows that semantic activation feeds directly to lexical phonological forms (as proposed by Caramazza and colleagues, e.g., Caramazza, 1997). It is worthy of note that Chinese has no inflected forms of words, and this may further argue against the necessity for a separate syntactically based lemma level, as lemmas would contain exactly the same information as Chinese lexemes.

Importantly, Figure 3.1 shows the points of contact between object naming and word reading, which is important to be able to explain various priming effects on picture naming times. The interpretation of priming effects relies on the assumption that activation from the prime stimulus *persists* at particular processing stages to facilitate the subsequent processing of a target object.

Repetition priming is the facilitation of picture naming times by the prior presentation of the same object (e.g., Bartram, 1974; Durso & Johnson, 1979; Mitchell & Brown, 1988(Francis T. Durso & Marcia K. Johnson, 1979)(Francis T. Durso & Marcia K. Johnson, 1979)(Francis T. Durso & Marcia K. Johnson, 1979)). Such picture-to-picture repetition priming can last for a long interval between prime and target objects; for example, Cave (1997) found that repetition priming persisted when the target was presented 6 to 48 weeks after the prime. Repetition priming could result from persisting activation from the first object operating at any or all of the stages involved in object naming. Word-to-picture priming effects have also been found, although these effects are generally of a smaller magnitude than picture-to-picture repetition (e.g., Durso & Johnson, 1979). For example, Barry, Johnston, Hirsh, and Williams (2001) presented prime and test stimuli in two blocks of trials, and found (in their experiment 2) that the picture-picture priming effect was 102ms, whereas the word-to-picture effect was 46ms. Semantic priming of object naming, either from objects or words, is generally of a smaller magnitude than repetition priming and lasts for a shorter interval between prime and target, which suggests that it arises from generally short-lived activation at the semantic level.

Homophone priming of object naming has been reported for speakers of European languages (e.g., Ferrand, Humphreys, & Segui, 1998). However, Wheeldon and Monsell

(1992) found that although object naming was primed by the prior production of the same name in response to a definition (e.g., *an __ a day keeps the doctor away* –> "apple", followed by naming a picture of APPLE), there was no priming by the prior production of a heterographic homophone of the object name (e.g., *white power used to make bread* –> "flour", followed by naming a picture of FLOWER). Wheeldon and Monsell (1992) also found that Welsh-English bilinguals showed no priming of object naming in Welsh by the prior production of the English name of the object (in response to a definition in English). These results suggest that the repetition priming of object naming does not operate solely at the semantic level (as cross-language translations are assumed to have the same semantic representation) and, further, does not operate solely at the lexical phonological level (as there was no homophone priming effect). The time interval between the production of prime words and the subsequent object naming responses was rather long in these studies (of about 10 minutes), which leaves the question open as to whether homophone priming would be found at shorter intervals.

The two experiments reported in this chapter will examine word-to-picture immediate priming, that is, the target object will be presented very shortly after the prime stimulus (see Figure 3.2). The experiments will compare three main prime-to-target relationships: (1) Repetition or identity priming, where the prime word is the name of the target object (an English example would be reading *oar* followed by naming the picture OAR). (2) Homophone priming, where the prime word is a heterographic homophone of the name of the target object (an English example would be reading *or* or *awe* followed by naming the picture OAR). The homophones will also vary of their relative word frequency of the word compared to target names, and will be a higher-frequency homophone, a homophone of the same frequency, and a homophone of lower-frequency. (3) Phonological priming, where the prime word shares the same atonal syllables as the name of the target object. (There is no English example of this form of phonological priming. Phonological priming in English could use either initial-position overlap, e.g., *two*–TOOTH or rhyming words with the final-position overlap, e.g., *uncouth*–TOOTH.) All three priming conditions will be compared against unrelated control words.

Word-to-object priming effects will be determined both by the number of common processing stages underlying word reading and object naming, and by the persisting strength of activation at these stages. Repetition priming would reflect persisting activation at all levels common to word reading and object naming (i.e., semantics, lexical phonological forms, sub-lexical component phonology [syllables, segments and tones], and motor articulation), and is therefore predicted to show the largest effect size. Homophone priming would reflect persisting activation at a subset of the levels common to word reading and object naming (i.e., lexical phonological forms, sub-lexical component phonology, and motor articulation), and is therefore predicted to show a smaller effect than repetition priming. Phonological priming would reflect persisting activation at a smaller subset of the levels common to word reading and object naming (i.e., only sub-lexical component phonology and motor articulation), and so is predicted to show a smaller effect than homophone priming. As the phonologically related prime words in the experiments will share the same syllables as the target name but have different tones, the study of priming from these words will be relevant to the claims made by Roelofs (2015) and Zhang, Zhu, and Damian (2018), and illustrated in Figure 1.2, that *atonal* syllable nodes are represented in the phonological encoding of Chinese spoken word production.

The investigation of word frequency effects in homophone priming may also shed light on the question of whether homophones have shared or independent representations of their lexical phonological forms. The shared representation hypothesis claims that all homophones have the same, common phonological representation and so the reading aloud of both high- and low-frequency homophone primes would be expected to prime object naming equally. (To use an English example, the reading aloud *or* and *ore* should be expected to prime naming the picture OAR equally.) In contrast, the independent representation hypothesis would expect that the magnitude of the homophone priming effect should be larger for high- than for low-frequency words. However, these expectations are based on the critical assumption that the strength of the *persisting* activation of phonological words forms (and their sub-lexical component phonological features) will be related to the frequency of the word prime.

Experiment 3 will examine priming from the oral reading of prime words, and Experiment 4 will examine priming from silent word reading. In oral reading, there must be explicit activation of the stages of sub-lexical component phonology encoding and motor articulation. These are not explicitly required in the task of silent reading, but it is possible that they may be activated automatically. A comparison of the magnitudes of the priming effects in the two experiments may, therefore, provide an evaluation of the role of post-lexical articulatory processes in the priming of object naming.

There are two novel contributions of these experiments. First, the homophone priming naming in Chinese has not been studied previously, despite the fact that the Chinese language and orthography has very many heterographic homophones. The findings will, therefore, be important for the general issue of whether there are universal features of speech production or whether there may be language-specific constraints on the cognitive processes that underlie spoken word production. Second, the phonologically related prime words used in the experiments will share exactly the same syllables as the target name but have different tones. This contrasts with the phonologically related words used in experiments with speakers of European and non-tonal languages, which, typically, would be words that share a subset of common phonemes. These findings will, therefore, be directly relevant to the theory proposed by Roelofs (2015) that *atonal* syllable nodes are represented in the phonological encoding of Chinese spoken word production.

Experiment 3 Priming Object Naming by Reading Word Aloud

1. Method

1.1 Participants.

Forty-three students, aged between 18 and 25 years, were recruited from the University of Essex in the UK. All were native Mandarin speakers who had lived in Mainland China for more than 15 years. Sixteen took part in the pre-experiment of naming agreement, and 28 took part in the priming experiment.

1.2 Stimulus materials.

Forty-one black-and-white line drawings of common objects were selected from Snodgrass and Vanderwart (1980). In the pre-experiment conducted to rate their name agreement, sixteen participants were asked to describe these pictures using one character as quickly as they can. Fourteen pictures were excluded from the experiment either because they had a name agreement of less than 80% or because it took a very long to name the picture. The picture names were all monosyllabic, and their written names had a mean frequency per million of 38.83 (from the frequency counts in the SUBTLEX-CH corpus, Cai and Brysbaert (2010).

For each object name, 10 Chinese one-character words were chosen that varied in their phonological relationship to the object name and their word frequencies. First, there was the object name itself. Second, there were three homophones of the object name, all of which shared exactly the same phonemes and tone. One homophone had a higher frequency than the target name, another homophone was of the roughly the same frequency as the target name, and a third homophone had a lower frequency than the target name. Third, there were three phonologically related words, which shared the same phonemes but *not* the tone of the target names. Again, one related word had a higher frequency than the target name, one was roughly the same frequency, and one had a lower frequency. Fourth, there were three unrelated single-character Chinese words that had a different initial and a different final phoneme than the target object names.

Except for the repeated target names, none of the prime characters had any semantic relationship with the objects. The priming words were also matched on visual complexity (indexed by the number of strokes). A summary of the characteristics of the prime words (along with an example in each condition) is presented in Table 3.1.

Table 3.1 The Characteristics of the Prime Words Used in Experiments 3 and 4

Condition	**Example (with English translation)**			**Mean Number of Strokes**	**Mean Frequency per Million**	**Frequency Range per Million**
Target Name	鼠	shu3	Rat	9.7	38.26	7.1 - 323
High Hom	属	shu3	Category; Belong to	9.91	174.76	26.6 - 2941
Same Hom	署	shu3	Government office	9.52	34.52	6.5 - 249
Low Hom	蜀	shu3	Sichuan Province	11.15	0.6	0.1 - 1.5
High Phon	输	shu1	Lose	8.82	173.37	49.2 - 1389
Same Phon	述	shu4	State; narrate	8.94	45.41	9.2 - 417
Low Phon	孰	shu2	Who; which	10.64	0.56	0.1 - 1.8
High Unrel	技	ji4	Skill	7.93	174.01	38 - 1954
Same Unrel	踢	ti1	Kick	10	38.02	8 - 318
Low Unrel	佯	yang2	Pretend	11.33	0.56	0.1 - 1.4

(1) Target Name: The names of the target objects.
(2) Homophonic characters with a higher, same or lower word frequency: High Hom, Same Hom, or Low Hom in Table 3.1.
(3) Phonologically similar characters, which had the same phonemes but with a different tone, with a higher, same or lower word frequency: High Phon, Same Phon, and Low Phon in Table 3.1.
(4) Phonologically unrelated words, with a higher, same or lower word frequency: High Unrel, Same Unrel and LowUnrel, in Table 3.1.

1.3 Procedure

The experimental stimuli were presented using the software Superlab (run on an Apple Macintosh computer). Superlab also recorded all naming responses via the computer's built-in microphone. Participants were tested individually in a partially soundproofed booth. Before the experiment, participants were familiarised with picture materials, and were asked to provide the most appropriate monosyllable names; they were corrected if they did not produce the target name. The participants were then familiarised with the experimental procedure. Figure 3.2 shows the trial structure of the two experiments. Each participant first heard a warning tone for 360ms, followed by the prime character which was presented for a fixed time of 1000ms. After the character disappeared, there was an inter-stimulus interval of 500ms. Then the target picture was presented. The picture remained on the screen until the participant's response was detected by the computer's built-in microphone that was used as the voice-key of the Superlab software. After the participant's response was detected, there was an inter-trial interval of 700ms.

Naming times were measured from the onset of the character or picture to the onset of articulation. The whole experiment was recorded for later inspection. The experimenter stayed outside the booth but could hear the participants and control the experiment. Any hesitation or error was noted and these trials were excluded from the analysis. There was a total of 270 trials (10 conditions x 27 trials in each), and a different random order of presentation was used for each participant.

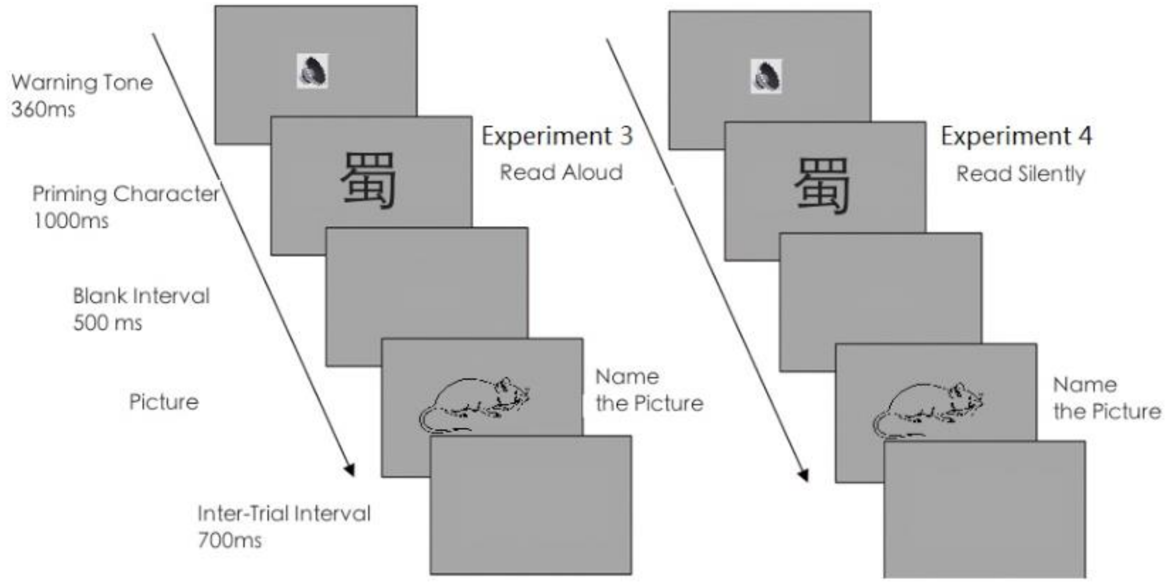


Figure 3.2 The Procedure of Experiment 3 And Experiment 4

2. Results

Prior to statistical analysis, the words and object naming times were collated and errors were identified. All responses were excluded that involved a voice-key failure, or any stuttering, utterance repairs, or production of nonverbal sounds that falsely triggered the voice key. Responses were also excluded if they involved pronouncing the character incorrectly or producing an unexpected name and/or pronouncing the name incorrectly. Also, naming times that were over 2000ms or less than 300ms were removed, followed by the responses of more than three standard deviations from the individual means. One participant's data were excluded because of the high error rate (18.6%). For the picture naming latencies, 7.76% of responses were removed, and 1.16% of the eliminated data were outliers. For the word reading times, errors accounted for 4.10% of responses and outliers 4.54%.

2.1 Object naming times.

The mean naming latency of pictures following identical prime words was 620ms. Compared to the mean of the three unrelated conditions, this produced a repetition priming effect of 106ms. The mean naming latencies in the other priming different are presented in Table 3.2.

Table 3.2 Mean Object Naming Latencies (with Standard Deviation) in Experiment 3, Along with the Homophone and Phonological Priming Effects Calculated from These Means.

	Higher Frequency	**Same Frequency**	**Lower Frequency**	**Mean**
Homophonic	641 (73)	641 (76)	656 (78)	647
Phonologically Related	693 (74)	670 (63)	698 (65)	687
Phonologically Unrelated	717 (66)	725 (66)	737 (58)	726
Mean	684	679	697	
Homophone Effect	76	84	81	80
Phonological Effect	24	55	39	40

The harmonic mean RTs of naming responses were analysed by 3 (Frequency: Higher, Same, and Lower) × 3 (Prime type: Homophone, Phonetically related, and Unrelated) repeated measures ANOVAs, both by participants and by items. The main effect of prime type was significant: $F_1(2, 50) = 47.0$, $MSE = 4368.8$, $p < .001$, $\eta_p^2 = .65$; $F_2(2, 52) = 205.2$, $MSE = 892.9$, $p < .001$, $\eta_p^2 = .89$. Object naming latencies were faster in the homophone priming condition (M= 639, SE= 5.2) than in phonological priming condition (M= 676, SE= 6.2), which were also faster than in the unrelated condition (M= 725, SE= 6.8). All pairwise contrasts were significant ($p < .001$). The homophone priming effects had an effect size of 1.11, 1.20 and 1.20 in higher, same, and lower frequency conditions, respectively. The phonological priming effects had effect sizes of 0.35, 0.87 and 0.65 in higher, same, and lower frequency conditions.

The main effect of character frequency was also significant: $F_1(2, 50) = 10.79$, $MSE = 740.8$, $p < .001$, $\eta_p^2 = .30$; $F_2(2, 52) = 15.27$, $MSE = 787.6$, $p < .001$, $\eta_p^2 = .37$. There was no significant difference between same frequency priming words (M= 675, SE= 6.5) and higher frequency priming words (M= 672, SE= 6.2), but lower frequency words were slower (M= 694, SE= 5.7). The interaction between prime type and frequency was just significant in the analysis by participants, $F_1(4, 100) = 4.62$, $MSE = 445.6$, $p < .05$, $\eta_p^2 = .16$, but was absent in the analysis by items, $F_2(4, 104) = 0.85$.

An analysis of the potential influence of the frequency differences among pictures was conducted using a linear mixed-effects model on picture naming latencies using MLwiN (Charlton et al., 2017). The model included priming condition (homophone, phonetic, and unrelated), prime word frequency (log-transformed and Laplace transformed), and the frequency of the picture name (log-transformed and Laplace transformed) as fixed effects, along with all their interactions. Continuous predictors were centred before entered the model. As the condition had three levels, contrast coding was chosen such that the phonetic and homophonic conditions were compared to the unrelated condition. The model contained random intercepts for participants and pictures, and random slopes allowing every fixed effect to vary by participants and pictures (the fixed effect of picture-name frequency only had the random slope by participants, but by not objects). The results were similar to the ANOVA results. The main effect of priming frequency was observed such that higher prime word frequency leads to faster picture naming latency (B= -10.6, SE= 2.14, p < .001). A main effect of the picture name frequency was observed such that object names with higher word frequency had faster naming latencies, (B= -27.2, SE= 5.529, p < .001). Object naming times also showed a significant effect of priming condition. Homophonic and phonological primes produced faster naming times compared to the unrelated condition (B= -82.0, SE= 11.5, p < .001 and B= -48.7, SE=5.837, p < .001,). A second contrast, between homophonic and phonetic similar condition, also showed a significant difference (B = -33.2, SE= 6.85, p < .001). None of the interactions yielded a significant effect.

2.2 Prime word reading times

The mean prime word reading latencies are shown in Table 3.3. Similar ANOVAs analyses were performed on the prime word reading times. The main effect of frequency was highly significant: $F_1(2, 50) = 382.2$, MSE=702.8, p< .001, η_p^2= .94; $F_2(2, 52) = 186.5$, MSE= 1159.2, p< .001, η_p^2= .88. Reading times were faster for higher frequency characters (M=564, SD= 3.5) than for same frequency ones (M= 584, SD= 5.0), and responses for same frequency characters were faster than for lower frequency ones (M= 658, SD= 5.4). The main effect of character type was significant by participants, $F_1(2,50) = 19.2$, MSE= 205.1, p< .001, η_p^2= .43, but not by items, $F_2(2,25) = 1.07$, p= .358. Homophonic priming words tended to be read faster than words in phonological similar and unrelated condition, which did not differ. The interaction was significant by participants, $F_1(4, 100) = 2.90$, $MSE = 319.170$, $p = 0.026$, $\eta_p^2 = .10$, but not by items, $F_2(4, 23) = 0.30$.

Table 3.3 Mean Prime Word Reading Latencies (with Standard Deviation) in Experiment 3.

	Higher Frequency	Same Frequency	Lower Frequency	Mean	Frequency Effect
Homophonic	563 (48)	579 (49)	642 (53)	**595**	**78**
Phonetic Related	566 (50)	585 (52)	658 (65)	**603**	**92**
Phonetic Unrelated	569 (46)	590 (46)	665 (54)	**608**	**96**
Mean	**566**	**584**	**655**		

Note: The Frequency Effect Is the Difference Between Higher and Lower Frequency Conditions.

3. Discussion

Experiment 3 produced a very clear pattern of results. First, there was a large repetition priming effect of 106ms. Second, there was a clear homophone priming effect. The mean magnitude of this effect was 80ms, and there were no reliable differences in the magnitude of this effect from words that were either of higher or lower frequency than the object's name. Third, there was a smaller phonological priming effect, with a mean magnitude of 40ms, and again the size of this effect was not reliably affected by the relative frequency of the prime words and target objects.

These results can be explained fairly straightforwardly by appeal to the model presented in Figure 3.1 in terms of the number of processing levels that are common to both word-reading and object naming. Repetition priming reflects persisting activation at all levels common to word reading and object naming (namely (semantics, lexical phonology, sub-lexical component phonology, and articulation), and so shows the largest effect. Homophone priming reflects persisting activation at the processing levels of lexical phonology, sub-lexical component phonology, and articulation, and so shows a smaller effect than repetition priming. The numerical difference between the repetition and homophone priming effects may be seen as an index of the contribution of semantic-level processing, and is 26ms (i.e., 106 minus 80ms). Phonological priming reflects persisting activation at the processing levels of sub-lexical component phonology and articulation, and so shows a smaller effect than homophone priming. The numerical difference between the homophone and phonological priming effects may be seen as an index of the contribution of lexical phonological level processing, and is 40ms (i.e., 80 minus 40ms). The results show that the priming effects (and so the contributions of semantic-level and lexical phonology-level processing) were not affected by the frequency of the prime words.

In homophone priming, the prime words have identical pronunciation as the target name. In Roelofs' (2015) model of phonological encoding in Chinese speech production (see Figure 1.2), producing a spoken word involves the activation of the word's atonal syllable (and its corresponding sub-syllable phoneme segments) in parallel with the tonal frame of the word, and will also activate the corresponding syllable motor programs that merge phoneme segments into the tonal frame. For reading aloud homophone prime words, all of these components will be primed. For reading aloud phonetically related prime words, which share the syllable but not the tone, the atonal syllable (and its corresponding segments) will be activated, but not its association with the tonal frame. Thus, homophones will activate the same syllable motor patterns as the target object names, but phonetically related prime words will not. The difference between the magnitudes of the priming effects from homophones and phonetically related words may, therefore, be attributed solely to differences in activation of the syllable motor patterns.

One interesting finding of Experiment 3 was the absence of any interaction between prime word frequency and the magnitude of the homophone or phonological priming effect observed; both high- and low-frequency words primed object naming times equally. The most likely explanation of this result is that, whereas the phonological forms of high-frequency words are accessed and produced faster than those of low-frequency words (as confirmed by the results of the prime word reading latencies), the strength of the persisting activation of prime words to influence subsequent object naming is not affected by word frequency. It would appear that once the phonology of a word has been fully activated, then this it will affect subsequent naming, at least over the time interval between the prime words and target object as used in Experiment 3. (Future research will be required to decide whether word frequency affects the strength of persisting activation over shorter word to object intervals.) The fact that prime word frequency did not interact with the magnitude of phonological priming (which shows that strength of *persisting* activation of lexical phonological forms is *not* related to prime word frequency in the experiment) means that it is not possible to test the contrasting predictions derived from the shared and independent representation hypotheses. Therefore, the results of Experiment 3 cannot speak further on this issue.

Experiment 4 Priming Object Naming by Silent Word Reading

The interpretation of the phonological priming effect found in Experiment 3 relied on the pre-activation of syllable motor patterns from reading aloud the prime word, that is, from its explicit articulation in overt speech. The task of silent word reading does not require participants to actually articulate the visually presented words. If the act of explicit articulation is essential for the pre-activating syllable motor patterns, then the pattern of priming effects may be different in silent reading. The possibility that homophone priming will be reduced –and even that phonological priming will be eliminated– is explored in Experiment 4 using the same stimulus materials as in Experiment 3.

1. Method

1.1 Participants.

A new group of twenty-four students (aged between 18 and 25 years) was recruited from the University of Essex. All were native Mandarin speakers who had lived in Mainland China for more than 15 years.

1.2 Stimuli and procedure.

The stimuli were identical to those used in Experiment 3. The procedure was similar to Experiment 3, with the exception that participants were asked to read the one-character prime words silently instead of reading them aloud. At the end of the experiment, participants were tested on their reading the words aloud to ensure that they knew the correct pronunciation, and their reading latencies were recorded.

2. Results

2.1 Object naming times.

For picture naming, 7.27% of responses were removed using the same criteria, and 1.40% were eliminated were outliers. The mean naming latency of objects following identical prime words was 623 ms. Compared to the mean of the three unrelated conditions, this produced a repetition priming effect of 130ms. The mean object naming latencies in each condition are presented in Table 3.4. Harmonic mean latencies were analysed using 3 (Frequency: Higher, Same, Lower) × 3 (Prime type: Homophonic, Phonetic, Unrelated) related ANOVAs by participants and by items.

The main effect of character prime type was significant: $F_1(2, 44) = 65.2$, $MSE = 1761.8$, $p < .001$, $\eta_p^2 = 0.75$; $F_2(2, 52) = 125.8$, $MSE = 1126.9$, $p < .001$, $\eta_p^2 = 0.83$. Naming times were faster after homophone primes (M= 674.5, SE= 13.6) than after phonetically related primes (M= 702.9, SE= 12.8) which also were faster than after unrelated primes (M= 754.9, SE= 12.1); all pairwise contrasts were significant ($p < .001$). Homophone priming had an effect size of 1.36, 1.34 and 1.16 in the higher, same, and lower frequency conditions, respectively. Phonetic priming had an effect size of 0.85, 0.95 and 0.76 in the higher, same, and lower frequency conditions. The main effect of prime word frequency was just significant by participants, $F_1(2, 44) = 4.1$, $MSE = 315.7$, $p = .023$, $\eta_p^2 = .16$, but not by items $F_2(2,52) = 1.85$, $p = .166$; there were only small differences between the higher (M= 709, SE= 12.3), same

(M= 708, SE= 12.4), and lower frequency words. The interaction of the two variables was absent: $F_1(4,88) = 1.00$, p=0.412, $F_2(4,109) = 1.29$, p= .278.

