Morphological Encoding of Mandarin Chinese

Evidence from Chinese Disyllabic Compound Words

Published by

LOT | phone: +31 20 525 2461
Binnengasthuisstraat 9
1012 ZA Amsterdam | e-mail: lot@uva.nl
The Netherlands | http://www.lotschool.nl

Cover illustration: photo taken by Jiaqi Wang on the way from her studio Xior to her office at Reuvensplaats in Leiden

ISBN: 978-94-6093-480-3
DOI: https://dx.medra.org/10.48273/LOT0695
NUR: 616

Morphological Encoding of Mandarin Chinese

Evidence from Chinese Disyllabic Compound Words

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 2 juli 2025
klokke 14.30 uur

door

Jiaqi Wang

geboren te Tianjin, China
in 1997

Promotores:	Prof. dr. N.O. Schiller
	Prof. dr. C.C. Levelt
Promotiecommissie:	Prof. dr. J. Grijzenhout
	Prof. dr. Y. Chen
	Prof. dr. A.G. Lewis (City University of Hong Kong)
	Prof. dr. M. Yan (Macau University)

Contents

Acknowledgements

From January 2021 to July 2025, I spent over four years completing my PhD journey. I still remember the first day I arrived in the Netherlands – a country entirely unfamiliar to me. Looking back, my time here has been filled with laughter and tears, moments of doubt and clarity, self-reflection, and growth. This journey has been one of the most transformative experiences of my life, and I will forever cherish every memory.

This thesis is the culmination of my PhD life in the Netherlands. It would not have been possible without the unwavering support and guidance of my supervisors. First and foremost, I extend my deepest gratitude to Prof. Dr. Niels O. Schiller, Prof. Dr. Claartje Levelt, and Dr. Rinus G. Verdonschot. Niels, your encouragement and mentorship have been invaluable – without you, this thesis would never have reached completion. Claartje, thank you for stepping in during my fourth year; your support helped me navigate countless challenges. Rinus, I am deeply grateful for your patience and insightful feedback on my experiments.

I would also like to express my appreciation to those who supported me during my research visits: Prof. Dr. Peter Hagoort and the NBL group in Nijmegen, Prof. Dr. Yanchao Bi for hosting me in Beijing, and the wonderful lab members at City University of Hong Kong for their constructive discussions. A special thanks to Prof. Dr. Xiufeng Zhao, my postgraduate supervisor, whose encouragement and guidance set me on this path and continued to inspire me throughout my PhD.

To my dear friends – you have been my greatest treasure in the Netherlands. Kexin, thank you for being my constant companion, confidant, and travel partner. Our adventures across the Netherlands and beyond are memories I will always hold close. Mengdi Zhu, Zheyu Shang, Ye Jiang, Xiaoqiang Meng, and Xinyuan Chen – our shared moments, deep conversations, gym sessions, or spontaneous trips, have given me strength and joy. A heartfelt thanks to Wenyu Wan and the Leiden Student Union members – Ying Yin, Roujia Feng, Wei Fang, Yao Yao, and others – for making my time here so vibrant.

To my lab mates – Sarah von Grebmer zu Wolfsthurn, Ruixue Wu, Yufang Wang, Yanan Wang, Jin Wang, Qi Zhang, Ziheng Cheng, Xuan Tang, and the rest of the team – thank you for your collaboration, feedback, and camaraderie. Sarah, your help in data collection and data analysis was indispensable. Ruixue, our travels to Lisbon, Paris, and beyond, as well as your unwavering support, mean a lot to me. My wonderful office mates – Qing Yang, Zhen Li, Elisabeth Kerr, Ou Ourilige, Ruoyu Shi, Jian Sun, and Jiahui Liang – you made the office feel like home.

To my LUCL colleagues - Jiang Wu, Xinyi Wen, Tingting Zheng, Renfei Bai, Xin Li, Mathew Sung, Jiaxin Sun, Meipan Yang, Jingtian Shi, and many more – thank you for the shared meals, board game nights, and countless coffee breaks. Jiang, you are talented in cooking and Dutch learning, thanks for the delicious food and for having Dutch class together A special thanks to Louise Friis, Hester Groot, Olga Nozdracheva, Perscilla Lam, Andrew Wigman and others for making the Hortus and the North End pub such welcoming spaces. Yu-Min and Marcel, our Dutch practice sessions were both fun and fruitful – *mijn Nederlands is zeker beter geworden dankzij jullie*!

To the strangers who greeted me with a warm “Goedemorgen” in Leiden, Nijmegen, Amsterdam, Den Haag, Utrecht, Delft, Texel, Maastricht and beyond – your kindness made the Netherlands feel like home.

Above all, I owe everything to my family – my beloved parents, sister, and brother. Your unconditional love and support carried me through the toughest times. Though I missed many family moments while abroad, I carry you all in my heart.

Finally, I draw inspiration from *Arya Stark* from *Game of Thrones* – a “nobody” who grew stronger through adversity. To my “Jaqen H’ghar” – thank you for teaching me to see the world differently, for the Vermeer exhibition, for the starry nights watching Aurora in Leiden, for the walks through various Dutch cities, for watching movies and playing video games together and for the unwavering support during my darkest days. Our travels – from Texel to Antwerp, Nijmegen to the Leiden coast – are memories I will forever treasure. Three-quarters of my PhD life and its outcomes should