Table 3.4 Mean Object Naming Latencies (with Standard Deviation) in Experiment 4, with Homophone and Phonological Priming Effects

	Higher Frequency	Same Frequency	Lower Frequency	Mean
Homophonic	668(69)	673(64)	682(68)	**674**
Phonologically Related	701(67)	697(65)	711(59)	**703**
Phonologically Unrelated	754(58)	756(62)	755(59)	**755**
Mean	708	709	716	
Homophone Effect	86	83	73	81
Phonological Effect	53	59	44	52

The same mixed effect model as in the results of Experiment 3 was conducted on the object naming latencies. The best-fitting model included a main effect of the name frequency and phonological relatedness. An effect of the word frequency of the object names was also observed such objects whose names were of higher word frequency were named faster (B= -24.8, SE= 5.745, p < .001). Naming also showed significant priming; both homophones (B= -78.3, SE= 7.153, p < .001) and phonetically related primes (B= -53.6, SE= 7.237, p < .001) produced faster naming times compared to unrelated primes. There was no improvement to the fit of the model when prime word frequency was added, and no significant interactions.

2.2 Post-experiment prime word reading times

For the word reading response times, assessed at the end of the main experiment, there were 3.27% and 3.27% outliers. ANOVAs showed a significant main effect of frequency: $F_1(2, 44) = 15.38$, $MSE = 7393.0$, $p < .001$; $F_2(2, 52) = 15.1$, MSE= 6912.8, $p < .001$. There was no main effect of post-prime type, $F_1(2,44) = 1.51$, $p = .232$, $F_2(2,52) = 1.10$, $p = .340$, and no interaction, $F_1(4,88) = 1.7$, $p = .157$, $F_2(4,108) = 1.99$, $p = 0.76$.

Table 3.5 Mean Post-experiment Word Reading Latencies (with Standard Deviation) in Experiment 4

	Higher Frequency	**Same Frequency**	**Lower Frequency**	**Mean**	**Frequency Effect**
Homophonic	534(67)	525(50)	593(103)	**551**	**59**
Phonologically Related	532(71)	530(60)	589(81)	**551**	**57**
Phonologically Unrelated	524(65)	579(94)	590(67)	**564**	**80**
Mean	**530**	**545**	**591**		

3. Comparison of the results of Experiments 3 and 4

The repetition priming effects were similar from reading aloud words in Experiment 3 and when the words are read silently in Experiment 4 (106 vs 130ms). In order to evaluate possible differences in priming from reading aloud compared to silently, another linear mixed effect model was constructed with same fixed effects and random effect as before, but also including experiment (reading aloud vs. silent read) and all interactions. The best-fitting model included the fixed effect of prime type, prime word frequency, picture naming frequency and the interaction between experiment and priming word frequency. Other interactions were found to be non-significant and were therefore removed from the model. Summary statistics for the model are shown in Table 3.6 and Figures 3.5 and 3.6 show the magnitudes of the priming effects observed in the two experiments.

However, it must be accepted that, as the two studies have n=28 and n=24 participants, this between-group contrast is statistically underpowered, which limits the conclusions that may be drawn from this contrast.

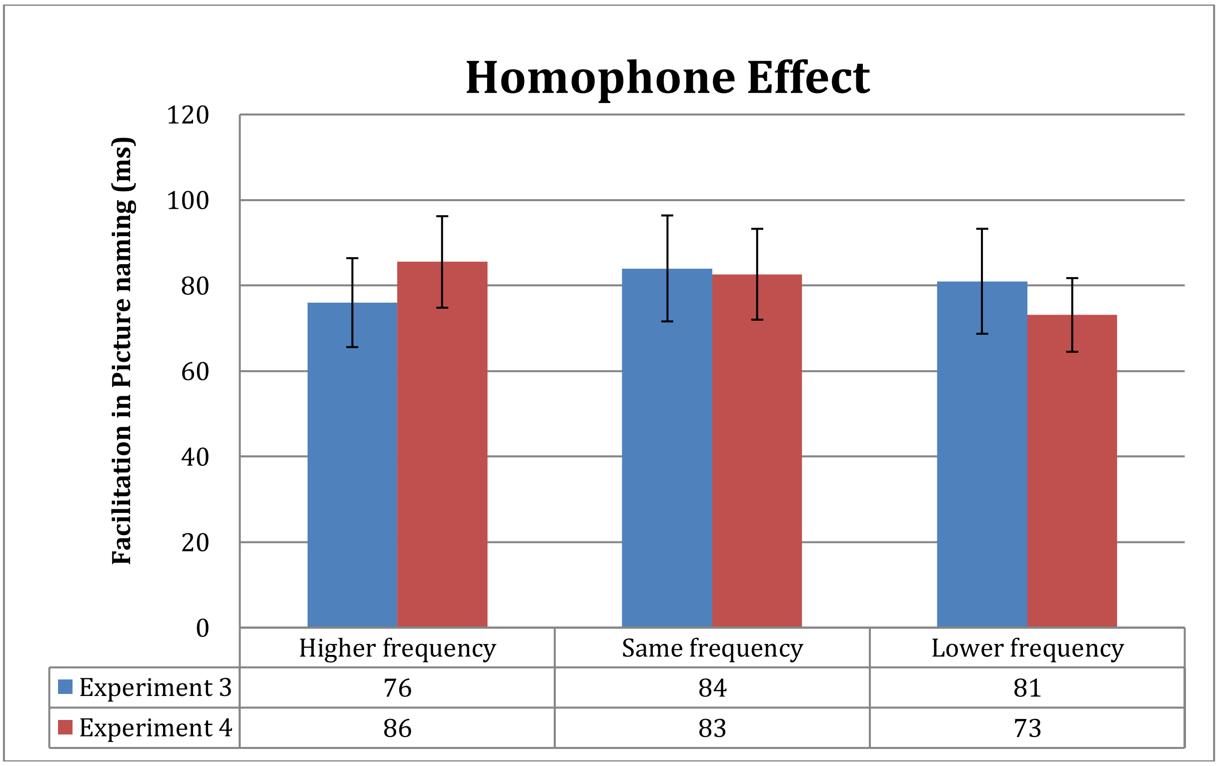


Figure 3.2 The Homophone Effects (With Standard Errors) in Two Experiments

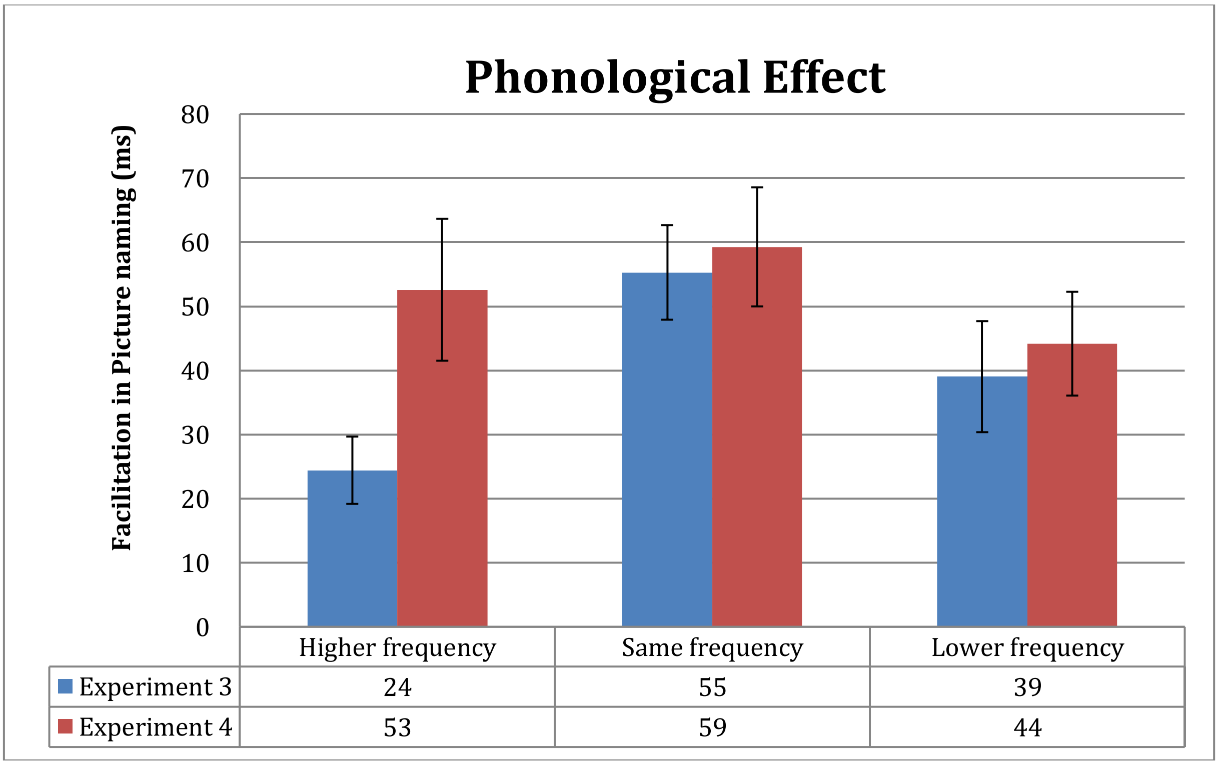


Figure 3.3 The Homophone Effects (With Standard Errors) in Two Experiments

Table 3.6 Summary Statistics of the Mixed-Effect Model of Picture Naming Latencies (N=10,819).

Fixed Part	B	S.E.	z-ratio	p-value
Intercept	758.56	9.15	82.87	<.001
Phonological Relatedness = Same syllable	-51.14	4.69	-10.91	<.001
Phonological Relatedness = Homophonic	-80.15	6.84	-11.72	<.001
Priming word frequency	-10.05	1.65	-6.11	<.001
Picture name frequency	-25.33	3.84	-6.59	<.001
Experiment condition (Silent read) × Priming word frequency	7.54	2.22	3.39	0.001

Random Part	Variance	SD
Level: Subject		
Intercept	3674.72	809.56
Phonological Relatedness = Same syllable	690.07	214.61
Phonological Relatedness = Homophonic	1854.13	454.29
Priming word frequency	7.87	13591.00
Picture name frequency	0.00	0.00
Level: Picture		
Intercept	2158.06	281.65
Phonological Relatedness = Same syllable	377.19	409.81
Phonological Relatedness = Homophonic	393.09	408.81
Priming word frequency	230.98	61.81
Level: Word		
Intercept	12648.07	230.293

General Discussion

Experiments 3 and 4 examined the priming of object naming by the prior reading of a single character (and so monosyllabic) Chinese word which was the name of the object, a homophone of the object's name, a phonetically related word to the object name (by having the same syllable but a different tone), or an unrelated word. Both experiments produced a clear and very similar pattern of results. There was a strong repetition (or identity) priming effect of 106ms in Experiment 3 and 130ms in Experiment 4. There was a smaller but still substantial effect of homophone priming of 80ms in Experiment 3 and 40ms in Experiment 4. Finally, there was a smaller, but still reliable, effect of phonological priming of 40ms in Experiment 3 and 52ms in Experiment 4.

Whether the prime word was read aloud or read silently appeared to make very little difference to either the magnitude or the pattern of priming effects observed. The homophone and phonetic priming effects observed do not require the explicit articulation of the prime words, and this suggests that in Chinese reading the phonology of the words becomes activated automatically, irrespective of reading task. Further, the results showed that the priming effects were not affected by the frequency of the prime words. The frequency effects observed on word reading latencies can be interpreted as reflecting the ease of access to, or retrieval from, permanent lexical-level representations in either the orthographic word recognition system or the phonological word production system, or both. The results of both experiments are more consistent with the notion that there are independent rather than shared phonological lexical representations of homophones. The word-specific frequency will affect the ease of accessing and/or retrieving the lexical phonology of homophones (and indeed all words); but once retrieved, the persisting activation produced to influence priming of object naming appears to be unaffected by frequency.

The general interpretation of these results is that priming can result from persisting activation, produced by word reading, at various levels of processing underlying spoken word production. Repetition priming reflects persisting activation at semantic, lexical, sub-lexical, and articulatory levels. Homophone priming reflects persisting activation at lexical, sub-lexical, and articulatory levels (including syllable motor programs). Phonetic priming reflects persisting activation at sub-lexical and articulatory levels, and in terms of Roelofs' (2015) and Zhu, Zhang, and Damian's (2016) models of phonological encoding in Chinese speech, how segmental phonemes within a monosyllabic word are integrated with the word's tonal frame to produce syllable motor programs. Homophone prime words activate the same syllable motor patterns as required to produce the target object names, whereas phonetically related prime words (that have the same phonetic syllable but not the same tone) will not. In the two experiments, only single syllable words were used. To test Roelofs' theory further, it might be interesting (if practically very difficult) to attempt to dissociate possible effects from atonal syllables (and their phoneme segments) and tonal frames. This would involve selecting probably multi-syllable words with the same syllables but different tones and words with the same tonal frames but different syllables. However, the possibility of detecting "tone-only" priming is limited due to the fact that there are only four tones used in Mandarin Chinese (although Cantonese has more tones).

Chinese reading is based primarily on the connections between orthography and meaning. This has led to the widespread assumption that skilled Chinese readers tend to rely only upon orthographic and semantic information when processing visually presented characters (Sze, Yap, & Rickard Liow, 2015). However, the results from the experiments presented here, along with others, suggest that phonological information is activated internally in the processing of Chinese characters. The results of Experiment 4 further extend this notion by showing that effects of a character's phonology can occur even in silent reading.

In a task in which participants were asked to judge if two words were semantically related, Xu, Pollatsek, and Potter (1999) found that phonological interference was observed only in exact homophones and not in characters that have same consonant and vowel. In Experiments 3 and 4, it is possible that the prime word activated a range of semantically related words, and also possible that the phonology of these related words were activated. However, in order for participants to read the character aloud, the presentation time of the prime word was set at 1000ms for both the reading aloud and silent reading conditions. One might argue that during this presentation time (and the 500ms of inter-stimulus interval), the participants had ample time to both finish the processing of the character *and* to fully activate the phonology of likely targets (and in the experiments there was a 4 in 10 chance that the target object name would be phonologically identical to the phonology of the prime word). Possibilities along these lines could be tested in future research by manipulating both the exposure duration of the prime word and, by the inclusion of semantically unrelated or related filler words, the probability of phonological match between prime word and target object name.

CHAPTER 4 MEDIATED PRIMING OF OBJECT NAMING

The experiments reported in Chapters 2 and 3 focused on the representation of homophones in lexical selection and phonological encoding during the word production. The experiments reported in the present chapter were designed to investigate how activation is transmitted from lexical to phonological levels in Chinese spoken word production, and the possibility that there are functional interactions between semantic, lexical, and phonological levels of processing.

Theories of speech production differ in terms of the proposed flow of activation from lexical to phonological levels. The main contrasting theories of the temporal dynamics of information flow are: (1) serial and discrete models, where one level of processing must be completed before the next level begins; (2) cascading models, where earlier stages continuously feed-forward information (before processing is fully completed at that level) to activate the next level; and (3) interactive models, which propose bi-directional flow of spreading activation between different levels of processing.

Debates between these models have focused on whether all semantically activated words or only a selected target word give rise to phonological activation, and whether there is feedback from phonological encoding to lexical selection. Central to these debates are studies that have claimed to find *mediated* priming effects.

1. Studies in European Languages That Use Alphabets

The issue of whether phonological activation is restricted to the target lexical node in spoken word production remains controversial. The serial discrete model of Levelt et al. (1999) claimed that only the selected lemma of a target word spreads activation to the phonological level, and semantic processing must be completed before phonological processing commences. Empirical support for such serial models came from the study reported by Levelt et al. (1991) They found no phonological activation for the non-target item in a lexical decision and picture naming dual task.

The investigation of mediated priming effects can be used to test theories of word production. According to the cascading and interactive models, in the picture-word task, where the target object could be, for example, *DOG*, the mediated distractor word could be *can*, which is phonologically similar to the semantically related word *cat* that is likely to be co-activated by the target object. While, a serial and discrete model would not predict such mediated effect, as the phonological representation of the non-target item would not be activated. For example, Damian and Martin (1998), using a picture-word task, found an interference effect on picture naming from distractors that were semantically related to the picture (e.g., TREE+*flower*). However, no effect was found for homophones semantically related words (e.g., TREE+*flour*). (This study is particularly relevant to the "homophone-to-semantic" mediated priming tested in Experiments 5 and 6 below.) Moreover, Jescheniak, Hahne, et al. (2003) found no effects of mediated distractors in the picture-word task that also recorded ERPs, which support serial-discrete models.

However, recent evidence has provided increasing support for cascading and interactive models of speech production in European languages. Peterson and Savoy (1998) studied the phonological encoding of names of pictures that had both a primary and a secondary name, such as "couch" or "sofa". Participants were required to name pictures but were required to switch the task to read aloud a printed word when, on some trials, this was presented superimposed upon a picture. The target words were phonologically related to the primary name of the picture (e.g., *count*), phonologically related to the secondary name (e.g., *soda*), or unrelated to either, and were presented at various stimulus-onset asynchronies (SOAs). Word naming times were faster when the words were phonologically related to *either* the primary or the secondary names of the object, and this effect was observed for SOAs of 150, 200, 300 and 400ms. At an SOA of 600ms only words phonologically related to the primary names showed facilitation. These results suggest that the phonological representations of both an object's alternative (or "near-synonymous") names are activated and so prime word reading-times.

In a picture-word task, Jescheniak and Schriefers (1998) presented auditory distractor words that were phonologically related either to the target name (e.g., *BURG* (castle) + "bursche") or to near-synonym names (e.g., *BURG*+ "fenchel", which is similar to "festung", an alternative name for castle). Naming times were speeded by distractors directly phonologically related to the target but were slowed by distractors phonologically related to the alternative synonymous name. This mediated interference effect is similar to one found when bilinguals perform the picture-word task. In the "phono-translation" interference effect (Hermans, Bongaerts, De Bot, & Schreuder, 1998) auditory distractor words phonologically related to non-target translations of the target object names slowed target naming; for example, MOUNTAIN+"berm" was slower than an unrelated control condition because *berm* is phonologically related to "berg", the translation of the target "mountain".

As reviewed in Chapter 1, studies using the picture-picture task (e.g., Morsella & Miozzo, 2002; Meyer & Damian, 2007; Roelofs, 2008) have found that the names of non-target objects are phonologically encoded, which is consistent with cascaded processing, and studies using Stroop-like colour word naming (e.g., Kuipers & La Heij, 2009; Roux, Bonin, & Kandel, 2014) have produced evidence consistent with "weakly" cascaded processing.

2. Are There Differences Between Chinese and European Languages?

Speech production in European languages that use alphabets seems to involve some form of cascaded information flow, especially from the semantic to the lexical level, and also some degree of feedback of information from the phoneme level to the lexical level. However, there may be differences between European languages and Chinese in the functional architecture of phonological encoding. O'Seaghdha et al. (2010) suggested that languages differ in their "proximate unit" of phonological encoding, defined as the primary selectable unit below the word level; this may be phonemes in English but syllables in Chinese (see Zhang, Zhu, & Damian, 2018). Another clear difference between European languages and Chinese concerns their orthography. The basic writing units of Chinese are characters, which have highly arbitrary symbol-to-sound correspondences (X. Zhou et al., 1999). Because symbol-to-sound correspondences are generally highly opaque in Chinese, it is common to find homophones with quite distinct visual forms, and so phonological and orthographic relatedness can be independent in Chinese.

Recently, attention has been paid to the possibility that the speech production system might differ in important aspects between European languages (that use alphabetic orthographies) and Mandarin Chinese (that has a non-alphabetic orthography). Zhuang and Zhou (2003) investigated whether processing between semantics and phonological encoding is interactive or discrete using the picture-word task in which a Chinese character was superimposed on a picture. They found that object naming was slowed by a character semantically related to the picture name and not by a character homophonic to a semantic competitor; an English example might be the picture of a MAN with the words *girl* or *buoy*. Thus, there was no homophone-mediated effect. In a second experiment, Zhuang and Zhou required participants to name the printed words and to ignore the pictures. The context pictures facilitated reading both words semantically related to the picture and homophones of semantically related words. Moreover, the facilitation effect sizes for the two types of words were similar. This result suggests that the context picture activated the concept nodes of its semantic related words and that the activation also cascaded to the phonological level. Hence, the retrieval of the semantically mediated homophones was facilitated. Zhuang and Zhou's third experiment used a semantic categorisation task that varied the difficulty of categorisation. Participants were asked to make speeded semantic judgment to pictures, onto which semantically related characters or homophones to the picture names were superimposed. Facilitation was observed for the semantic condition, but no effect was observed for the homophone condition. The authors suggested that phonological activation did not feedback to influence the semantic activation of the picture, casting doubt on the interactive view. Despite this rather unclear pattern of results, Zhuang and Zhou proposed that Chinese spoken word production may be explained by a serial and discrete model (although they accepted that a semantic representation could activate multiple phonological representations).

Q. Zhang and Yang (2006) extended this investigation by varying both the stimulus-onset asynchrony (SOA) and the picture-word relationship along different lexical dimensions. Their results showed that multiple phonological activations were observed one experiment one but not in another with a similar design. One possible explanation for this discrepancy was the difference in the proportion of related trials in the two experiments. The magnitude of priming effects has been shown to be larger when the proportion of related trials is increased (Navarrete & Costa, 2009). In their first experiment, only semantically mediated and unrelated distracters were used, while in their other experiment, semantic mediated distracters comprised only 20% of the experimental trials. It should be noted that the proportion of semantically mediated distracters in Zhuang and Zhou (2003) first experiment was also low (at 20%).

However, there are some problems with Q. Zhang and Yang (2006) study that makes their conclusion concerning mediated priming rather fragile, including the lack of strong statistically reliable evidence supporting the difference between mediated priming and unrelated words and the absence of any observed difference between priming from semantically related words and medicated words.

Two studies using the picture-word task have reported no evidence for cascaded activation in spoken Mandarin. Xuebing Zhu, Zhang, and Damian (2016) used distractors that were (a) semantically but not phonologically related, (b) phonologically but not semantically related, (c) both semantically and phonologically related, and (d) semantically and phonologically unrelated to the picture names. Their results showed that semantic and phonological variables did not interact, but rather exerted additive effects, which, following the additive-factors logic of Sternberg (1969), indicated that semantic and phonological relatedness affected different and serially organized processing stages in Chinese word production. Xuebing. Zhu, Damian, and Zhang (2015) provided further evidence for a serial model of Chinese speech using EEG measures by showing that semantic and phonological stages emerged in sequential time windows, which conflicts with the results of comparable EEG studies conducted on speakers of Western languages; e.g., Dell'Acqua et al. (2010) found a late semantic and phonological effects arising at around 320 ms post-stimulus onset, and having partially overlapping sources. The conclusion from these results is that Chinese word production reflects the operation of serial transmission of activation rather than cascading phonological activation.

3. Overview of the present studies

The four experiments reported in this chapter investigated whether the cascaded and interactive models could be applied to Chinese word production. In all experiments, as in the experiments reported in Chapter 3, participants read prime Chinese words (either aloud or silently) and then named picture objects in Chinese. (For ease of description, the examples given below will be equivalents in English.)

Experiments 5 and 6 examined *homophone-to-semantic* and *phonological-to-semantic* mediated priming. The prime words were either a homophone of a word semantically related to the object (e.g., *thyme—CLOCK*, via mediator 'time') or a phonologically related word semantically related to the object (e.g., *lime—CLOCK*, via the 'time'). The word frequency of the mediator words was also manipulated.

In addition to the conditions designed to assess mediated priming, these two experiments also included direct (i.e., non-mediated) priming conditions in order to provide a comparison with any effects of mediated priming. The direct priming conditions assessed repetition priming (e.g., *clock—CLOCK*) and semantic priming using as primes the semantically related mediator words themselves (e.g., *time—CLOCK*). The trials for these conditions were presented in a second block after those that assessed mediated priming. Figure 4.1 presents a framework for understanding the possible activation flow in Experiments 5 and 6.

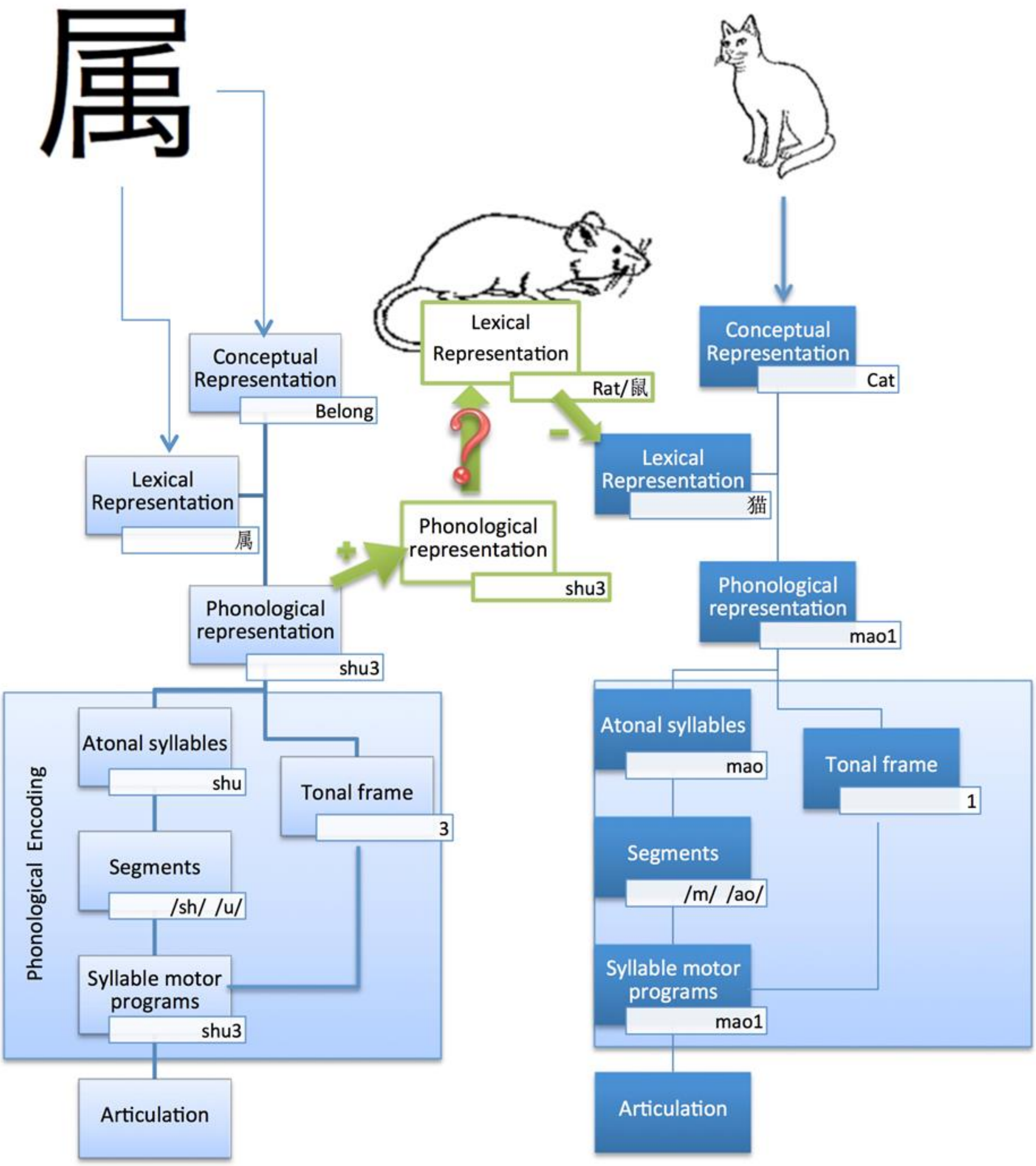


Figure 4.1 A framework of the possible activation involved in Experiments 5 and 6

When reading the words *thyme* or *lime* (aloud or silently), their phonological representations should become activated and so pre-activate the shared features with the phonological representation of the word 'time', as has been shown in Experiment 3 and 4. If there were feedback from the phonological to the lexical level, then the lexical node of 'time' should also be activated. On the assumption that, when naming a picture of a *CLOCK*, its semantic processing will activate a number of semantically related words at the lexical level (such a 'clock', 'time', 'watch', etc.), then increasing the lexical activation of 'time' (via feedback from phonology of 'thyme' and 'lime') should increase its status as a competitor in lexical selection, which should slow target naming time. It was expected that the feedback from homophones (e.g., 'thyme' to 'time') would be larger than for phonologically related words (e.g., 'lime' to 'time').

Experiments 7 and 8 examined *semantic-to-homophone* mediated priming. The related prime words were semantically related to a homophone of the target object name (e.g., *either—OAR*, via 'or'). The word frequency of the mediator words was also varied. The two experiments also included conditions designed to assess the direct (i.e., non-mediated) effects of repetition priming (e.g., *oar—OAR*) and homophone priming (e.g., *or—OAR*), using the same conditions also used in Experiments 3 and 4 (and so they also provide an opportunity to replicate the effects found in those earlier experiments). The trials for these conditions were intermixed with mediated trials within the experiment. Figure 4.2 presents a framework for understanding the possible activation flow in Experiments 7 and 8.

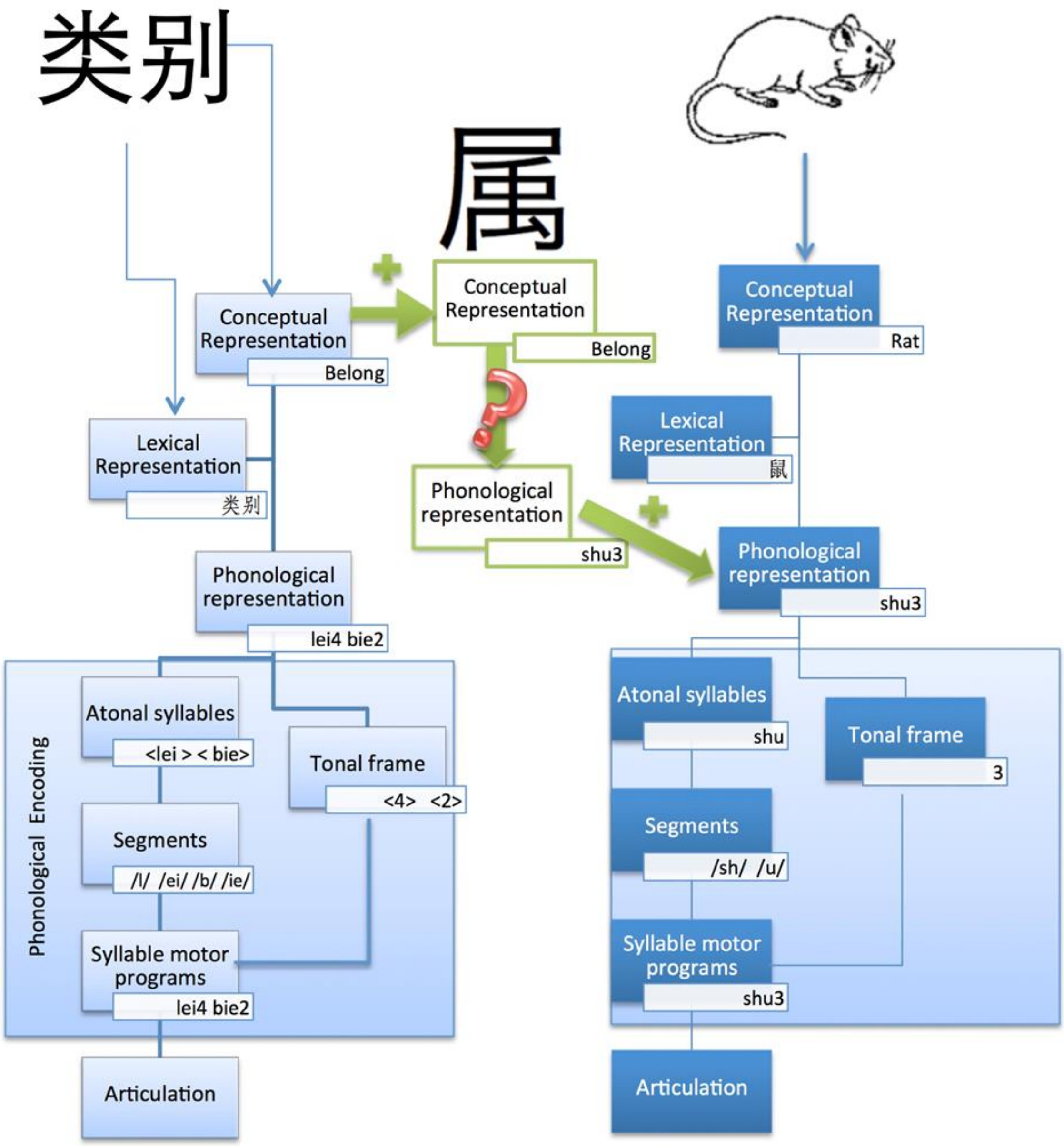


Figure 4.2 A framework of the possible activation involved in Experiments 7 and 8

When reading the word *either*, its conceptual/semantic representation should spread activation to semantically related nodes, including that of *or*. If all semantically activated words also activate their phonological representations, as predicted by cascaded models, then the phonological representation of /ɔː/ should be pre-activated and so facilitate the retrieval of picture name “oar”. As Experiments 3 and 4 found that there was an effect of word-frequency of object names on picture naming latencies when the priming task was reading aloud, but not reliably when the priming task was silent reading, it was expected that there might be a larger priming effect when the mediator was a higher frequency word.

In contrast to these predictions, a serial discrete model of Chinese word production would expect that there would be no reliable priming effect either from semantically mediated homophones and phonologically related words, or from phonologically-mediated semantically related words.

Experiment 5 Semantically Mediated Priming of Object Naming When Reading Prime Words Aloud

1. Method

1.1 Participants

Twenty-nine students, aged between 18 and 25 years, were recruited from the University of Essex. All were native Mandarin speakers who had lived in Mainland China for more than 15 years.

1.2 Stimuli and Design

The twenty-seven line-drawings of common objects, taken from Snodgrass and Vanderwart (1980), used in Experiments 3 and 4 were also used here. The previous experiments established that these pictures have high name agreements in Chinese.

A pre-experiment was conducted to select a semantically related word to each picture name. Ten participants were asked to report a monosyllable word that was semantically related to each picture and/or belong to the same category as the picture. Then an online survey was used to rate the relationships between the most commonly generated related words and the associated object name. Sixty-three participants responded in this survey. They were asked to rate the relatedness of paired words on a rank scale of 1 (*not related at all*) to 10 (*closely related*). Five pictures were excluded because the relatedness ratings were below seven. The picture names were monosyllabic or disyllable words, and their written names had a mean frequency of 73 per million, from SUBTLEX-CH corpus; Cai and Brysbaert (2010).

Each picture had four experimental prime words, which varied in word frequency and relationship to the mediator. The primes in the homophone-to-semantic (or "H2S") mediated condition shared the same syllable(s) and tone with the mediator; and one word was of high frequency and another of low frequency. The prime words in the phonological-to-semantic ("P2S") mediated condition shared the same syllable with the mediator but had a different tone; and one word was of high frequency and another of low frequency. Four unrelated prime words were also selected for each picture, which had different syllables than the mediator, and were matched to words in the H2S and P2S conditions on word frequency and their visual complexity (indexed by the number of strokes). A summary of the characteristics of the priming words (along with an example in each condition) is presented in Table 4.1.

Table 4.1 A Summary of the Characteristics of the Prime Words (Along with an Example in Each Condition)

Condition		Example (with English translation)			Mean Number of Strokes	Mean Frequency per Million	Frequency Range per Million
Target Name		猫	mao1	cat		38.26	7.1 – 323
Mediator		鼠	shu3	rat/ mouse	9.7		
H2S	Higher Frequency	属	shu3	category; belong to	9.91	174.76	26.6 - 2941
	Lower Frequency	蜀	shu3	Sichuan Province	11.15	0.6	0.1 - 1.5
P2S	Higher Frequency	输	shu1	lose	8.82	173.37	49.2 - 1389
	Lower Frequency	孰	shu2	who; which	10.64	0.56	0.1 - 1.8
Unrelated	Higher Frequency	技	ji4	skill	7.93	174.01	38 - 1954
	Lower Frequency	佯	yang2	pretend	11.33	0.56	0.1 - 1.4

1.3 Procedure

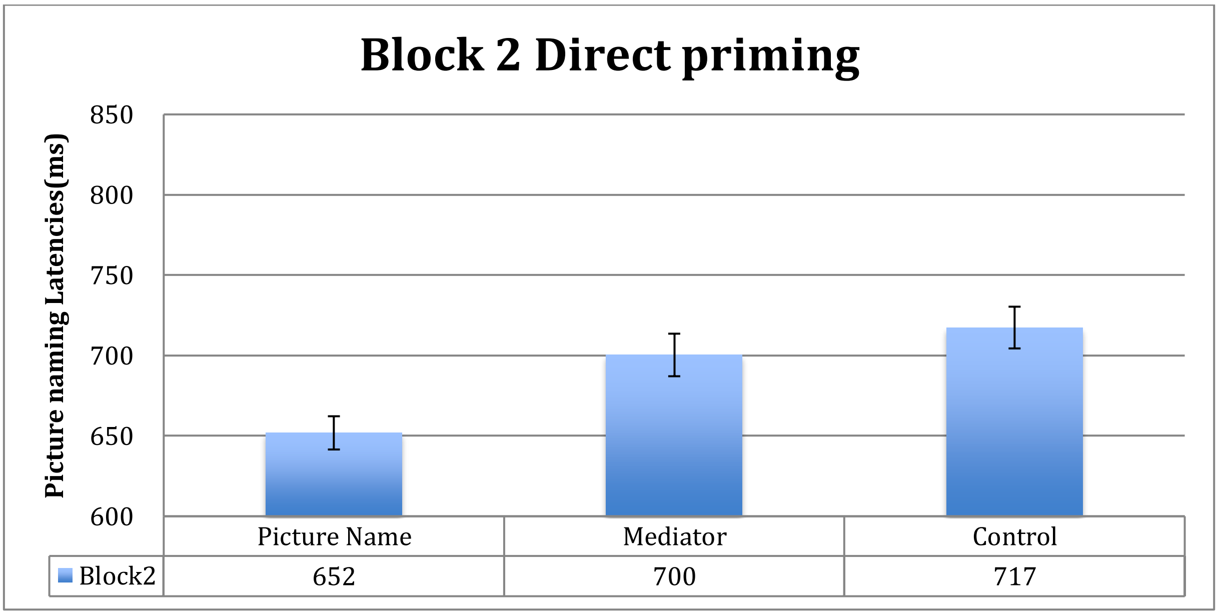
The experiment contained two blocks of trials. The first block contained the trials using prime words in the H2S, the P2S, and the unrelated sets of items. Two sets of fillers were created by randomly assigning two words from H2S or P2S groups to each picture. The second block contained the trials using picture names and the mediators themselves as the prime words, to examine repetition and direct semantic priming. A control group of 22 trials were created by randomly assigning one word from picture names or mediators to each picture. Therefore, there were 176 trials in the first block and 66 trials in the second block. Presenting the target-name and mediator-prime words in separate blocks ensured that made sure they were always presented after their H2S and P2S equivalents, so there were no reliable clues about the relationship between H2S and P2S words and their associated pictures.

The participants were asked to read aloud the characters and to name the pictures as soon as it presented, using the same temporal sequence of events and procedure as used in Experiment 3 (and illustrated in Figure 3.2). The dependent measure was object naming latency.

2. Results

The same data trimming procedure was used as in the previous experiments. The data from three participants were excluded; one due to high error rate and two dues to equipment problems. For the data from the 26 participants analysed, 7.74% of responses were removed. The mean object naming latencies in each condition are presented in Figure 4.3. The latencies from each block were analysed separately.

For the direct priming conditions (in Block 2), harmonic mean latencies in each condition were analysed by one-way related ANOVAs, with the factor of prime word type (identical, semantically related, and unrelated) both by participants and by items. The main effect of priming type was significant: $F_1(2, 50) = 28.2$, $MSE = 1081.7$, $p < .001$; $F_2(2, 42) = 29.5$, $MSE = 931.9$, $p < .001$. Pairwise comparisons (with Bonferroni corrections) showed that naming times were faster following identical prime words than both semantically related or unrelated prime words; there was a robustly significant repetition priming effect. However, despite a trend for a small semantic priming effect form the mediator words themselves presented for oral reading, the contrast between naming times following semantically related and unrelated prime was not significant (ps > 0.3).



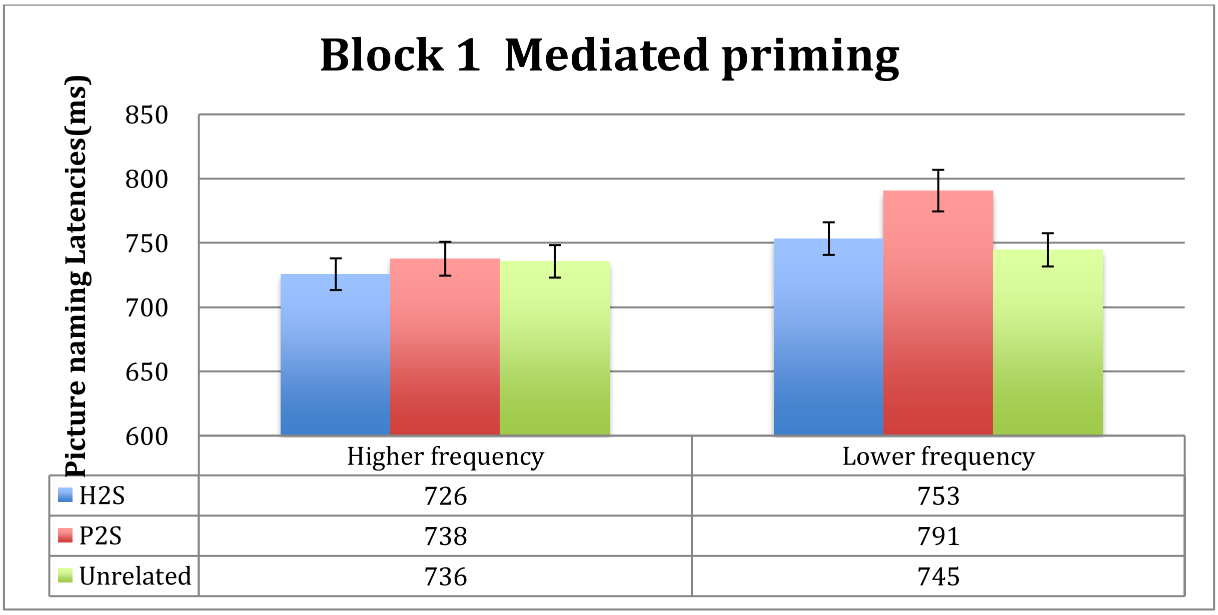


Figure 4.3 Mean Object-Naming Latencies (with Standard Errors) in Each Condition for Both Block 2 and Block 1 Of Experiment 5

For the mediated priming conditions (in Block 1), harmonic mean naming latencies in each condition were analysed by 2 x 3 ANOVAs, with the factors of prime word frequency (Higher vs. Lower) and Prime type (H2S, vs. P2S vs. Unrelated) both by participants and by items. The main effect of prime type was significant by participants, $F_1(2, 40) = 3.74$, $MSE= 759.3$, $p= 0.032$, $\eta_p2= .157$, but not by items, $F_2(2, 42) = 2.65$, $p = 0.082$. Naming latencies in the H2S condition were significantly faster than in the P2S condition, but no different than in the unrelated condition, and the difference between the P2S and unrelated conditions was not significant (*p*s > 0.28). The main effect of priming word frequency was significant; $F_1(1, 20) = 33.8$, $MSE= 646.0$, $p< .001$, $\eta_p^2= .63$; $F_2(1, 21) =20.0$, $MSE= 1190.7$, $p< .001$, $\eta_p^2= .49$. Object naming times were faster following high frequency than low frequency mediator words. There was no interaction between the two factors, $F_1 < 1$, $F_2 < 1$.

The data analysis of Block 2 showed no difference in naming latencies between the semantically related condition (i.e., the mediator words themselves presented as prime words) and the control condition (i.e., the same words but re-ordered so as to be unrelated to the target pictures). Thus, the experiment failed to find a direct semantic priming effect, which may have been due to the complex design adopted. In Block 2, 67% of the priming words were identical or semantically related to the picture names; participants may, therefore, have changed their strategies during response selection. Therefore, comparisons between the semantically related and H2S conditions, and between the semantically related and P2S condition (i.e., the priming words from Blocks 1 and 2), were conducted to investigate if mediators show similar priming effects as their homophones or phonologically related words. The naming latencies in the semantically related condition, the higher frequency H2S condition, and the lower frequency H2S conditions were analysed by a one-way ANOVA. This showed a significant main effect; $F_1(2, 40) = 35.0$, $MSE = 1545.9$, $p < .001$, $\eta_p^2 = .64$; $F_2(2, 42) = 42.58$, $MSE = 1450.7$, $p < .001$, $\eta_p^2 = .75$. The difference between the semantically related, and the higher and lower frequency H2S condition was significant (as was the difference between the higher and lower frequency H2S words). A similar analysis was conducted to compare the semantically related, and the higher and lower frequency P2S conditions, and this also showed a significant main effect: $F_1(2, 40) = 60.2$, $MSE = 1231.9$, $p < .001$, $\eta_p^2 = .75$; $F_2(2, 42) = 61.6$, $MSE = 1328.4$, $p < .001$, $\eta_p^2 = .75$. There were significant differences between the semantically related and the lower frequency P2S words (and between the higher and lower H2S words). These results showed that picture-naming was interfered less in semantic priming condition compare to that in any mediated priming conditions. These differences could not be due solely to word frequency, as the prime words in the higher frequency H2S and P2S conditions were of higher word frequency than their mediators (i.e., the semantically related words). However, the difference between the mediators and their homophones could not be compared, as they were presented in different blocks.

3. Discussion

The experiment produced a mixed set of results. First, although there was a clear and robust repetition priming effect (of 65ms), the smaller direct semantic priming effect (of 17ms), when the mediator words themselves were presented as prime words, failed to reach significance. Second, there were no reliable homophone-to-semantic or phonological-to-semantic mediated priming effects. Finally, there was an effect of prime word frequency but this did not interact with any of the mediated priming effects. This is similar to the finding of Experiment 3, where the frequency of the prime word also did not interact with the direct priming effects tested in that experiment.

Given that the experiment found no clear direct semantic priming effect, absence of any mediated priming could be attributed to an insufficiently high degree of semantic relatedness (despite the reasonably high ratings of relatedness provided by participants in the online survey). However, naming latencies primed by a directly semantically related word were faster than those primed by the homophone mediated semantically related word and faster than those primed by phonologically related words of the semantically related words. Overall, these results provide no support for the view that there is feedback of activation from the phonological to the lexical level during word production.

Experiment 6 Semantically mediated priming of object naming when reading prime words silently.

1. Method

1.1 Participants

Another twenty-eight students, aged between 18 and 25 years, were recruited from the University of Essex. All were native Mandarin speakers who had lived in Mainland China for more than 15 years.

1.2 Stimuli, Design, and Procedure

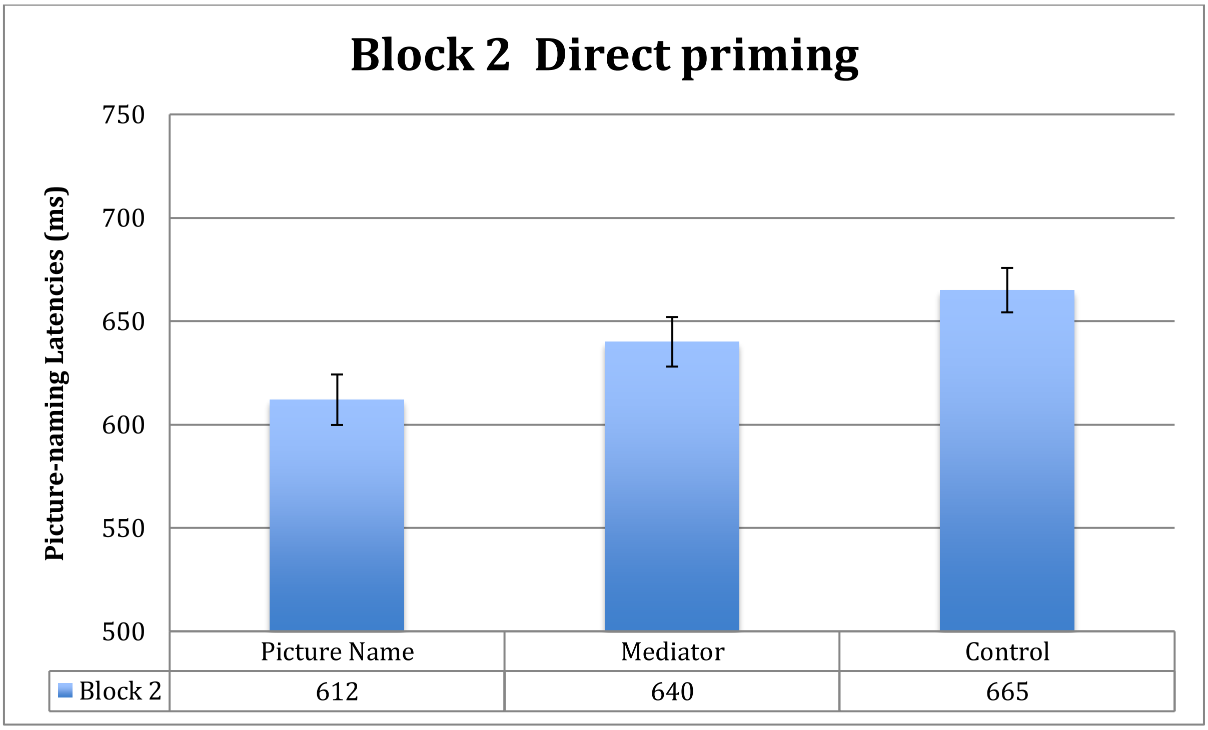
The stimuli and experimental design were the same as Experiment 5. The experiment procedure was the same as Experiment 4 in Chapter 3. The participants were asked to read the characters silently and then to name the pictures when presented. The dependent measures were the latencies for picturing naming.

2. Results

The same data trimming process was used as the previous experiments. For the picture naming latencies, 2.77% of responses were excluded. The mean object naming latencies in the different conditions are presented in Figure 4.4. The picture naming latencies in each block were analysed separately.

For the direct priming conditions (in Block 2), harmonic mean naming of each condition was analysed by one-way ANOVAs with the factor of prime type (identical, semantic, unrelated), with Greenhouse–Geisser correction for sphericity when required. The main effect of priming type was significant: $F_1(1.6, 43.6) = 16.8$, $MSE = 1687.9$, $p < .001$, $\eta_p^2 = .384$; $F_2(2, 42) = 55.5$, $MSE = 746.5$, $p < .001$, $\eta_p^2 = .725$. Pairwise comparisons showed that object naming times were significantly faster following identical prime words than both semantically related and unrelated words, and that naming times were faster following semantically related words (i.e., the mediator words themselves) than unrelated words (*p*s < 0.02); there were significant repetition and semantic priming effects.

For the mediated priming conditions (in Block 1), the harmonic mean naming times in each condition were analysed using 2x3 related ANOVAs with the factors of mediator word frequency (Higher vs. Lower) and prime type (H2S vs. P2S vs. Unrelated). Neither the two main effects nor the interaction between them was significant, $F_1 < 1$, $F_2 < 1$ for all effects. There were no reliable mediated priming effects.



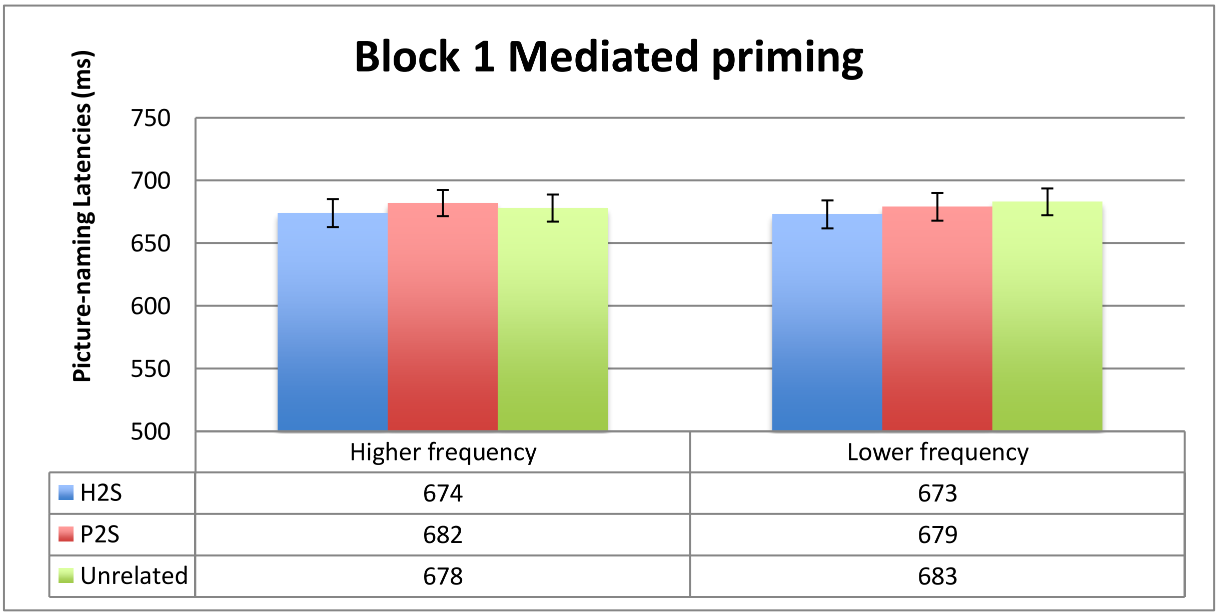


Figure 4.4 Mean object-naming latencies (with 95% confidence intervals) in each condition for both Block 2 and Block 1 of Experiment 6

3. Discussion

In Experiment 5 the prime words were read silently, and the results showed reliable effects of both repetition priming and semantic priming from the related mediator words themselves. However, there was no reliable priming effects mediated by homophones of the prime words or by words phonologically related to the primes. To use English examples, seeing the words *clock* and *time* both directly primed naming a picture of a CLOCK, but seeing homophones of these words (e.g., *thyme*) or words phonologically related to these words (e.g., *lime*) but shared atonal syllables did not prime naming CLOCK. There were no reliable mediated priming effects detected.

Although both Experiments 4 and 5 found no reliable mediated priming effects, there were some differences in the results they did find. In Experiment 5, participants read the prime words silently and here a significant direct semantic priming effect was observed; when reading aloud the prime words, the semantic priming effect did not achieve significance. It is possible that silent reading is more like normal reading and so is more tuned to accessing semantic representations. However, silent reading did not lead to the detection of either homophone-to-semantic or phonological-to-semantic mediated priming. Experiment 4 found that these mediated priming effects were equally difficult to detect. The null effects observed in Experiment 5 could not be attributed to the insensitivity of detecting direct semantic priming from the mediator words themselves in Experiment5. Overall, these results suggest that there is no feedback from the phonological level to the lexical-semantic level.

Experiment 7 Homophone Mediated Priming of Object Naming When Reading Prime Words Aloud

The following two experiments examined *semantic-to-homophone* mediated priming. The presence of this mediated priming effect is expected by interactive models of speech production where, in the act of naming a pictured object, a number of lexical-level representations become active by a process of cascade, which is then phonologically encoded. For example, the semantic representation (small mammal, pet, barks, etc.) would activate a number of related words (e.g., 'dog', 'hound', 'terrier', 'fox', 'cat', etc.), if to varying degrees. If one of these non-target names had had its phonological form pre-activated by the prime word, then it would be expected to become a stronger competitor in lexical selection that should slow target naming time.

1. Method

1.1 Participants

Twenty-seven students, aged between 18 and 25 years, were recruited from the University of Essex. All were native Mandarin speakers who had lived in Mainland China for more than 15 years.

1.2 Stimuli and Design

Twenty-eight black and white pictures of common objects were selected as candidates from Snodgrass and Vanderwart (1980). These pictures were used in the previous two experiments, which established that they have high name agreements in Chinese. Each picture was paired with two homophones, one with higher word frequency and one with lower word frequency, as the mediators. Both the mediators and the object names were monosyllabic words represented by a single character. Each mediator word was paired with two priming words: a semantically related prime word, and a semantically unrelated control prime word. The control words were matched to the experimental prime words on word frequency and visual complicity (number of strokes). The semantic relatedness between the chosen prime words and the mediators was then rated in an online survey in which sixty-three participants were asked to rate the relatedness of pairs words on a scale of 1 (not related at all) to 10 (closely related). Further, the semantic relatedness was also calculated using HowNet, which is an online (and common-sense) knowledge base showing inter-conceptual relations and inter-attribute relations of concepts as connoted in Chinese and English bilingual lexicons (Dong & Dong, 2003). An effort was made to ensure that the experimental prime words were as semantically close to the mediators in both the ratings from the online survey and HowNet ratings. Finally, 19 pictures with their related prime words were selected. The picture names were all monosyllabic words. The semantically related mediator prime words are all monosyllable. The characteristics of the priming words (along with an example in each condition) are summarized in Table 4.2.

Table 4.2 The mean characteristics of the prime words (with an example) in each condition.

Condition	Example (with English translation)		Mean Number of Strokes	Mean Frequency per Million	Frequency Range per Million
Target Name	鼠 shu3	Rat/ mouse			
Higher-Frequency Mediator	属 shu3	Category; Belong to		348.04	19.92-3015.59
S2H	类别 lei4bie2	Classification	7	59.01	0.33- 401.98
Control	糖衣 tang2yi1	Sugar coating	8.6	59.01	0.33- 401.98
Lower-Frequency Mediator	蜀 shu3	Another name for Sichuan Province		0.76	0.08 - 1.44
S2H	四川 si4chuan1	Sichuan Province	8.6	174.84	0.06-3226.48
Control	山药 shan1yao4	Chinese yam	8.6	174.84	0.06-3226.48

For each target picture, there were sixteen prime words. For the direct priming conditions, the prime words were: (1) the object name itself; (2-4) a higher frequency homophone, phonological related character and unrelated word; (5-7) a lower frequency homophone; phonological related character and unrelated word, and (8-12) unrelated filler words, which were the same prime words from but re-ordered to ensure that there was no phonological, semantic or semantically mediated phonological relatedness to the pictures. For the mediated priming conditions, the prime words were: (13) a semantic-to-homophone (S2H) high frequency mediator; (14) a S2H low frequency mediator; (15) an unrelated word matched to the S2H high frequency mediator; and (16) an unrelated word matched to the S2H mediator. Note that the high and low frequency homophone mediators, and their unrelated control words, were the same as those used in Experiments 3 and 4 reported in Chapter 3. There were 304 trials in total (19 target objects x 16 prime words), and all trials were presented randomly in the experiment.

1.3 Procedure

The experiment procedure was the same as Experiment 3 in Chapter 3 and Experiment 5 in this Chapter. On each trial, participants had to read aloud the prime word and to name the pictured object.

2. Results

The same data procedure for data trimming as in the previous experiments was applied, and 3.5% of picture naming responses were excluded. The mean naming latencies in the mediated conditions are presented in Figure 4.5.

The means of naming latencies for the direct priming conditions are shown (for both this experiment and the following Experiment 8) in Table 4.3. The statistical analyses (which are not presented here for the sake of brevity) confirmed the pattern of results found in Experiment 3 using the same materials: there was a repetition priming effect, and a smaller homophone priming effect (which tended to be smaller for low-frequency homophones).

The harmonic means of naming latencies in the mediated priming conditions were analysed by 2 (Mediator Frequency: Higher vs. Lower) $\times$3 (Prime type: H2S vs. Unrelated) related ANOVAs by participants and by items. There were no significant main effects of priming type, $F_1 < 1$, $F_2 < 1$, or mediator word frequency, $F_1 < 1$, $F_2 < 1$. Also, there was no interaction between the two factors, $F_1 < 1$, $F_2 < 1$.

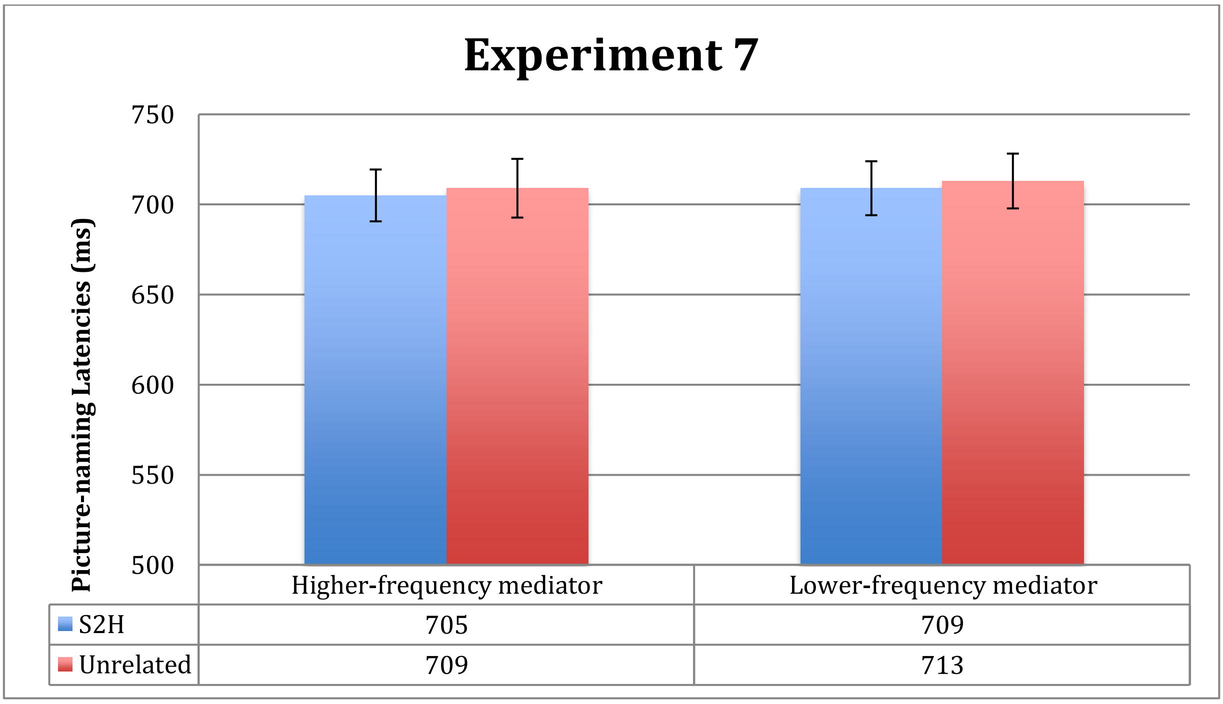


Figure 4.5 Mean Naming Latencies (ms) With Standard Errors in Each Condition

Table 4.3 Mean Picture-Naming Latencies (M) with Standard Deviation (SD) for the Direct Priming Conditions in Experiment 7 and 8

		Experiment 7 (N = 26)		**Experiment 8 (N = 19)**	
		M	**SD**	**M**	**SD**
	Picture name	673	164	614	163
	Control[a]	714	162	684	153
Word Frequency	**Priming Type**				
High					
	Homophone	700	163	652	160
	Phonological	709	166	666	152
	Unrelated	708	160	675	155
Low					
	Homophone	699	168	662	153
	Phonological	726	174	675	160
	Unrelated	739	177	684	148

a: The words in the homophone, phonological and picture name priming groups were pooled together and then each word was paired with a picture that is phonologically unrelated to the word.

3. Discussion

Experiment 7 examined possible semantic-to-homophone mediated priming, but found that presented prime words that were semantically related to homophones of the target object names (e.g., *either—OAR*, via 'or') did not influence object naming times. There was no S2H mediated priming. It is likely that reading words activate their semantic representations that then produce, by a process of semantic-level spreading activation, an increase in the activation levels of semantically related lexical representations. (This is the explanation of the direct semantic priming effect found in Experiment 6, although this was not found in Experiment 5.)

Reading aloud words does not appear to activate the semantics of their homophones; for example, reading the word *either* may activate the lexical representation of 'or', but either this does not then prime its phonological-level forms or that any activation it produces does not persist to affect subsequent object naming times to OAR. This suggests either that any internally activated semantically related word does not activate its lexical phonology, or that such an effect was not detected in the word-to-picture priming studied here. It is important to note that the word frequency manipulation in the current experiment refers to the frequency of the mediator words (and not to the actual prime words used), and so the absence of an effect of this factor does not dispute the word-frequency effects described in Chapter 3.

Experiment 8 Homophone Mediated Priming of Object Naming When Reading Prime Words Silently

1. Method

1.1 Participants

Another nineteen students, aged between 18 and 25 years, were recruited from the University of Essex. All were native Mandarin speakers who had lived in Mainland China for more than 15 years.

1.2 Stimuli, Design and Procedure

The stimuli, experimental design, and experimental procedure were the same as in Experiment 7 with the exception that participants did not read aloud the prime words; instead, they were instructed to read them silently.

2. Results

The same procedure for data trimming as in the previous experiments was applied, and 7.9% of picture naming responses were excluded. The mean naming latencies in the mediated priming conditions are presented in Figure 4.6. The harmonic means of naming latencies were analysed using 2 (Mediator frequency: Higher vs. Lower) × 3 (Prime type: H2S vs. Unrelated) related ANOVAs. There were no main effects of priming type, $F_1 < 1$, $F_2 < 1$, or of mediator frequency, $F_1 < 1$, $F_2 < 1$. Also, there was no interaction between these two factors, $F_1 < 1$, $F_2 < 1$.

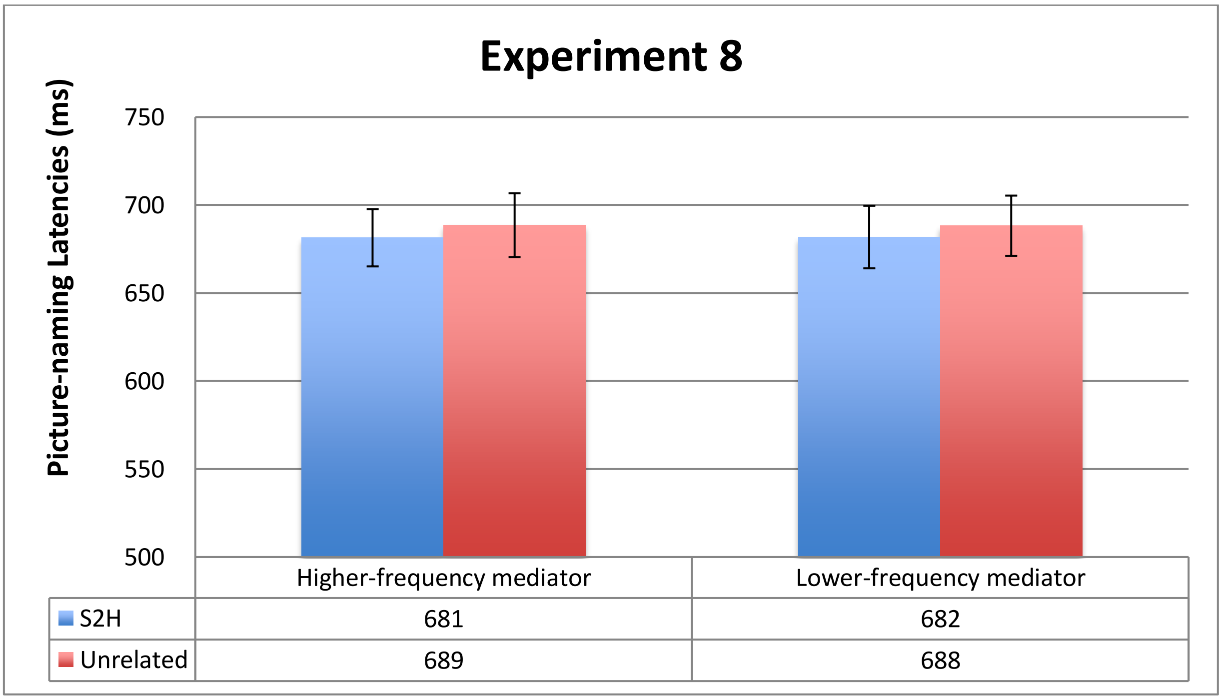


Figure 4.6 Mean Naming Latencies (ms) with standard Errors in Each Condition

3. Discussion

The results of Experiment 8 were, disappointingly, in line with the findings from Experiment 7, where prime words were read aloud. When reading a word silently, its semantically related words might be activated, but the homophones of internally generated semantically related did not become phonologically activated to an extent that was detected by the priming study used here. As such, the results provide no support for the cascading and interactive models of spoken word production.

The results from the identical primes and the high- and low-frequency homophones (and their unrelated controls), and presented in Table 4.3, essentially replicate those already reported from Experiment 4.

General Discussion

The general conclusion of the results of the four experiments reported here is that they provide no support for the existence of any mediated priming effects. Experiments 5 and 6 did not detect any evidence of either homophone-to-semantic (H2S) or phonological-to-semantic (P2S) mediated priming. Experiments 7 and 8 did not detect any evidence of semantic-to-homophone (S2H) mediated priming. When an experiment fails to detect a theoretically predicted effect, there are always two possible reasons: they may be no effect to detect, or the experiment, unfortunately, failed to detect it. There is accumulating evidence to suggest that spoken word production in Chinese reflects serial and discrete processing (e.g., Zhang et al., 2018), and so there really may be no mediated effect indicative of interactive processing to detect. It is always difficult to conclude that an experiment that fails to detect an effect was the "perfect" test of its detection. After all, all experiments could be improved in terms of their selection of stimulus materials and their experimental power.

The experimental paradigm used in the current experiments was that of sequential priming: on each trial, participants read a word and then named a picture. The reasoning behind this task was that the processing of the prime word would leave persisting activation at stages of processing that are also used in the naming of picture objects, which should, therefore, benefit from this pre-activation. It is possible that, if mediated priming effects are actually very short-lived, then the priming task used here was not able to detect them (as the prime word appeared too long before the onset of the stimulus object).

However, it should be noted that the priming task used here was able to detect effects of repetition, homophone, and phonetic priming (in Experiments 3 and 4), and repetition and semantic priming in Experiment 6. It is possible that mediated priming effects may only be observed under particular circumstances, particularly when either semantic or phonological activation is boosted compared to usual circumstance (e.g., Zhang et al., 2017; Abdel Rahman & Melinger, 2008). For instance, Oppermann et al. (2010) observed a mediated effect of auditory-presented distracters only when the mediator was presented as context picture. Future research will be required to further test for mediated priming effects, perhaps using the picture-word task where the words used as primes in Experiments 5 to 8 could be presented as distractors in conditions designed to contextually boost either semantic or phonological activation.

One interesting and surprising result revealed by the contrast between Experiments 5 and 6 was that the semantic priming effect was observed when the prime word was read silently but not when it was read aloud. (This also showed the opposite pattern to that observed concerning the effect of prime word frequency in Experiments 3 and 4, where it was suggested that the prime word frequency effect has its locus at the stage of activating syllable motor programs by the act of actually articulating prime words.) The detection of a semantic facilitation priming effect when reading word silently but not aloud may find an explanation in the concept selection model (CSM) proposed by Starreveld and La Heij (1996), which has been developed and modified by (Bloem & La Heij, 2003; Bloem, van den Boogaard, & La Heij, 2004). This proposes that semantic facilitation is localized at the conceptual level and only one concept is selected for lexicalization, and that lexical representations decay faster than conceptual representations. During the silently reading task, the priming word was presented 1.5 seconds before the presentation of the target picture. If the prime word activated both its concept and lexical nodes, concept-level activation may either be less active in the task of reading a word aloud or it may decay more rapidly. This account could be tested in future experiments by the manipulation of the SOA of the prime and target object.

In conclusion, the results of the present series of experiments found no evidence to support the idea that there is feedback from the level of phonological encoding to the level of lexical selection during Chinese word production. The results obtained are more consistent with the serial discrete model of word production in Chinese. Indeed, it is entirely possible that there are differences between languages in whether processing is cascaded or discrete (Zhang et al., 2017).

CHAPTER 5 PRIMING OBJECT NAMING VIA TRANSLATIONS IN CHINESE-ENGLISH BILINGUALS

There is a large body of evidence showing that proficient bilinguals activate words in both of their two languages (and also to the phonological level), even when only one language is required (e.g., Colomé, 2001; Colomé & Miozzo, 2010; Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, De Bot, & Schreuder, 1998). Colomé (2001) tested Catalan-Spanish bilinguals who operated entirely within their first language in a task where participants had to perform phoneme detections in the names of presented objects (e.g., is /d/ in the name of *DOG*). She found that the bilinguals were slower to respond "no" when their second language translation of the object name contained the target phoneme (e.g., deciding if /p/ in the name of a picture of a DOG, whose name is "gos" in Catalan and "perro" in Spanish). These results show that target objects activate the phonological forms of their names in both languages. Colomé and Miozzo (2010) used the picture-picture task (e.g., Morsella & Miozzo, 2002), in which participants are presented with two spatially overlapping line drawings and were instructed to name only the one shown in green ink. Colomé and Miozzo tested Spanish-Catalan bilinguals who named the objects only in Catalan and found that distractor pictures whose names in Spanish were phonologically related to the targets facilitated naming. These results show that even non-target objects activate the phonological forms of their names in both languages. Indeed, Meuter (2005) concluded that all related lexical representations in both languages are activated and remain available until fairly late in the selection process.

Costa, Caramazza, and Sebastian-Galles (2000) compared naming times to objects with cognate and non-cognate names. Cognates are words that sound similar in both languages. For Catalan-Spanish bilinguals, naming latencies were shorter for pictures whose names are cognates rather than non-cognates. These results suggest that both Spanish and Catalan names were activated in by the bilinguals during the lexical selection. When picture names were cognates, both their Spanish name and Catalan name activate shared phonological-level representations that facilitate naming. The cognate facilitation in bilinguals shows that non-target (and so non-selected) lexical nodes are also phonologically encoded.

The previous chapter reported four experiments whose results provided no evidence of any semantically or phonologically mediated priming in the priming of naming in Chinese. The absence of mediated priming is consistent with the view that Chinese spoken word production operates by serial and discrete processing rather than by the cascading and interactive information flow proposed to explain word production in European languages. However, it is possible that the mediated prime words used in the experiments reported in Chapter 4 may be somehow inefficient in their activation of non-target lexical nodes; the presumed mediator words may have been only weak (or quickly fading) activation. Cross-language translations are necessarily very closely related semantically; indeed, for many translations, they will be semantically identical. The experiments reported in this chapter will investigate possible mediated cascading of activation via translations. All of the participants in the experiments reported in this thesis were tested in Essex; they had Chinese as their first language but were currently living in the U.K. and studying in English. As such, they are Chinese-English bilinguals, and it is possible that they will co-activate both Chinese and English words when processing words and object names in Chinese.

Q. Zhang and Zhu (2016) investigated the possible cascading of activation via translations using a word translation variant of the picture-word task, in which an English target word (presented superimposed on a context picture) had to be translated into Chinese. In their first experiment, the target word appeared on a semantically related or unrelated context picture, or upon a picture with a phonologically related or unrelated name. (All picture names and word translations were disyllable words.) In the phonologically related condition, the first characters of the Chinese translations and picture names shared the same syllable but not the same tone; for example, the translation of the stimulus word *banana* is 香蕉, pronounced /xiang1 jiao1/, and the Chinese name of a picture of a CAMERA is 相机, pronounced /xiang4 ji1/, which has the same atonal first syllable as the translation. Zhang and Zhu found that translation latency was facilitated when the context picture was semantically related to the word compared to when the picture was unrelated in meaning. This effect suggests that the semantic information provided by the context picture was activated and assisted the translation process. According to Roelofs (2006a), context pictures activate their concept nodes and the lexical nodes of their names regardless of whether a speaker actually produces these names. Zhang and Zhu found no difference in translation times when the picture names were phonologically related or unrelated. Further, this null effect of phonological relatedness was replicated when the proportion of phonologically related trials was increased from 25%

in the first experiment to 50% in the second experiment. The absence of a phonological effect indicated that the phonological representation of context picture's name was not activated in Chinese (despite the activation of its semantics). These findings support a serial discrete view of processing in which only the phonology of a lexically selected word becomes activated.

The absence of a phonological effect was also observed in a similar translation task with both Dutch-English (Bloem & La Heij, 2003; Bloem et al., 2004) and Spanish-Catalan bilinguals (Navarrete & Costa, 2009). Bloem et al. (2004) explained this absence of any phonological effect from context pictures using a modified version of Starreveld and La Heij (1996) conceptual selection model (CSM). According to the CSM, non-verbal contextual stimuli will active their conceptual-semantic representations but will not automatically activate their lexical representations. During lexical access for speech, only one concept, the one that receives "task activation" or "signalling activation", reaches a threshold for selection (i.e., entry into the process of lexicalization). This view contrasts with cascading models that propose that *all* activated concepts will activate lexical-level (and phonological-level) representations. One difference between selection at the concept-level compared to selection at the lexical-level is that the CSM proposes that there is no competition at the conceptual level, whereas lexical representations will compete for selection (but see Miozzo & Caramazza, 2003, for a different view). It should be noted that the modified CSM also assumes some degree of cascaded processing; this explains Bloem and La Heij's (2003) finding that context words produce a phonological facilitation effect in word translation, whereas context pictures did not.

However, the CSM fails to explain the phonological facilitation effect observed in the picture-picture task (Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005; Roelofs, 2008). This phonological effect suggests that activation spreads continuously from concept to lexical to phonological levels, even for the names of non-target pictures. A possible explanation of the discrepancy between studies that show phonological effects and those that do not is that target selection was harder when target pictures must be perceptually distinguished from distractor (or context) pictures, while target selection was easier in picture-word interference task. A difficulty in target selection would result in more attention being allocated to the context picture and so increases the activation of its name. In order to avoid this problem of divided attention, the present study will continue to use the prime word to object naming paradigm, in which the concept representation of the prime words will be certainly activated, but not necessarily to the level of their phonological representations.

The Present Study

The two pairs of experiments reported in this chapter investigated whether cascading processing operates in Chinese object naming preceded (i.e., primed) by word reading or word translation. In Experiments 9a and 9b, participants were asked to read aloud or translate a Chinese prime word into English (e.g., 零 is pronounced as /ling2/, and would be translated to “zero” in English), and then name a picture in Chinese that had either a homophone name (e.g., 铃, which is pronounced as /ling2/, and means “bell”) or a phonologically unrelated name. The results of Experiments 3 and 4 (reported in chapter 3) showed that reading a Chinese homophone primes (i.e., facilitates) object naming, by the pre-activation of its phonological-level representation. According to the cascading processing view, when translating the prime word 零, the phonological representation of its Chinese name /ling2/ (and its homophones) should also be activated, even though /ling2/ is not the target word to produce (as an English translation must be spoken). If this is true, then translating 零 to say “zero” should also prime the naming of a picture of 铃 (a homophone of 零). According to the serial discrete processing view, no such homophone priming should be observed, because the Chinese word 零 is not a target response.

In order to rule out the possibility that any pre-activation of shared phonological nodes comes directly from orthographic nodes, the prime words in Experiments 10a and 10b were English words (and English and Chinese are very distinct). The Chinese translations of the stimulus English words were either homophones of the picture names or phonologically unrelated words. According to the cascading processing view, when translating the word *zero* into Chinese word 零, its phonological representation /ling2/ must be activated and so should facilitate the time to name the picture 铃, with a homophonic name. The serial discrete processing view can only explain homophone priming in terms of persisting activation of post-lexical phonological encoding stages necessary for speech production.

Experiments 9a and 9b Reading and Translating Chinese Words

1. Method

1.1 Participants

Twenty-eight students (20 women, aged between 18 and 25 years) were recruited from the University of Essex. All were native Mandarin speakers who had lived in Mainland China for more than 15 years. All were also highly proficient in English and had an overall IELTS score of 6.5 and above. Most participants were postgraduate students studying Chinese-English translation and interpreting. English-to-Chinese and Chinese-to-English translation tests were performed when recruiting. 50 Chinese words (characters) and 50 English words (which were created during the preparation of stimuli, but not used in the experiments) were given to candidate participants. Only candidates with an error rate of less than 10% were selected for participation in the experiments. The same recruitment procedure was used for Experiment 10a and 10b.

The participants tested in the previous experiments reported in this thesis were also bilingual in English, as they were from the student community of the University of Essex (and were studying in English). However, their proficiency in English was not formally assessed, as the previous experiments were conducted solely in Chinese.

1.2 Stimuli and Design

The experiment contained two blocks of trials. In each trial of Experiment 9a, participants had to read a written Chinese word and then name a picture in Chinese. In each trial of Experiment 9b, participants had to translate a written Chinese word into English and name a picture in Chinese. For the picture naming task, twenty-eight pictures used in previous experiments were initially selected from Snodgrass and Vanderwart (1980). The names of these pictures were all monosyllabic words. Each picture was combined with four prime words: a higher frequency homophone; a lower frequency homophone; and two unrelated control words matched on word frequency and the number of strokes to the homophone primes. Prior to the experiments, two postgraduate students studying Chinese-English translation and interpreting were asked to translate all the prime words. Words that were hard to translate or had too many alternative meanings were excluded. Finally, thirteen pictures with their prime words were selected in the present experiment. The picture names had a mean frequency per million of 76.93 (SUBTLEX-CH corpus; Cai and Brysbaert (2010). Table 5.1 provides a summary of the characteristics of the prime words used in Experiment 9.

Table 5.1 Mean Characteristics of The Prime Words Used in Experiments 9a and 9b

Condition	Example (with English translation)			Mean Number of Strokes	Mean Frequency per Million	Frequency Range per Million
Picture Name	铃	ling2	Bell		38.26	7.1 – 323
Homophone						
Higher Frequency	零	ling2	Zero	8.85	366.86	25.81 – 1830.13
Lower Frequency	陵	ling2	Mount; tomb	11.60	2.63	0.23-18.85
Phonologically Unrelated						
Higher Frequency	频	pin2	Frequency	6.77	378.76	45.41-1389.40
Lower Frequency	噎	ye1	Choke	10.85	3.06	0.23-18.85

In order to reduce the proportion of homophone word-object pairs (to mitigate against any strategic bias towards predicting the names of pictures), unrelated filler trials were also included. In Experiment 9a, the related prime-pictures pairs were randomly re-ordered to generate 26 unrelated filler trials, ensuring that these were not phonologically or orthographically related. The same procedure was used to generate another 26-unrelated filler trials for Experiment 9b. Therefore, a total of 78 word-picture stimulus sets were generated for each experiment, with 26 homophone trials (13 with higher and 13 with lower frequency homophones), 26 control trials (13 higher and 13 lower frequency), and 26 fillers. The presentation order of Experiments 9a and 9b was counter-balanced over sub-groups of participants. Within each experiment, the order of trials was pseudo-randomized for each participant, with the constraint that pictures did not repeat in three consecutive trials.

1.3 Procedure

The experimental stimuli were presented using the software Superlab, run on an Apple Macintosh computer. Superlab also recorded the latencies of all responses. Participants were tested individually in a partially soundproofed booth. Before the experiment, participants were familiarized with both the pictures and words. They were asked to provide the most appropriate monosyllable names to the pictures and were corrected if their responses were not the target names. Then they were asked to provide an English translation to each word, and were corrected if they did not produce the intended word.

For each trial, the participant first saw a fixation point for 360ms. The prime word was then presented and remained on the screen until the participant's response was detected by the computer's built-in microphone that activated Superlab's voice-key. After the character disappeared, there was an inter-stimulus interval for 500ms. The target picture was then presented, which picture remained on the screen until the participant initiated a vocal response. After the participant's response was detected, there was an inter-trial interval of 700ms. The same procedure was used in both Experiments 9a and 9b, but with different instructions. In Experiment 9a, participants were asked to read the Chinese word presented aloud and then name the picture in Chinese as quickly as possible. In Experiment 9b, participants were asked to translate the presented Chinese word into English and then name the picture in Chinese as quickly as possible. A short break was given after 40 trials and a longer break was given between experiments. Response times were measured from the onset of the character or picture to the onset of articulation. The whole experiment was recorded for later inspection. The experimenter stayed outside the booth but could hear the participants and control the experiment. Trials on which there were hesitations or errors were excluded from the analysis.

2. Results

Responses were excluded using the same criteria as in previous experiments. Four participants were excluded because of a high error rate. For the picture naming latencies, 2.5% of responses were removed in Experiment 9, and 2.9% were removed in Experiment 10.

2.1 Experiment 9a: word reading times

The mean word reading latencies in the different conditions are presented in Figure 5.1. Harmonic mean times in each condition were analysed using 2x2, Frequency (higher vs. lower) x Word type (homophone vs. phonological unrelated) ANOVAs. The main effect of prime type was marginally significant by participants, $F_1(1, 23) = 4.32$, $MSE = 1166.9$, $\eta_p^2 = .16$, $p = .049$, but not by items, $F_2 < 1$. Word reading latencies were slightly faster to homophone primes than to unrelated words. The main effect of frequency was significant: $F_1(1, 23) = 72.5$, $MSE = 2550.6$, $p < .001$, $\eta_p^2 = .76$; $F_2(1, 12) = 25.5$, $MSE = 3843.5$, $\eta_p^2 = .68$. Word reading latencies were faster to higher frequency prime words than to lower frequency words. The interaction between the two factors was significant by participants, $F_1(1, 23) = 5.57$, $MSE = 775.3$, $p = .027$, $\eta_p^2 = .20$, but not by items, $F_2 < 1$.

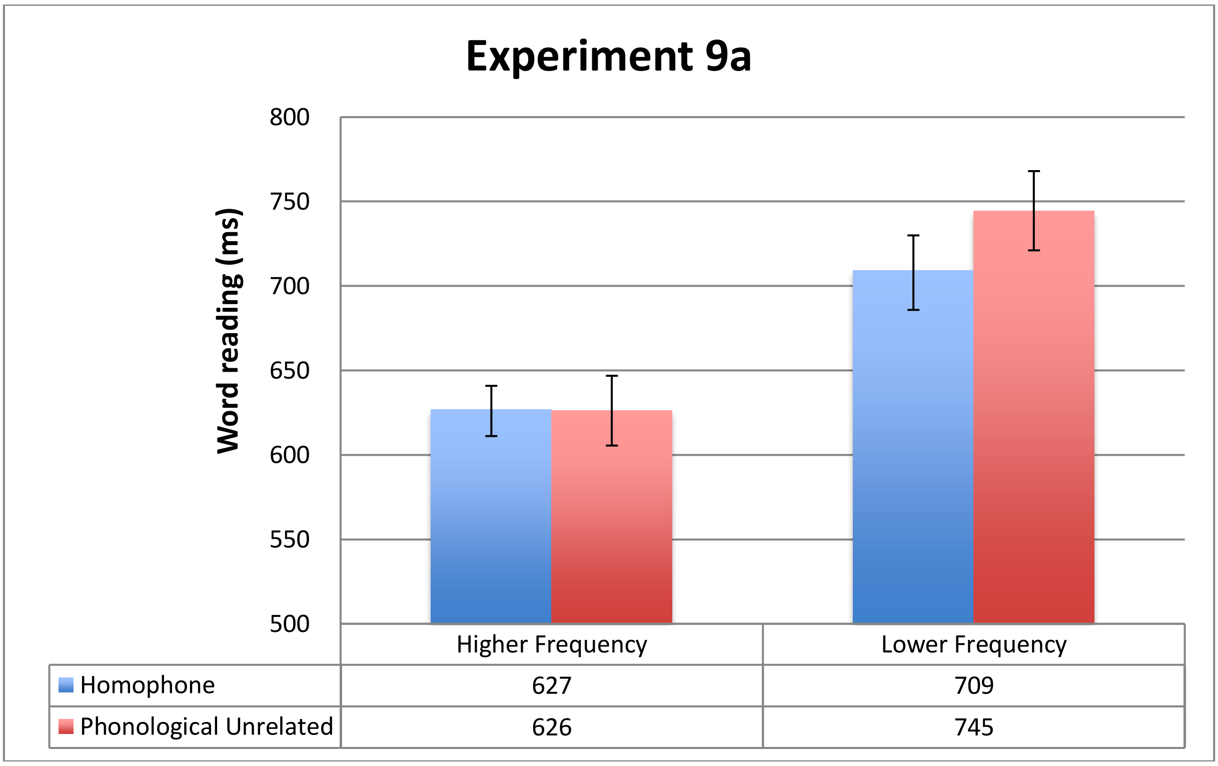


Figure 5.1 Mean Reading Times (with Standard Errors) to Prime Words in Experiment 9a

2.2 Experiment 9a: picture naming times

The mean naming latencies in each condition are shown in Figure 5.2. Harmonic means of naming times were analysed by 2 x 2 frequency (higher vs. lower) by prime type (homophone vs. unrelated) ANOVAs. The main effect of prime type was significant: $F_1(1, 23) = 5.07$, $MSE = 1312.0$, $p < .001$, $\eta_p^2 = .18$; $F_2(1, 12) = 9.97$, $MSE = 7742.6$, $p = .008$, $\eta_p^2 = .45$. Naming times were faster following a homophone prime compared to an unrelated prime, showing a clear homophone priming effect. The main effect of frequency was significant in the participants analysis, $F_1(1, 23) = 8.77$, $MSE= 1312.0$, $p=.007$, $\eta_p^2= .28$, but not in items analysis, $F_2(1, 12) = 2.452$, $p= .0131$. Naming times were faster with higher frequency primes compared to lower frequency primes. The interaction between the two variables was absent, $F_1 < 1$, $F_2 < 1$.

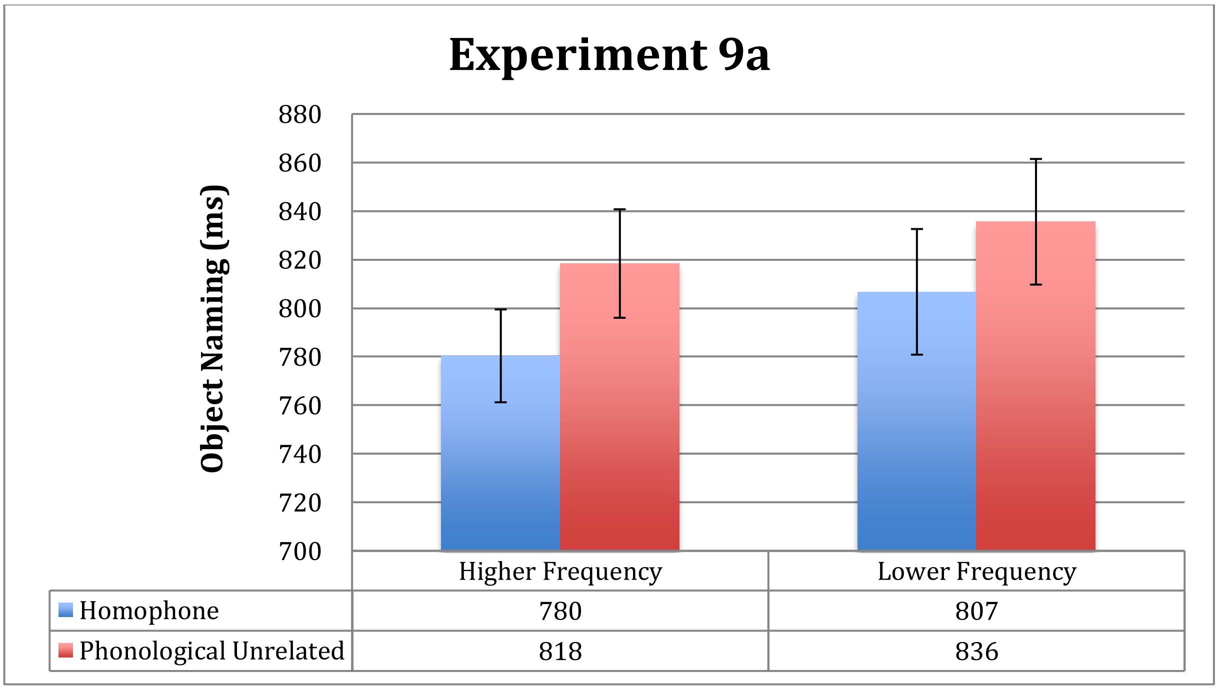


Figure 5.2 Mean Object Naming Times (with Standard errors) by Conditions in Experiment 9a

2.3 Experiment 9b: word translation times

The mean word translation times in each condition are presented in Figure 5.3. Harmonic means of translation times were analysed by 2 x 2 frequency (higher vs. lower) by word type (homophonic vs. controls) ANOVAs. There was no significant main effect of word type; $F_1(1, 23) = 1.32$, $p = 0.26$; $F_2 < 1$; homophones took no longer to translate than frequency matched control words. The main effect of frequency was significant by participants, $F_1(1, 23) = 12.7$, $MSE = 2091.9$, $p = .002$ $\eta_p^2 = .36$, but not by items, $F_2 < 1$. The interaction of these two variables was marginally significant by participants, $F_1(1, 23) = 4.35$, $MSE = 3221.0$, $p = .048$, $\eta_p^2 = .16$, but not by items, $F_2(1, 12) = 1.50$, $p = .244$

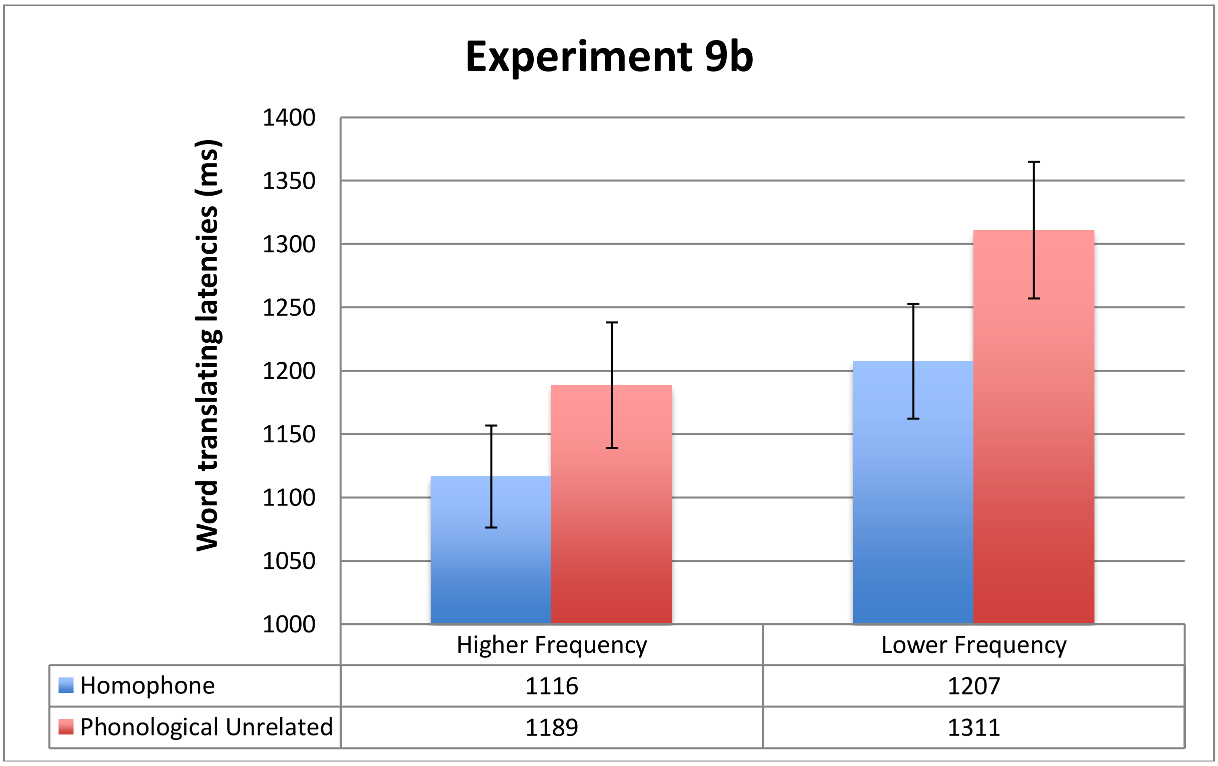


Figure 5.3 Mean Translation Times (With standard errors) to Prime Words in Experiment 9b

2.4 Experiment 9b: picture naming times

The mean object naming latencies in each condition are presented in Figure 5.4. Harmonic mean naming times in each condition were analysed by 2 x 2 frequency (high vs. low) by prime word type (homophone vs. unrelated) ANOVAs. The main effect of prime type was significant; $F_1(1, 23) = 24.6$, $MSE = 3931.6$, $p < .001$, $\eta_p^2 = .52$, $F_2(1, 12) = 21.90$, $MSE = 2606.1$, $p = .001$, $\eta_p^2 = .63$. Object naming times were faster following homophone primes than unrelated words, showing a clear homophone priming effect. The main effect of frequency was absent significant, $F_1 < 1$, $F_2 < 1$. The interaction between the two factors was just significant by participants, $F_1(1, 23) = 4.67$, $MSE = 1834.6$, $p = .041$, $\eta_p^2 = .17$, but not by items, $F_2(1, 12) = 1.69$, $p = .218$.

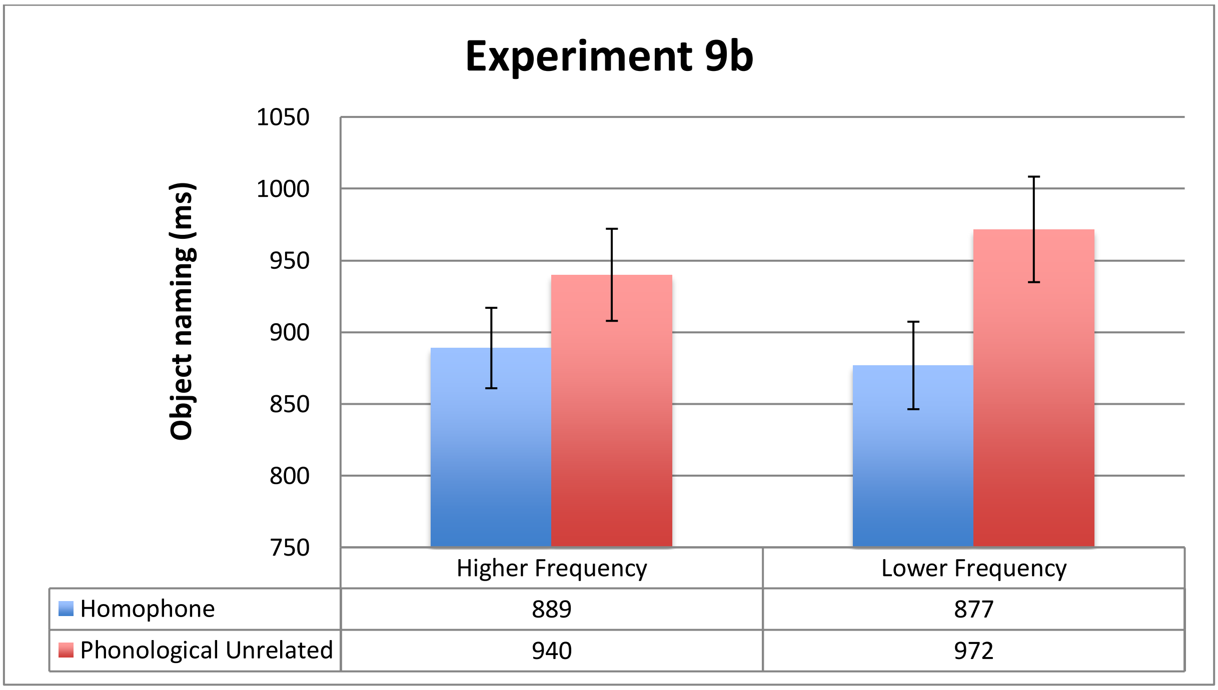


Figure 5.4 Mean Object Naming Times (with Standard Errors) to target pictures in Experiment 9b.

3. Discussion

Experiment 9a replicated the word reading results found in Experiments 1 and 2: reading times showed a clear frequency effect, but no frequency by homophone interaction. Object naming times showed a clear homophone priming effect (that was essentially the same for high- and low-frequency homophone words), which replicated the results found in Experiment 3 and 4. Experiment 9b showed that word translation times did not differ for homophone and unrelated words but were slightly faster to high frequency words. The word translating latencies (about 1200ms) were longer than simple reading times (about 670ms) or naming latencies (around 800-900ms). The pattern was similar to the finding by Q. Zhang and Zhu (2016). In their translation experiment, average latencies were about 1100ms, while the picture naming latencies were about 900ms. Many studies with English speakers have found that word reading times are faster than object naming times. The fact that translation times are slower than naming times probably reflect the greater level of both semantic and linguistic processing required to translate words.

Despite the fact that no Chinese words were articulated in Experiment 9b, the visually presented words produced a homophone priming effect on object naming times. In fact, the homophone priming effect was somewhat larger in Experiment 9b than in 9a, where the prime words were read aloud; 63 vs. 24ms. This difference in the magnitudes of the priming effects suggests that translating words may be a more cognitively taxing task than reading words aloud, but the key finding is homophone priming of object naming is found even though the participants do not actually articulate the Chinese word. This shows that the phonological facilitation from homophones emerges even when a lexical node is not selected (and, as some theories of bilingual speech have suggested, may have had to be inhibited in order to produce the English translation). Word translation was faster for higher frequency than for lower frequency words, but the frequency of the prime word had no effect on picture naming latencies. The analysis of word response times ruled out the possibility that the phonological effect arose because homophones are intrinsically easier or harder to translate.

These results show that non-selected lexical nodes can pass activation to their corresponding phonological nodes as predicted by the cascading activation view. When translating a Chinese word 零 to its English equivalent "zero", the visual word recognition system would activate the words' common semantic representation. Semantic information would then activate the lexical representations of both English and Chinese names. A language control system is responsible for bilinguals choosing to speak in one language, to translate words, and to switch between languages when required (e.g., Green, 1998; Dylman and Barry (2018). This language control system would increase the activation level of the target lexical representation "zero" which activates its phonological encoding and articulation. The lexical node for 零 may also receive extra activation from the direct reading route (if it exists), and also some from any inter-lexical translation connection (if these exist). The phonological facilitation on picture naming suggests that the phonological representation of "ling2" was pre-activated by the non-target lexical nodes of 零, indicating cascading activation from lexical level to phonological level representation.

A possible source of phonological pre-activation is from a non-lexical reading route, but there exists no convincing evidence that such a process exists in fluent Chinese readers (Law et al., 2009; Perfetti et al., 2005). The basic writing units of Chinese are characters, which have highly arbitrary symbol–sound correspondences (X. Zhou et al., 1999).

However, to exclude the possibility that phonological-level pre-activation arises from a non-lexical reading route, the prime words in Experiment 10a and 10b will be English words, whose translations were entirely phonologically unrelated to the picture names.

Experiments 10a and 10b Reading and translating English words.

In this pair of experiments, participants were presented with printed English prime words, which they either had to read aloud or to translate into Chinese. The Chinese translations of these words were either homophones of the names of the subsequently presented pictures or were phonologically unrelated.

Reading aloud an English word could not produce any activation at the orthographic stage of Chinese word recognition, and the only way it would activate the stages of Chinese word production would be if automatically activated its non-target Chinese translation. Whether such lexical-level activation then cascades to the phonological-level will be a theoretically important question.

Translating an English word into a Chinese word to produce in speech will involve both language-independent semantic processing, and language-specific phonological processing to say the Chinese translation. It is likely to also involve lexical-level representations of both English and Chinese words. As translating English words into Chinese involves the actual production of Chinese words, these will be expected to produce homophone priming of object naming (in much the same way as reading aloud a homophone would).

The task of translating English words into Chinese also returns the theoretical focus to the nature of the lexical representation of homophones. Chapter 1 reviewed the evidence from studies using this task that attempted to arbitrate between the shared and independent representation hypotheses of how homophones are represented in the speech production system. Jescheniak and Levelt (1994) asked Dutch-English bilinguals to produce the Dutch translation of printed English words. They found that Dutch translations that were low-frequency (homographic) homophones that had a high-frequency homophone mate were produced faster than non-homophones matched on specific-word frequency (and no different than non-homophones matched on cumulative homophone frequency). This frequency inheritance effect provides strong support for the view that homophones share a common lexical phonological representation. Jescheniak, Meyer, and Levelt (2003) replicated this pattern of results with English-German bilinguals, but Caramazza et al. (2001) failed to replicate it with both English-Spanish and Chinese-English bilinguals.

1. Method

1.1 Participants

Another group of twenty-seven students (17 women; aged between 18 and 25 years), were recruited from the student body of the University of Essex. All were native Mandarin speakers who had lived in Mainland China for more than 15 years and were currently studying in English. They were all highly proficiency in English with an overall IELTS score of 6.5 and above.

1.2 Stimuli and Design

The experiment comprised two blocks of trials. In Experiment 10a, each trial involved reading aloud an English word and then naming a pictured object in Chinese. In Experiment 10b, each trial involved translating visually presented English words into Chinese followed by object naming in Chinese.

For the picture naming task, twenty-four pictures were selected from Snodgrass and Vanderwart (1980) used in previous experiments. The picture names were all monosyllabic. Two master students attending the Chinese-English translation and interpreting course generated Chinese translations, which were all disyllable words. Each picture was combined with four English prime words. For the homophone condition, the first character of the Chinese translation was a homophone (e.g., *beard* = 胡子, /hu2 zi1/) of the picture name (壶, /hu2/), and this homophone of either higher or lower word frequency. For the translations of unrelated prime words (e.g., *villa* = 别墅, /bie2 shu4/), neither of character was phonologically related to the picture name. The words in the homophone and unrelated conditions were matched in word length and word frequency (from Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The picture names had a mean frequency of 64.73 per million. Table 5.2 shows the mean characteristics of the stimuli.

Table 5.2 Characteristics of the Stimuli Used in Experiments 10a and 10b

Condition	Example (With Translation)		Mean English Word Frequency (Per Million) (SD)	Mean Chinese Translation Word Frequency (Per Million) (SD)
Picture Name	Rat; mouse	鼠 shu3		
Higher frequency homophone	Category	属 shu3		
Phonological related	Metal	金属 jin1 shu3	10.37(10.42)	24.97(34.37)
Phonological Unrelated	Trade	交易 jiao1 yi4	10.17(13.64)	10.02(11.80)
Lower frequency homophone	Potato	薯 shu3		
Phonological related	Chips	薯条 shu3 tiao2	11.57(14.29)	5.68(5.15)
Phonological Unrelated	Aunt	阿姨 a1 yi2	10.19(7.95)	15.04(21.42)

For Experiment 10a, related prime words and pictures were randomly crossed to form an additional 48 filler trials, which were neither phonologically nor orthographically related to the picture names. The same procedure was used to form another 48 fillers for Experiment 10b. Therefore, a total of 144 word-picture stimulus sets were generated for each experiment. The order of presentation of the experiments was counter-balanced over sub-groups of participants. Within each experiment, the order of trials was pseudo-randomized for each participant, with the constraint that pictures did not repeat in three consecutive trials.

1.3 Procedure

The same experimental procedure as in Experiments 9a and 9b was used, but with different instructions. On each trial in Experiment 10a, participants were required to read the English word and then to name the picture in Chinese as quickly as possible. On each trial in Experiment 10b, participants were required to translate the presented English word and then to name the picture in Chinese. A short break was given after 40 trials and a longer break was given between experiments. Naming times were measured from the onset of the character or picture to the onset of articulation. The experimenter could hear the participants and control the experiment, and trials involving speech hesitations or errors were excluded from the analysis.

2. Results

The same data cleaning procedure was used as in previous experiments. The results of three participants were excluded due to high error rates. For the picture naming latencies, 3.5% of responses were removed in Experiment 10 and 2.9% of were removed in Experiment 10b.

2.1 Experiment 10a: English word reading times

The mean reading latencies to English words in each condition are shown in Figure 5.5. Harmonic mean reading times in each condition were analysed by 2 x 2, ANOVAs, both by participants and by items, with the factors of English word type (homophones vs. unrelated) and frequency of the shared characters in the Chinese translation (higher vs. lower). The main effect of word type was not significant, $F_1 < 1$, $F_2 < 1$. The main effect of character frequency was not significant: $F_1(1, 23) = 3.11$, $p = .091$; $F_2 < 1$. The interaction between the two variables was significant by participants, $F_1(1, 23) = 8.67$, $MSE = 370.5$, $p = .007$, $\eta_p^2 = .27$, but not by items, $F_2 < 1$.

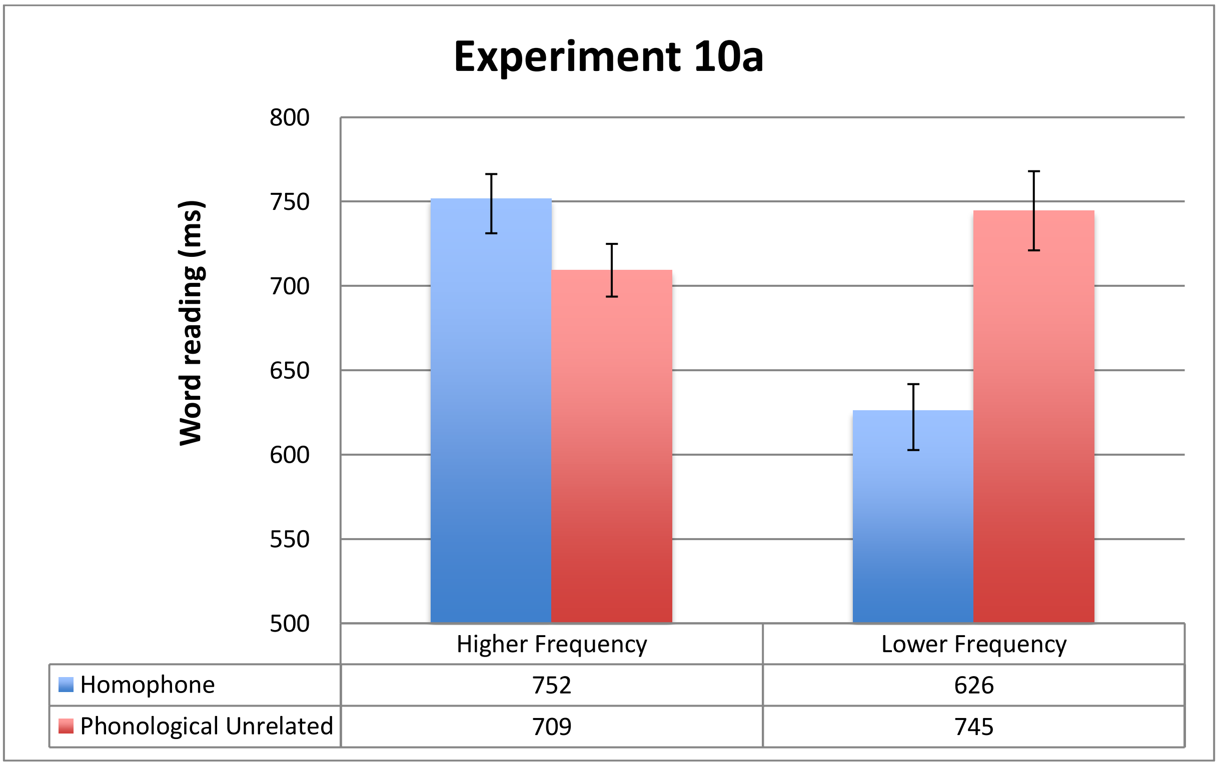


Figure 5.5 Mean Reading Times (with Standard Errors) for English Prime Words in Experiment 10a.

2.2 Experiment 10a: Object naming times

The mean Chinese naming latencies in each condition are shown in Figure 5.6. Harmonic means of naming times were analysed using 2 × 2 ANOVAs with the factors of prime type (homophones vs. unrelated) and frequency (higher vs. lower). The main effect of prime type was significant by participants, $F_1(1, 23) = 7.42$, $MSE= 1588.3$, $p=.012$, $\eta_p^2= .24$, but not quite by items, $F_2(1, 23) = 3.1$, $p= .092$. Object naming times were faster following the production of a homophone of the name than an unrelated word, showing a homophone priming effect. The main effect of character frequency was absent, $F_1 < 1$, $F_2 < 1$. The important interaction between these two factors was significant by analysis of participants, $F_1(1, 23) = 8.56$, $MSE= 701.5$, $p=.008$, $\eta_p^2= .27$, but did not achieve significance in the analysis of items, $F_2(1, 23) = 2.72$, $p= .113$.

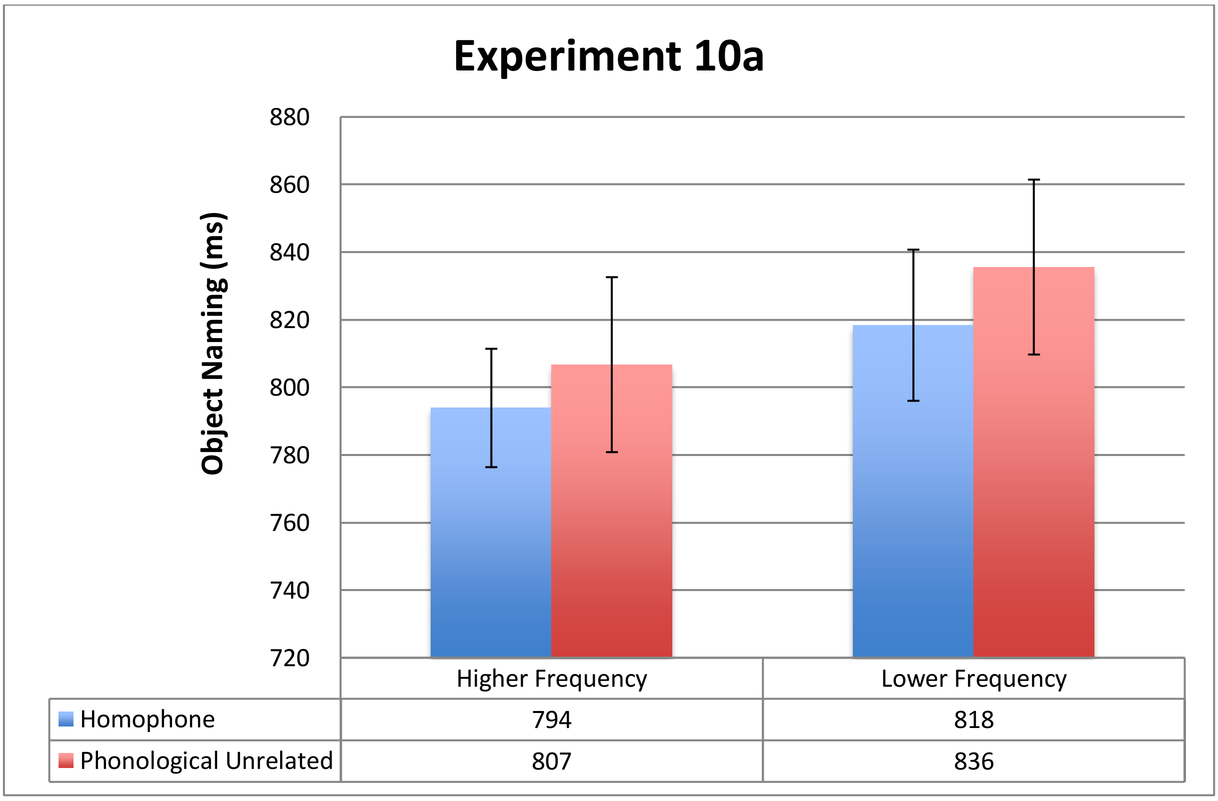


Figure 5.6 Mean Object Naming Times (with Standard Errors) in Experiment 10a.

2.3 Experiment 10b: Word translation times

The mean word translation latencies in each condition are shown in Figure 5.7. Harmonic means of latencies were analysed by 2 x 2 ANOVAs with the factors of frequency (higher vs. lower) and word type (Chinese translation are homophones of, vs. phonologically unrelated to the target object names). The main effect of frequency was significant by participants, $F_1(1, 23) = 12.7$, $MSE = 2092$, $p = .002$, $\eta_p^2 = .36$, but not by items, $F_2 < 1$. Word translating times were faster to higher frequency than to lower frequency words. The main effect of word type was not significant, $F_1(1, 23) = 1.33$, $p = .261$, $F_2 < 1$. The interaction between these two factors was marginally significant by participants, $F_1(1,23) = 4.35$, $p = .048$ and absent by items, $F_2(1,23) = 1.02$, $p = .32$.

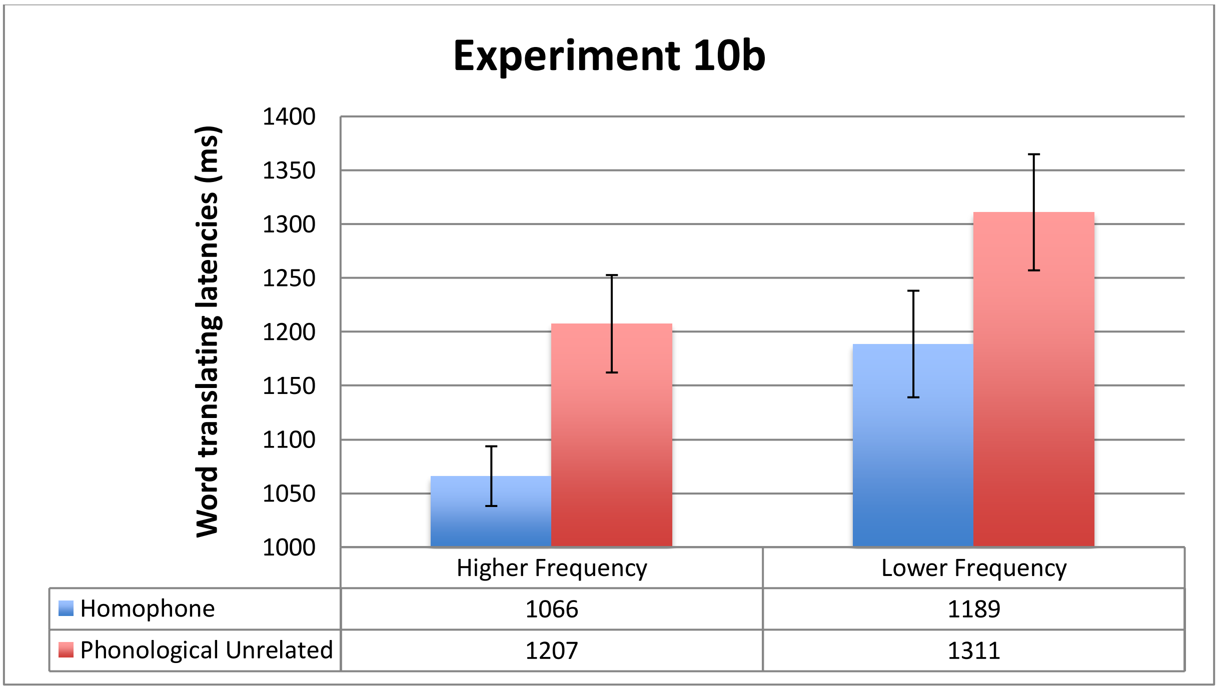


Figure 5.7 Mean English-To-Chinese Translation Times (with Standard Errors) to Prime Words in Experiment 10b

2.4 Experiment 10b: Object naming times

The mean object naming latencies in each condition are presented in Figure 5.8. Harmonic mean naming times were analysed by 2 x 2 ANOVAs with the factors of shared character frequency (higher vs. lower) and prime word (homophones vs. unrelated). The main effect of prime word type was significant: $F_1(1, 23) = 12.62$, $MSE= 2060.2$, $p=.002$, $\eta_p^2= .35$; $F_2(1, 23) = 10.22$, $MSE= 3272.8$, $p=.004$, $\eta_p^2= .31$. Naming times were faster following the production of a homophone than an unrelated word, a clear homophone priming effect. The main effect of character frequency was not significant, $F_1 < 1$, $F_2 < 1$. The interaction between the two factors was not significant, $F_1(1,23) = 2.5$, $p = .13$, $F_2(1,23) = 1.11$, $p = .30$.

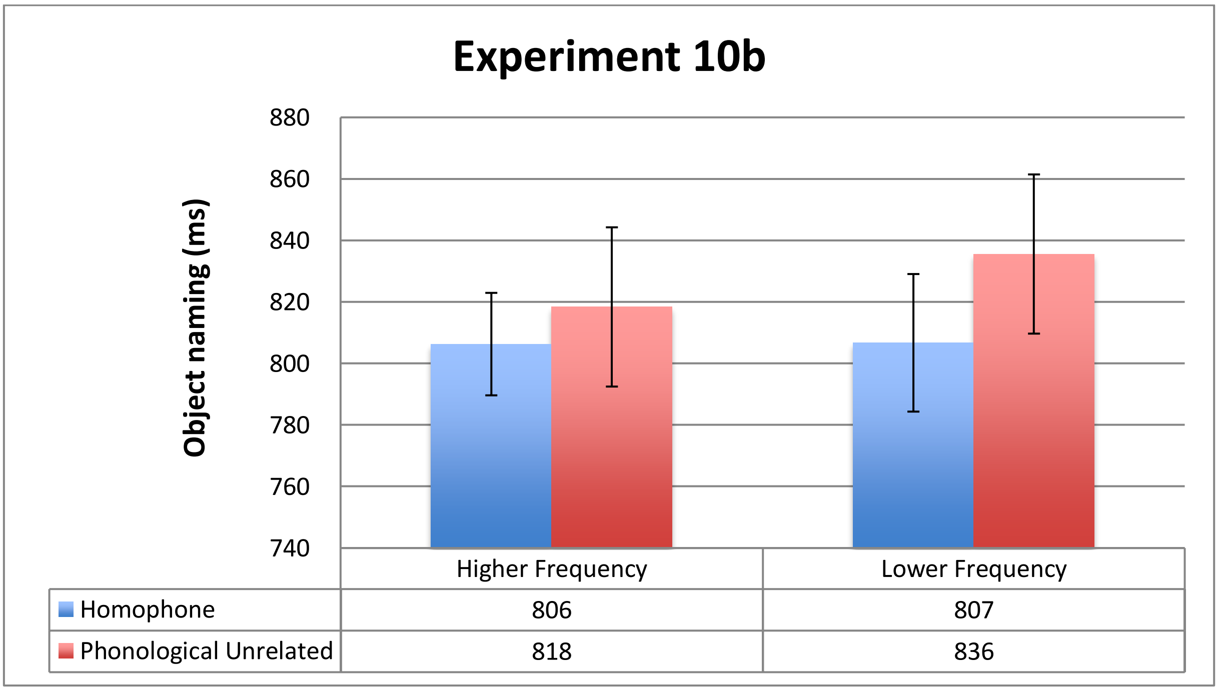


Figure 5.8 Mean Object Naming Latencies (with Standard Errors) in Experiment 10b

2.5 Regression analyses

The word frequencies of the English prime words were matched between the homophone and unrelated sets of words, but the Chinese translation-equivalents of these words were not exactly matched on Chinese word frequency. To exclude any potential influence of this, two linear mixed effects models were performed that included fixed effects of prime word type (homophones vs. unrelated), character frequency (higher vs. lower), and the Log transformed English word frequency or Log transformed English reading latency. The maximal random effect structure, random slopes for all fixed effects and all interactions for pictures, and random intercept for subjects, was used. The main effect of English word frequency was not significant (B= -19.03, SE= 15.01, p = 0.21), while the English word reading latency predicted the picture naming latencies (B= -308.83, SE= 55.92, p < .001). A significant main effect of prime type (B= -33.39, SE= 10.25, p < 0.001) and a significant interaction between prime type and character frequency (B= -35.57, SE=14.55, p = 0.015) were observed in models with English word frequency and interactions. Only a significant main effect of prime type (B= -33.39, SE= 10.25, p < .001, B= -29.98, SE= 8.68, p < 0.01) was observed in the model with word reading latency and interactions. Other main effects and interactions were not significant.

As the word frequency of the English prime words in the higher character frequency and lower character frequency group was matched, the reading latencies were not different between these two groups. Moreover, there was no difference between phonologically related and unrelated groups. For picture naming, there were significant differences between different types of priming characters. When reading an English such as *zero*, the phonological representation of its Chinese translation-equivalent “ling2” was also activated, which facilitated the retrieval of the picture’s name. The analysis of word reading time also excluded the possibility that the effect may come from a difference in translating words.

3. General Discussion

The discussion of the results of the two pairs of experiments reported in this chapter will first concentrate on the priming effects found for object naming latencies from the various tasks performed upon the prime stimuli. It will then concentrate on the response latencies to the words used as primes in the various word processing tasks studied.

3.1 Homophone priming of object naming times

Two pairs of experiments investigated whether phonological activation of both target words and non-target (i.e., non-selected) words presented in the prime task will affect lexical selection of object names in the probe task. Experiment 9b showed that when bilingual participants saw written Chinese words but had to actually select and produce their English translations (i.e., the Chinese words were seen but not spoken), this produced a homophone priming effect on object naming latencies (of 73ms). This finding is consistent with the results of Experiment 4 (reported in Chapter 3) that demonstrated that the silent reading of a Chinese word produced a homophone priming effect (of 80ms). As a baseline, Experiment 9a showed that when participants saw written Chinese words and read them aloud (i.e., Chinese words were both seen and spoken), this also produced a homophone priming effect (of 34ms). This finding replicates the results of Experiment 3 (reported in Chapter 3). The fact that the priming observed in Experiment 9b was not smaller than in Experiment 9a (and indeed was actually larger) shows that the phonological activation of non-selected words persists to affect subsequent picture naming times. Further, the fact that homophone priming of object naming was found when the word prime was read aloud (in Experiment 3) and both when it was read only silently (in Experiment 4) and when a phonologically different word was produced (in the translation experiment of Experiment 9a) shows that the locus of the homophone priming effect is highly unlikely to be solely articulatory.

Experiment 10a showed that reading English words aloud produced a homophone priming effect on object naming times in Chinese; this rather small effect, of 16ms, was significant by participants but not by items. (Items showed a rather high degree of variability in these experiments, probably due to the relatively small number used that satisfied the stringent selection constraints for use with bilinguals.) Finally, Experiment 10b found that translating an English word into Chinese also produced a homophone priming effect (of 21ms). In Experiment 10a, the internally activated Chinese translation-equivalents were neither seen nor spoken, and in Experiment10b the prime words were not seen but were spoken.

The combination of these results suggests that, in the prime task of each trial, the phonological representations of Chinese words are activated, and that this irrespective (a) of whether the word is seen or not, (b) of whether it is spoken or not, or (c) of the task performed on the prime trial. These results are consistent with a cascaded processing model of word production in Chinese in which semantically co-activated lexical representations are phonologically encoded (and that this phonological activation persists to affect subsequent object naming). Further, the results suggest that bilinguals always activate the lexical representation of words in both their languages.

This conclusion appears to be at odds with those studies that found no effects of phonological activation of context pictures in word translation and word association tasks in Chinese speakers (Q. Zhang & Zhu, 2016). This discrepancy may be explained in terms of the potentially important differences between the tasks used (and perhaps other aspects of methodology). The sequential prime word to object naming paradigm used in the experiments here (where the prime word appeared about 1.5 seconds before the target object) detects persisting activations from prime to probe task, which are separated in time and where *both* tasks require lexical selection and word production. While Q. Zhang and Zhu (2016) used word translation task with picture superimposed task, where the co-activated target and distractor representations can produce conflicting activation patterns; this will increase the difficulty of resolving competition in order to achieve a *single* lexical selection.

3.2 Word processing times in the various prime tasks.

The prime task in Experiment 9a required participants to read aloud visually presented Chinese words, which was the same task as used in Experiments 1 and 2, and the prime task in Experiment 3. Experiment 9a found that reading times to high frequency words were faster than to low frequency words, which was also observed in the earlier three experiments. Reading times to homophones were only slightly faster than to control words, and the difference between homophones and controls tended to be slightly larger for low- than for high-frequency words; however, these results were not consistently reliable in the analyses both by participants and by items. The results of Experiments 1 and 2, where the homophones and non-homophone controls were matched on frequency and a wide range of other factors, found no main effect of frequency or any interaction between word type and frequency. Although less clear, the results of Experiment 9a show no convincing evidence for any effect of frequency inheritance on homophone reading times. The overall conclusion of all these results supports the hypothesis that homophones have independent lexical representations

The times taken to translate Chinese into English words in Experiment 9b are faster for high than low frequency words, but there are no differences in translating homophones and control words. This suggests that the process of translation operates at the semantic (and also potentially at lexical) levels, but not at the phonological level.

The times taken to read aloud English words by the Chinese-English bilingual participants showed no effects of either frequency or word type. It is noteworthy that the overall reading times for English words were much slower than for Chinese words, which probably reflects the facts that English is the second language of the bilinguals and that reading English is a hard and less practised activity.

The times taken to translate English into Chinese words in Experiment 10b showed a frequency effect, but no effect of word type: these bilinguals do not produce homophones and faster than control words. That is, there is no evidence that the spoken production of homophones enjoys any frequency inheritance effect from their homophone twins or families. This result is contrary to the results of Jescheniak and Levelt (1994) and Jescheniak, Meyer, and Levelt (2003), where the homophones in their experiments were all homographic due to the consistent spelling-sound correspondences of the orthographies of the languages of the bilinguals studied. In Chinese, all the homophones were heterographic (and were visually quite different). It is interesting that Caramazza et al. (2001) failed to replicate Jescheniak and colleagues' results with both English-Spanish and Chinese-English bilinguals. Overall, the results from the translation times in Experiment 10b, like those of Experiments 1, 2, 3 and 9a, support the hypothesis that homophones have independent lexical representations.

CHAPTER 6 OVERVIEW AND CONCLUSIONS

Spoken word production in Chinese has attracted less cognitive psychological research than speech in European languages, and there has not been very much work on the production of homophones which are very prevalent in Chinese (and are rare in languages with alphabets that have consistent correspondences between spellings and sounds, unlike English). The experimental work reported in this thesis explored the time to produce spoken words in Chinese, and was designed to address two major questions for theoretical accounts of spoken word production: (1) how homophones are represented in the speech production system, and (2) how activation is transmitted from lexical to phonological levels in spoken Chinese.

1. The Lexical Representation of Homophones

Homophones are two or more words with different meanings but whose pronunciations are identical. Homophones can be homographic, having the same spelling (e.g., the *palm* of the hand vs. the *palm* tree, and the *nurse* vs. to *nurse*), or heterographic with different spellings or written forms (e.g., *hair* vs. *hare*, and *rain* vs. *rein* vs. *reign*). In Chinese, there are very, very many heterographic homophones, and there can be many different characters that have the same pronunciation. For example, the Cihai dictionary of Standard Mandarin Chinese lists 149 characters representing the syllable “yì”. Some

Chinese characters may be pronounced with the same syllable but not the same tone, and so will have different meanings. In the work reported in this thesis, *homophones* will refer to heterographic homophones that have both the same syllable (or syllables) and the same tone.

An early proposal (e.g., Jescheniak & Levelt, 1994) was that homophones share the same, common lexical phonological representation for spoken word production. For example, *rain*, *rein* and *reign* would share the same phonological representation /rain/. An alternative proposal (e.g., Caramazza, 1997) was that homophones, just like all words, have their own word-specific, independent lexical phonological representations (even if the content of these will be identical for homophones). Critical evidence for the arbitration between the shared and independent representation hypotheses concerns whether there are "frequency-inheritance" effects in spoken word production tasks.

In support of the shared representation hypothesis, Jescheniak and Levelt (1994) found that the times taken by bilinguals to produce words in a translation task were affected by the *cumulative* frequency of homophones (i.e., the sum of the frequencies of *rain*, *rein*, and *reign*), and not by the *specific* frequency of the particular word produced. Low frequency words (e.g., "rein") were named faster than non-homophones matched on word-specific frequency; they appeared to "inherit" the advantage from their higher-frequency homophone mates ("rain"). These results support the view that homophones share the same lexical representation. In support of the independent representation hypothesis, Caramazza, Costa, Miozzo, and Bi (2001) and others found that the times to name objects with homophone names (e.g., *nun*) were not any faster than those with non-homophone names matched on word-specific frequency; they did not "inherit" any advantage from their higher-frequency homophone mates (*none*). Further, the lower-frequency names were produced slower than non-homophone names matched on homophone cumulative frequency. These results support the view that homophones have independent lexical representations. It would, therefore, appear that the Jescheniak and Levelt results are exceptional in the literature, although the reasons for this remain unclear. A possible line of inquiry for future research to illuminate this issue would be to examine the role played by the level of proficiency with the orthographic-to-phonological consistency of the bilingual's second language.

Another source of evidence relevant to the representation of homophones comes from analyses of the articulation durations of spoken words. In analyses of conversational speech, Gahl (2008) found that the durations of higher-frequency members of homophone pairs (e.g., "time") were shorter than their lower-frequency counterparts (e.g., "thyme"). Bell et al. (2009) found shorter durations for high- than low-frequency words, and it remains unclear whether Gahl's results are due solely to homophony.

Experiments 1 and 2 attempted to arbitrate between the shared and independent representation hypotheses of homophones in Chinese. The experiments analysed both the latencies to read aloud written two-character Chinese homophones and matched non-homophone words, and the durations (and intensities) of the reading responses. The experiments compared two types of homophones: homophone *twins*, where there are only two words sharing the same pronunciation, and homophone *families*, that contain three or more identically pronounced words.

The latency results from reading both homophone twins and families showed very similar results. Homophones were not named faster than non-homophones for either high-and, crucially, low-frequency words. Frequency had clear effects on reading all words, but the difference between homophone and control words was the same for both high- and low-frequency words. Thus, lower-frequency homophones did not inherit any processing advantage from their higher-frequency homophones. This result was confirmed by mixed-effects analyses that included a measure of cumulative homophone frequency. The results support the theory that all Chinese homophones have independent representations in the speech production system.

The articulation duration results showed similar trends as the reading latency data but were not consistently reliable across both experiments. The results do not support those reported by Gahl (2008). The results from the intensities of reading responses were also not consistent across experiments, and the trend was for non-homophones to be pronounced somewhat louder than homophones, for both high- and low-frequency words. The duration and intensity results offer no convincing support for the view that homophones have a shared representation. It is possible that there are differences in how English and Chinese homophones are articulated. At present, it is unclear whether the articulation-level features of duration and intensity reflect (i) “output” processes in the execution and expression of speech (e.g., for prosody and for emphasis) when producing words within sentences in conversational contexts, or (ii) “central” features of lexical phonological representations as detected in the production of single word reading responses.

The results of Experiments 1 and 2 support the idea that there are independent representations of homophones in the speech production system in Chinese. This conclusion is based on assumptions concerning both the use of word frequency used to index ease of lexical processing and the use of the task of reading aloud. First, the effects of word frequency have been commonly observed in very many tasks, including visual word recognition and spoken word production. The interpretation of the results of Experiments 1 and 2 assumes that the critical variable that determines the ease of retrieval of lexical phonological representations in Chinese is word frequency. However, some studies have claimed that a word's age-of-acquisition (AoA) rather than frequency is the major determinant of lexical retrieval in speech (e.g., Barry et al., 2001; Bonin, Barry, Méot, & Chalard, 2004). Measures of AoA, frequency and imageability correlate quite highly in large samples of words (e.g., Gilhooly & Logie, 1980), and it is difficult to construct factorial sets of words that distinguish the effects of these variables. A potential criticism of the results reported in the thesis is that they are compromised by the failure to match homophones and non-homophones for AoA or, indeed, for other potentially important variables. Frequency and AoA are quite highly correlated in Chinese. The homophones and non-homophones were matched on many lexically relevant factors, including familiarity, number of meanings, and concreteness; and both sets of words showed a strong and reliable effect of word frequency. There are very many homophones in Chinese, and so they not an unusual set of words. It is unlikely that there are systematic differences between homophones and non-homophone controls on AoA.

Second, it may be possible that reading aloud Chinese words do not engage the stored lexical phonological representations involved in speech production. Although such a possibility may be plausibly the case for reading English words, where sub-lexical phonological recoding would activate phoneme-level information during reading, it is much less likely to occur in reading Chinese words. There is very little evidence that skilled adult Chinese readers rely upon the inconsistent phonological correspondences of sub-character segments (such as phonetic radicals). Reading-aloud Chinese words involve semantic and lexical processes, just like self-generated speech. The evidence presented in Chapter 2 shows that homophones have independent lexical phonological representations.

2. The flow of Activation From Lexical to Phonological Levels in Spoken Word Production

An important difference between theories of spoken word production is their conception of how activation is transmitted between the processing stages that operate upon different types of representations involved in speech production. All models make a distinction between semantic, lexical, and phonological levels of representation and processing (and some additionally propose syntactic and morphological levels).

The highly influential and well specified theoretical framework advanced by Levelt et al. (1999) proposed that the cognitive processing responsible for speech is *serial* and *discrete*, where one processing level must be completed before the next can begin. This claims that only the selected lexical representation of the target is phonologically encoded. In contrast, *cascaded* processing models of speech production (e.g., Humphreys et al., 1988; Morsella & Miozzo, 2002) propose that once processing begins at one level, activation flows continuously —and in cascade— to the next level, and so multiple representations will be activated to various degrees. This approach proposes that all activated lexical representations are phonologically encoded to some extent. *Interactive* models (e.g., Dell, 1986) additionally propose that there is bidirectional transmission of activation, and in particular that activation at the stage of phonological encoding can feedback to influence lexical selection.

Critical evidence for the arbitration between these models comes from studies of the facilitation and interference effects from distractors presented along with a stimulus designed to elicit a target response (as in the picture-word task) and from the study of various semantic priming effects from one trial to another, as in studies using the semantic blocking procedure (e.g., Damian & Als, 2005) and from studies finding cumulative semantic interference effects (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010), where naming times increased linearly as a function of the number of previously named pictures in the same category (such as fruits, animals, vehicles, etc.).

In this research, interest has focused on the effects that non-target items, related to the target at different levels of representation, have on the time to produce the target word. Non-target words are presented as explicit distractors in the picture-word and picture-picture tasks and are assumed to be generated internally (as mediators) in priming tasks. An example of a priming study that investigated the effects of semantically related words is the study by Wheeldon and Monsell (1994). On prime trials, participants were presented with definitions and were asked to say the defined name (e.g.*, the largest creature that swims in the sea* –> “whale”). On subsequent probe trials, participants were asked to name pictured objects (e.g., *SHARK* –> “shark”). Wheeldon and Monsell found that probe naming times were slower when the object was semantically related to the word produced on the prime trial. This was interpreted to result from competition within cascaded processing from the semantic to the lexical level, and specifically to the proposed inhibition of related non-target words. For example, when selecting the word “whale” to produce to a definition, semantically related words, such as 'shark', would also be activated, but these semantic competitors must be actively suppressed in order to select the target word "whale". When, on the following probe trial, 'shark' was the target, participants were slower to produce it because it had just been actively suppressed, and the effect of this inhibition persisted to the probe trial.

An important topic for study is whether (and how) words assumed to be internally co-activated affect the time taken to select, encode, and produce target words. Cascaded models propose that they would (by spreading activation and competition operating in the flow of activation between different levels), but serial discrete models propose that they would not. Interactive models further propose that co-activated presentations at one level may feedback activation to an earlier level.

A number of studies using the picture-word task have investigated the effects of distractor words that are phonologically related to a presumed semantically co-activated word. For example, when naming the pictured object *BOAT*, conceptual and semantic processing —and spreading activation at this level— will activate more than one lexical-level representation, such as 'boat', 'ship', 'oar', 'sail' etc., if to varying degrees. As activation builds up in these representations, it will cascade to activate the corresponding sub-word phonological-level representations of all these words. If these also feedback activation to influence lexical selection times, then mediated effects should be observed on target naming times. Semantic-to-phonological interference effects have been reported for mediated related distractors, such as SHIP+ *bow* (via 'boat') (e.g., Jescheniak & Schriefers, 1998). Semantic-to-phonological interference effects have also been reported for words phonologically related to translation-equivalents in Dutch-English bilinguals by Hermans et al. (1998) who called it the "phono-translation" interference effect. For example, the Dutch word *berm* when superimposed on a picture of MOUNTAIN (to name in English) slowed naming times, because *berm* is phonologically related to "berg" (the Dutch translation of the target "mountain") and so increased the status of *berg* as a lexical competitor to the target. This effect was replicated in balanced Spanish-Catalan bilinguals naming in Catalan studied by Costa, Colomé, Gómez, and Sebastián-Gallés (2003), although they attributed it to cross-language influence at the post-lexical level of phonological encoding of a language-specific selected naming response.

3. Word-To-Picture Priming in Chinese Word Production

The experiments reported in chapters 3, 4, and 5 used a priming procedure that manipulated the type of relationship between a prime word and a probe object to-be-named. As this is a sequential prime-to-probe paradigm, it involves two tasks, each requiring lexical activation, selection, phonological encoding, and the production of a word. (The only exception was Experiment 4, where the prime task was the silent reading of a prime word, where there was lexical activation, but no word production.) The two tasks were temporally sequential and discrete, as the first had to be performed before the second (and so there were two serial and discrete sets of lexical selections in a row). However, it is generally assumed that changes in activation levels during the processing of the prime word will persist to affect the naming of the probe object. Changes of patterns of activation within the lexical processing system do not return (or are reset) to baseline once a response is made to the prime word; the slate is not wiped clean, and the processing of the prime word will leave persisting activation to prime the naming of picture objects.

A variety of types of prime-to-target relationships were manipulated in the experimental work reported in this thesis. The purpose was to determine whether these produce priming of object naming latencies, and to assess the relative magnitudes of their effects in order to explore the extent of cascaded or interactive processing in spoken word production in Chinese. A range of both direct and indirect, mediated relationships were studied. These are listed below (along with English examples).

(1) Direct priming relationships involved only a 'one-step' prime-to-probe connection between the presented prime word and the target name of the object. The direct relationships studied were: (a) *Repetition* (or identity), where the prime word was the name of the target (e.g., *oar*—OAR); (b) *Semantic*, where the meanings of the prime word and object were related or associated (e.g., *boat*—OAR); (c) *Homophone*, where the prime word was a homophone of the object's name (e.g., *awe*—OAR); and (d) *Phonological*, where the prime word was phonologically related only in terms of segmental, atonal syllables (and so has the same phonemes but not the same tone) (e.g., *orb*—OAR; in Chinese /shu1/— /shu3/).

The interpretation of priming effects from these direct relationships, which have been widely demonstrated by previous research, is that they result from persisting activations at semantic, lexical or sub-lexical representations that facilitate their subsequent retrieval when required during the process of object naming.

(2) Indirect or *mediated* relationships involve a 'two-step' prime-to-probe connection between the presented prime word and the object to name; it is a prime-to-mediator-to-target relationship. The mediator would be a word, internally activated by the prime, that will pre-activate representational levels that are also contacted during the process of naming the target object. This mediated activation may produce facilitation if it operates at levels common to naming (e.g., homophone prime words pre-activating the same syllable motor patterns to be used in producing the object's target name), or may produce interference if it operates at levels that conflict with naming (e.g., by increasing the activation of a lexical competitor that slows the selection of the object's target name).

The mediated relationships studied were: (a) *Homophone-to-semantic* (or H2S), where the prime word was a homophone of a word semantically related to the target object (e.g., *thyme*—[mediator=time]—CLOCK). If this condition produced a priming effect, it would suggest that the prime word activated the lexical representations of its homophones that then feedback to, and interacted with, the semantic system. (b) *Phonological-to-semantic* (or P2S), where the prime word shared an atonal syllable with the target object name (e.g., *lime*—[mediator=time]—CLOCK). If this condition produced a priming effect, it would suggest that the prime word activated sub-lexical phonological representations that, via interactive processing, feedback to lexical and semantic levels. (c) *Semantic-to-homophone* (or S2H), where the prime word was semantically related to a homophone of the target object name (e.g., *either*—[mediator=or]—OAR). If this condition produced a priming effect, it would suggest that the prime word would activate words semantically related to it and that these would then activate their homophones. (This type of mediated relationship is, perhaps, one of the most "mediated" of all.)

(3) Cross-language mediated relationships in Chinese-English bilingual speakers also involve a 'two-step' prime-to-probe connection between a presented prime word and the object to name, via the prime word's translation-equivalent; it is a prime-to-translation-to-target mediated relationship. The experiments reported in Chapter 5 examined possible translation mediated homophone priming of object naming times in Chinese.

Results of the priming experiments.

Direct priming. The results of the experiments produced a generally consistent pattern of results. A direct repetition priming effect was observed in Experiments 3 and 4 (in Chapter 3) and in Experiments 5, 6, 7, and 8 (in Chapter 4). A direct homophone priming effect was observed in Experiments 3 and 4, in Experiments 7 and 8, and in Experiment 9a (in Chapter 5). A direct phonological priming effect, from prime words with the same syllable but not the tone as the target name, was observed in Experiment 3 and 4, and in Experiments 7 and 8. A direct semantic priming effect (of 25ms) was observed in Experiment 6 for silent reading, but the effect (of 17ms) did not reach significance in Experiment 5 where the prime words were read aloud. These results show quite convincingly that the priming paradigm used in this work is most certainly sufficiently powerful to detect direct priming effects on probe object naming times. Whether the prime word was read aloud or read silently had very little difference on the magnitudes of the priming effects, apart from the semantic priming effect, which was the one with the smallest magnitude (and which may have the fastest rate of decay).

These direct priming effects show an orderly pattern and can be interpreted in terms of persisting activation from word reading operating at different levels of representation within the lexical system. Repetition priming reflects persisting activation at semantic, lexical, and sub-lexical phonological levels. Homophone priming reflects persisting activation at lexical and sub-lexical phonological levels. Both repetition and homophone priming will activate the syllable motor programs (that include tonal information) in the models of Chinese phonological encoding proposed by Roelofs (2015) and Zhang, Zhu, and Damian (2018). Phonological priming will activate atonal syllables but not the syllable motor programs. The order of the magnitudes of direct priming effects may be interpreted in terms of both discrete and cascaded models by persisting activation operating selectively at different representational levels: the more levels are activated, then the larger the priming will be.

Mediated priming. The research presented here detected no reliable mediated priming effects for either homophone-to-semantic or phonological-to-semantic priming in Experiments 5 and 6, or for semantic-to-homophone priming in Experiments 7 and 8. In combination, these results offer no support at all for interactive processing (and arguably only limited support for strong cascaded processing) in the lexical to phonological encoding stages underlying Chinese spoken word production.

Translation-equivalent mediated priming. The results of Experiments 9b and 10b show that both translating written Chinese words into English (in 9b) and translating printed English words into Chinese (in 10b) produce reliable mediated homophone priming of object naming. These results suggest that Chinese-English bilinguals will activate the phonological word forms of Chinese word both when they see a Chinese prime word but do not have to say it (as they must produce an English word) and when they do not see a Chinese prime word (as they see an English word) but do say aloud the prime word.

3. General Conclusions

The results from both word reading latencies and from translating English words into Chinese support the hypothesis that homophones have independent lexical phonological representations. It remains unclear whether the articulation durations (or the intensity) of reading responses reflect features of the independent lexical representations within the speech production system (as opposed to post-lexical prosodic processes). It would appear that the "proximate units" of Chinese phonological encoding (i.e., the primary selectable sub-lexical unit) are atonal syllables, although these must be integrated into a tonal frame for Chinese speech.

The patterns of the priming effects observed were broadly the same when participants read words aloud compared to reading them silently. This shows that the act of explicit word production is not necessary for the detection of homophone and phonological priming and suggests that, in silent reading, processes of phonological encoding are activated automatically.

The absence of mediated priming effects in the experiments reported here offers no support for the idea that Chinese spoken word production operates by an interactive model. Although there may be cascaded processing from the semantic to the lexical levels, it would appear that processing from the lexical to sub-lexical phonological levels operates in a serial and discrete fashion in Chinese. However, future work will be required to strengthen this conclusion. In particular, it is possible that mediated effects may have shorter effective time courses such that they would not be detected by the priming paradigm employed in the experimental work reported in this thesis. This could be usefully explored by investigating direct and mediated relationships in comparisons between the priming paradigm and presenting the words used as primes here (and the mediators) as distractors in the picture-word task, and to study possible differences in their time course (by manipulations of the stimulus onset asynchrony of the target object and distractor word and of the inter-stimulus intervals in the priming paradigm).

Language universal or language-specific processing?

An important general theoretical question for language production is whether the cognitive processes that underlie and allow the production of words in speech are *universal* (i.e., obey the same basic computational principles irrespective of the content of the representations of the specific language used) or are *language-specific* (i.e., that different cognitive processes are applied as determined by the features of, and linguistic constraints imposed by, specific languages). Languages differ in a variety of ways, for example in terms of their orthography, phonology, vocabulary, and grammar. It is clear that there will be constraints on particular aspects of cognitive processes imposed by these various linguistic features of a language. For example, the role of phonological recoding when reading aloud of printed words printed words will obviously be different in alphabetic orthographies (and especially those with highly consistent letter-to-sound relationships, as in Italian, Turkish, and Welsh) than in "logographic" orthographies (such as the characters of Chinese). Thus, the particular process of visual word recognition may be language-specific. Further, there must be language-specific constraints upon grammatical processes, as some languages have particular grammatical features that are not present in other languages (e.g., grammatical gender). Thus, some languages will involve particular processes (e.g., to ensure 'gender agreement' when adding articles and adjectives to nouns) that other languages simply do not require.

Concerning word production in speech, it is highly likely that the process of semantic-to-lexical activation operates in cascade and does so universally for all languages. It is also highly likely that, for proficient bilinguals, lexical representations in both languages are activated in parallel. Concerning the lexical representation of phonological forms, it might be expected that the pervasive existence of homophones in Chinese would exploit the advantage of economy of storage by having shared lexical representations of homophones. However, the results of the first experiments reported in this thesis show that in Chinese, as has also been found for English, there are independent lexical phonological representations for homophones; therefore this also appears to be a universal feature of speech production.

All spoken languages have lexical and sub-lexical phonological forms, and so a critical question is whether the functional organisation of these levels is universal or language-specific. This important question cannot yet be answered definitively, but a potentially interesting speculation is that there is cascaded lexical-to-phonological processing in European languages but serial and discrete lexical-to-phonological processing in Chinese. There is some evidence to support this speculation from the study by Zhang, Zhu, and Damian (2018) described earlier. They combined the use of the semantic blocking procedure with the picture-word task in which the distractor words were phonologically related to a semantic competitor of the target object names (e.g., ARM+*note*, which is phonologically similar to the name of the same category member 'nose'). For English-speaking participants, such "mediated" distractor words slowed object naming, but only when presented in semantically homogeneous blocks, which indicates "limited" cascaded processing. For Chinese-speaking participants, such phonological-to-semantic mediated distractors had no effect for either homogeneous or heterogeneous blocks, which indicates that in Chinese there is serial and discrete processing.

Whether this apparent language-specific processing difference is due to the tonal nature of Chinese phonology or its essentially syllable-based "proximate" phonological units (compared to phoneme-based units in English) is not clear, and so studies of speakers of other tonal languages (such as Thai) would be required to test this further.

The experiment planned to explore Zhang et al.'s work will use the picture-word task (presented in both semantically homogeneous and heterogeneous cyclic blocks of trials) to compare distractors that are homophones of same-category members in both English (e.g., APPLE+*pair*, HORSE+*dear*) and Chinese. As homophones have complete overlap with another member of the semantic cohort, in Chinese they will have both the same atonal syllable and the same tone in common, and so might be expected to show mediated interference effect. Such experiments will be necessary in order to arrive at a satisfactory answer to the question of whether cascaded or serial processing is a universal feature or whether there are language-specific constraints on spoken word production.

REFERENCES

Abdel Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin & Review, 15*(6), 1055-1063. doi:10.3758/PBR.15.6.1055

Alario, F. X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition, 102*(3), 464-475. doi:10.1016/j.cognition.2006.02.002

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. J. B. R. M. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445-459. doi:10.3758/bf03193014

Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of Acquisition, Word Frequency, and the Locus of Repetition Priming of Picture Naming. *Journal of Memory and Language, 44*(3), 350-375. doi:https://doi.org/10.1006/jmla.2000.2743

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cogn Psychol, 6*(3), 325-356. doi:https://doi.org/10.1016/0010-0285(74)90016-4

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology, 58*(4), 667-692.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*(1), 92-111. doi:http://dx.doi.org/10.1016/j.jml.2008.06.003

Bi, Y., Xu, Y., & Caramazza, A. (2009). Orthographic and phonological effects in the picture–word interference paradigm: Evidence from a logographic language. *Applied Psycholinguistics, 30*(4), 637-658. doi:10.1017/S0142716409990051

Biedermann, B., Blanken, G., & Nickels, L. (2002). The representation of homophones: Evidence from remediation. *Aphasiology, 16*(10-11), 1115-1136. doi:10.1080/02687030244000545

Biedermann, B., Coltheart, M., Nickels, L., & Saunders, S. (2009). Effects of homophony on reading aloud: Implications for models of speech production. *Language and Cognitive Processes, 24*(6), 804-842. doi:10.1080/01690960802597243

Biedermann, B., & Nickels, L. (2008a). Homographic and heterographic homophones in speech production: Does orthography matter? *Cortex, 44*(6), 683-697. doi:https://doi.org/10.1016/j.cortex.2006.12.001

Biedermann, B., & Nickels, L. (2008b). The representation of homophones: More evidence from the remediation of anomia. *Cortex, 44*(3), 276-293. doi:https://doi.org/10.1016/j.cortex.2006.07.004

Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition, 42*(1), 23-60. doi:https://doi.org/10.1016/0010-0277(92)90039-K

Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive neuropsychology, 22*(6), 695-717.

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language, 48*(3), 468-488. doi:https://doi.org/10.1016/S0749-596X(02)00503-X

Bloem, I., van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language, 51*(2), 307-323. doi:https://doi.org/10.1016/j.jml.2004.05.001

Boersma, P. P. G. (2002). Praat, a system for doing phonetics by computer. *Glot international, 5*.

Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language, 50*(4), 456-476.

Bonin, P., & Fayol, M. (2002). Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology, 14*(3), 289-313. doi:10.1080/09541440143000078

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

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS ONE, 5*(6), e10729. doi:10.1371/journal.pone.0010729

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive neuropsychology, 14*(1), 177-208.

Caramazza, A., Bi, Y., Costa, A., & Miozzo, M. (2004). What Determines the Speed of Lexical Access: Homophone or Specific-Word Frequency? A Reply to Jescheniak et al. (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 278-282. doi:10.1037/0278-7393.30.1.278

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1430-1450. doi:10.1037/0278-7393.27.6.1430

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex, 26*(1), 95-122.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *nature, 349*, 788. doi:10.1038/349788a0

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon. *Cognition, 64*(3), 309-343.

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science, 8*(4), 322-325.

Charlton, C., Rasbash, J., Browne, W. J., Healy, M., & Cameron, B. (2017). MLwiN Version 3.00. *Bristol: Centre for Multilevel, Modelling University of Bristol*.

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. doi:10.1080/01690960902804515

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751-781. doi:10.1006/jmla.2001.2825

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. In *Laboratory Phonology* (Vol. 6, pp. 243).

Coles, M. G. H. (1989). Modern mind - brain reading: psychophysiology, physiology, and cognition. *Psychophysiology, 26*(3), 251-269.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45*(4), 721-736. doi:10.1006/jmla.2001.2793

Colomé, À., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology Learning Memory and Cognition, 36*(1), 96-109. doi:10.1037/a0017677

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological review, 100*(4), 589.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*: Lawrence Erlbaum Associates.

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.

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of experimental psychology. Learning, memory, and cognition, 26*(5), 1283-1296. doi:10.1037//0278-7393.26.5.1283

Costa, A., Colomé, À., Gómez, O., & Sebastián-Gallés, N. (2003). Another look at cross-language competition in bilingual speech production: Lexical and phonological factors. *Bilingualism: Language and Cognition, 6*(3), 167-179.

Cuetos, F., Bonin, P., Alameda, J. R., & Caramazza, A. (2010). The specific-word frequency effect in speech production: Evidence from Spanish and French. *The Quarterly Journal of Experimental Psychology, 63*(4), 750-771. doi:10.1080/17470210903121663

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 416-431. doi:10.1037/0278-7393.29.3.416

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1372.

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language, 49*(1), 119-132. doi:10.1016/S0749-596x(03)00008-1

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

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 345.

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition, 81*(3), B77-B86. doi:https://doi.org/10.1016/S0010-0277(01)00135-4

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP Evidence for Ultra-Fast Semantic Processing in the Picture–Word Interference Paradigm. *Frontiers in Psychology, 1*, 177. doi:10.3389/fpsyg.2010.00177

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review, 93*(3), 283.

Dell, G. S. (1990). Effects of Frequency and Vocabulary Type on Phonological Speech Errors. *Language and Cognitive Processes, 5*(4), 313-349. doi:10.1080/01690969008407066

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Memory and Language, 20*(6), 611.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review, 104*(4), 801-838. doi:10.1037/0033-295X.104.4.801

Dérouesné, J., & Beauvois, M.-F. (1979). Phonological processing in reading: data from alexia. *Journal of Neurology, Neurosurgery & Psychiatry, 42*(12), 1125-1132.

Dong, Z., & Dong, Q. (2003, 26-29 Oct. 2003). *HowNet - a hybrid language and knowledge resource.* Paper presented at the International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003.

Durso, F. T., & Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory, 5*(5), 449.

Durso, F. T., & Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory*

Memory & Cognition, 5(5), 449.

Dylman, A. S., & Barry, C. (2018). When having two names facilitates lexical selection: Similar results in the picture-word task from translation distractors in bilinguals and synonym distractors in monolinguals. *Cognition, 171*, 151-171. doi:https://doi.org/10.1016/j.cognition.2017.09.014

Edwards, J., Pexman, P., & Hudson, C. (2004). Exploring the dynamics of the visual word recognition system: Homophone effects in LDT and'naming. *Language and Cognitive Processes, 19*(4), 503-532. doi:10.1080/01690960344000215

Feng, S. (1995). Prosodic structure and prosodically constrained syntax in Chinese. *Dissertations available from ProQuest*.

Feng, S. (2001). The multidimensional properties of "word" in Chinese. *Contemporary Linguistics, 3*(3), 161-174.

Ferrand, L., Humphreys, G. W., & Segui, J. (1998). Masked repetition and phonological priming in picture naming. *Perception & Psychophysics, 60*(2), 263-274.

Fidelholtz, J. L. (1975). *Word frequency and vowel reduction in English.* Paper presented at the Chicago Linguistic Society.

Fowler, C. A. (1988). Differential Shortening of Repeated Content Words Produced in Various Communicative Contexts. *Language and Speech, 31*(4), 307-319. doi:10.1177/002383098803100401

Fry, D. B. (1954). Duration and Intensity as Physical Correlates of Linguistic Stress. *26*(1), 138-138. doi:10.1121/1.1917773

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology, 74*(2), 159-180. doi:10.1111/j.2044-8295.1983.tb01851.x

Gagnon, D. A., Schwartz, M. F., Martin, N., Dell, G. S., & Saffran, E. M. (1997). The Origins of Formal Paraphasias in Aphasics' Picture Naming. *Brain and Language, 59*(3), 450-472. doi:https://doi.org/10.1006/brln.1997.1792

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*(3), 474-496.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*(4), 789-806.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395-427.

Glaser, W. R. (1992). Picture naming. *Cognition, 42*(1), 61-105. doi:https://doi.org/10.1016/0010-0277(92)90040-O

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes, 21*(7-8), 817-855. doi:10.1080/01690960600824112

Gomez, P., Shutter, J., & Rouder, J. N. (2008). Memory for objects in canonical and noncanonical viewpoints. *Psychonomic Bulletin & Review, 15*(5), 940-944.

Harley, T. A. (1999). Will one stage and no feedback suffice in lexicalization? *Behavioral and Brain Sciences, 22*(1), 45-45. doi:10.1017/S0140525X99291777

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). *Journal of Memory and Language, 52*(1), 58-70. doi:https://doi.org/10.1016/j.jml.2004.07.006

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition, 1*(3), 213-229. doi:10.1017/S1366728998000364

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. *Current progress in historical linguistics, 96105*.

Howard, D., & Orchard-lisle, V. (1984). On the origin of semantic errors in naming: Evidence from the case of a global aphasic. *Cognitive neuropsychology, 1*(2), 163-190. doi:10.1080/02643298408252021

Hsiao, J. H.-w., & Shillcock, R. (2006). Analysis of a Chinese Phonetic Compound Database: Implications for Orthographic Processing. *Journal of Psycholinguistic Research, 35*(5), 405-426. doi:10.1007/s10936-006-9022-y

Humphreys, G. W., & Bruce, V. (1989). *Visual cognition: Computational, experimental and neuropsychological perspectives*: Psychology Press.

Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *2*(255). doi:10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144. doi:https://doi.org/10.1016/j.cognition.2002.06.001

Jacobs, M., Singer, N., & Miozzo, M. (2004). The representation of homophones: Evidence from anomia. *Cognitive neuropsychology, 21*(8), 840-866. doi:10.1080/02643290342000573

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes, 23*(7-8), 1191-1223. doi:10.1080/01690960802250900

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29*(6), 687-715. doi:https://doi.org/10.1016/0749-596X(90)90044-Z

Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: evidence from event-related brain potentials. *Cognitive Brain Research, 15*(3), 261-276.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 824-843.

Jescheniak, J. D., Meyer, A. S., & Levelt, W. J. M. (2003). Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 432-438.

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1256-1274. doi:10.1037/0278-7393.24.5.1256

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language, 65*(4), 406-430. doi:https://doi.org/10.1016/j.jml.2011.07.003

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones*: na.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *J Acoust Soc Am, 118*(2), 1038-1054.

Kuipers, J.-R., & Heij, W. L. (2009). The limitations of cascading in the speech production system. *Language and Cognitive Processes, 24*(1), 120-135.

Law, S.-P., Weekes, B. S., Wong, W., & Chiu, K. (2009). Reading aloud pseudo-characters by individuals with acquired dyslexia: Evidence for lexically mediated processes in reading Chinese. *Language and Cognitive Processes, 24*(7-8), 983-1008. doi:10.1080/01690960802193696

Lee, C.-Y., Tsai, J.-L., Su, E. C.-I., Tzeng, O. J., & Hung, D. L. (2005). Consistency, regularity, and frequency effects in naming Chinese characters. *Language and Linguistics, 6*, 75-107.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. ACL. In: MIT Press Series in Natural-Language Processing. MIT Press, Cambridge, Massachusetts.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38. doi:undefined

Levelt, W. J. M., Schreuder, R., & Hoenkamp, E. (1978). Structure and use of verbs of motion. In *Recent advances in the psychology of language* (pp. 137-162): Springer.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological review, 98*(1), 122.

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods, 39*(2), 192-198. doi:10.3758/bf03193147

Lohmann, A. (2017). Cut (n) and cut (v) are not homophones: Lemma frequency affects the duration of noun–verb conversion pairs 1. *Journal of Linguistics*, 1-25.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing, 19*(1), 1.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* (pp. 122-147). Cambridge, MA, US: The MIT Press.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69-69.

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35*(3), 494-503. doi:10.3758/BF03193289

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General, 132*(2), 228-252. doi:10.1037/0096-3445.132.2.228

Miozzo, M., & Caramazza, A. (2005). The Representation of Homophones: Evidence From the Distractor-Frequency Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1360-1371. doi:10.1037/0278-7393.31.6.1360

Mitchell, D. B., & Brown, A. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, Cognition, 14*(2), 213.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 555-563. doi:10.1037/0278-7393.28.3.555

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*(3), 359-377. doi:https://doi.org/10.1016/j.jml.2005.05.001

Navarrete, E., & Costa, A. (2009). The Distractor Picture Paradox in Speech Production: Evidence from the Word Translation Task. *Journal of Psycholinguistic Research, 38*(6), 527. doi:10.1007/s10936-009-9119-1

Newcombe, F., & Marshall, J. C. (1980). Transcoding and lexical stabilization in deep dyslexia. *Deep dyslexia*, 176-188.

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282-302. doi:https://doi.org/10.1016/j.cognition.2010.01.001

Oppermann, F., Jescheniak, J. D., Schriefers, H., & Görges, F. (2010). Semantic relatedness among objects promotes the activation of multiple phonological codes during object naming. *The Quarterly Journal of Experimental Psychology, 63*(2), 356-370.

Patterson, K., Coltheart, M., & Marshall, J. C. (1980). *Deep dyslexia*: Routledge & Kegan Paul.

Patterson, K., Marshall, J. C., & Coltheart, M. (1985). Surface dyslexia: Cognitive and neuropsychological studies of phonological reading. *London: LEA*.

Pease, D. M., & Goodglass, H. (1978). The Effects of Cuing on Picture Naming in Aphasia. *Cortex, 14*(2), 178-189. doi:https://doi.org/10.1016/S0010-9452(78)80043-4

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The Lexical Constituency Model: Some Implications of Research on Chinese for General Theories of Reading. *Psychological review, 112*(1), 43-59. doi:10.1037/0033-295X.112.1.43

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 539.

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and Visual Word Recognition: Can Feedback Explain Both Homophone and Polysemy Effects? *Canadian Journal of Experimental Psychology, 53*(4), 323.

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 139-156. doi:10.1037/0278-7393.27.1.139

Pexman, P. M., Lupker, S. J., & Reggin, L. D. (2002). Phonological effects in visual word recognition: investigating the impact of feedback activation. *Journal of experimental psychology. Learning, memory, and cognition, 28*(3), 572-584. doi:10.1037/0278-7393.28.3.572

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review, 107*(3), 460-499. doi:10.1037/0033-295X.107.3.460

Roelofs, A. (1997). A Case for Nondecomposition in Conceptually Driven Word Retrieval. *Journal of Psycholinguistic Research, 26*(1), 33-67. doi:10.1023/A:1025060104569

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological review, 110*(1), 88.

Roelofs, A. (2005). The visual-auditory color-word Stroop asymmetry and its time course. *Memory & Cognition, 33*(8), 1325-1336. doi:10.3758/BF03193365

Roelofs, A. (2006a). Context effects of pictures and words in naming objects, reading words, and generating simple phrases. *Quarterly Journal of Experimental Psychology, 59*(10), 1764-1784. doi:10.1080/17470210500416052

Roelofs, A. (2006b). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review, 13*(1), 33-37. doi:Doi 10.3758/Bf03193809

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye-movements. *Journal of Experimental Psychology A-Learning, Memory and Cognition, 34*, 353 - 368

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*(1), 22-37. doi:10.1111/jpr.12050

Roux, S., Bonin, P., & Kandel, S. (2014). The “BIG BIRD” of the “YELLOW YOUNG” man: Do nontarget properties cascade? *The Quarterly Journal of Experimental Psychology, 67*(4), 763-784.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10*(6), 645-657. doi:https://doi.org/10.1016/S0022-5371(71)80071-3

Sasanuma, S. (1980). Acquired dyslexia in Japanese clinical features and underlying mechanisms. *Deep Dyslexia: 2 nd Ed, 48*, 1987.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86-102. doi:https://doi.org/10.1016/0749-596X(90)90011-N

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*(1), 140-155.

Shu, H., Meng, X., Chen, X., Luan, H., & Cao, F. (2005). The subtypes of developmental dyslexia in Chinese: Evidence from three cases. *Dyslexia, 11*(4), 311-329.

Snodgrass, J. G., & McCullough, B. (1986). The role of visual similarity in picture categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(1), 147-154. doi:10.1037/0278-7393.12.1.147

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174-215. doi:10.1037/0278-7393.6.2.174

Starreveld, P. A., & La Heij, W. (1995). Semantic Interference, Orthographic Facilitation, and Their Interaction in Naming Tasks. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*(3), 686-698. doi:Doi 10.1037/0278-7393.21.3.686

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(4), 896-918. doi:10.1037/0278-7393.22.4.896

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276-315. doi:https://doi.org/10.1016/0001-6918(69)90055-9

Stone, G. O., Vanhoy, M., & Orden, G. C. V. (1997). Perception Is a Two-Way Street: Feedforward and Feedback Phonology in Visual Word Recognition. *Journal of Memory and Language, 36*(3), 337-359. doi:https://doi.org/10.1006/jmla.1996.2487

Surendran, D., & Levow, G.-A. (2004). *The functional load of tone in Mandarin is as high as that of vowels.* Paper presented at the Speech Prosody 2004, International Conference.

Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods, 46*(1), 263-273. doi:10.3758/s13428-013-0355-9

Sze, W. P., Yap, M. J., & Rickard Liow, S. J. (2015). The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis. *The Quarterly Journal of Experimental Psychology, 68*(8), 1541-1570. doi:10.1080/17470218.2014.985234

Taft, M. (1988). A morphological-decomposition model of lexical representation. *Linguistics, 26*(4), 657-668.

Taft, M. (2003). Morphological representation as a correlation between form and meaning. In *Reading complex words* (pp. 113-137): Springer.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 638-647.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior, 15*(6), 607-620.

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., Hoosain, R., & Siok, W. W. T. (1996). Activation of phonological codes before access to character meaning in written Chinese. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(4), 865-882.

Tsang, Y.-K., Huang, J., Lui, M., Xue, M., Chan, Y.-W. F., Wang, S., & Chen, H.-C. (2018). MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods, 50*(5), 1763-1777. doi:10.3758/s13428-017-0944-0

Uttl, B., Graf, P., & Santacruz, P. (2006). Object color affects identification and repetition priming. *Scandinavian Journal of Psychology, 47*(5), 313-325.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical Gender Is on the Tip of Italian Tongues. *Psychological Science, 8*(4), 314-317. doi:10.1111/j.1467-9280.1997.tb00444.x

Vitevitch, M. S. (2002). The Influence of Phonological Similarity Neighborhoods on Speech Production. *Journal of experimental psychology. Learning, memory, and cognition, 28*(4), 735-747. doi:10.1037//0278-7393.28.4.735

Vitkovitch, M. S., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology, 48*(4), 822-848.

Wang, L.-C., & Yang, H.-M. (2014). Classifying Chinese children with dyslexia by dual-route and triangle models of Chinese reading. *Research in developmental disabilities, 35*(11), 2702-2713.

Wang, W., Lu, A., He, D., Zhang, B., & Zhang, J. X. (2017). ERP Evidence for Chinese Compound Word Recognition: Does Morpheme Work all the Time? *NeuroQuantology; Vol 15, No 3 (2017)*.

Weekes, B., & Chen, H. Q. (1999). Surface dyslexia in Chinese. *Neurocase, 5*(2), 161-172.

Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *The Quarterly Journal of Experimental Psychology Section A, 44*(4), 723-761.

Wong, A. W.-K., & Chen, H.-C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1172-1190. doi:10.1037/a0013000

Xu, Y., Pollatsek, A., & Potter, M. C. (1999). The activation of phonology during silent Chinese word reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 838-857. doi:10.1037/0278-7393.25.4.838

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology, 97*(2), 259-268. doi:10.1348/000712605X70066

Yao, Y., & Sharma, B. (2017). What is in the neighborhood of a tonal syllable? Evidence from auditory lexical decision in Mandarin Chinese. *2017, 2*. doi:10.3765/plsa.v2i0.4090

Yin, W., & Butterworth, B. (1992). Deep and surface dyslexia in Chinese. *Advances in psychology, 90*, 349-366.

Yin, W. G., & Weekes, B. S. (2003). Dyslexia in Chinese: Clues from cognitive neuropsychology. *Annals of Dyslexia, 53*(1), 255-279.

Zevin, J. D., & Balota, D. A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 121.

Zhang, B., & Peng, D. (1992). Decomposed Storage in the Chinese Lexicon. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Advances in psychology* (Vol. 90, pp. 131-149): North-Holland.

Zhang, Q., Chen, H.-C., Stuart Weekes, B., & Yang, Y. (2009). Independent Effects of Orthographic and Phonological Facilitation on Spoken Word Production in Mandarin. *Language and Speech, 52*(1), 113-126. doi:10.1177/0023830908099885

Zhang, Q., & Weekes, B. S. (2009). Orthographic facilitation effects on spoken word production: Evidence from Chinese. *Language and Cognitive Processes, 24*(7-8), 1082-1096. doi:10.1080/01690960802042133

Zhang, Q., & Yang, Y. (2006). The Interaction of Lexical Selection and Phonological Encoding in Chinese Word Production. . *Acta Psychologica Sinica, 38*(4), 480-488.

Zhang, Q., & Zhu, X. (2016). It Is Not Necessary to Retrieve the Phonological Nodes of Context Objects for Chinese Speakers. *Frontiers in Psychology, 7*, 1161.

Zhang, Q., Zhu, X., & Damian, M. F. (2018). Phonological activation of category coordinates in spoken word production: Evidence for cascaded processing in English but not in Mandarin. *Applied Psycholinguistics*, 1-26.

Zhao, H., La Heij, W., & Schiller, N. O. (2012). Orthographic and phonological facilitation in speech production: New evidence from picture naming in Chinese. *Acta Psychologica, 139*(2), 272-280. doi:https://doi.org/10.1016/j.actpsy.2011.12.001

Zhou, X., & Marslen-Wilson, W. (2000). Lexical representation of compound words: Cross-linguistic evidence. *Psychologia, 43*(1), 47-66 %@ 0033-2852.

Zhou, X., Shu, H., Bi, Y., & Shi, D. (1999). Is there phonologically mediated access to lexical semantics in reading Chinese? In *Reading Chinese script: A cognitive analysis.* (pp. 135-171). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Zhou, Y. (2003). *The Historical Evolution of Chinese Languages and Scripts. Pathways to Advanced Skills Series. Volume 8*: ERIC.

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.

Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain Lang, 144*, 16-25. doi:10.1016/j.bandl.2015.03.007

Zhu, X., Zhang, Q., & Damian, M. F. (2016). Additivity of semantic and phonological effects: Evidence from speech production in Mandarin. *The Quarterly Journal of Experimental Psychology, 69*(11), 2285-2304. doi:10.1080/17470218.2015.1129427

Zhuang, J., & Zhou, X. (2003). The interaction between semantics and phonology in the speech production of Chinese. *Acta Psychologica Sinica, 3*, 002.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The Feedback Consistency Effect in Lexical Decision and Naming. *Journal of Memory and Language, 37*(4), 533-554. doi:https://doi.org/10.1006/jmla.1997.2525

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. doi:10.1111/1467-9280.00247

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Oxford, England: Addison-Wesley Press.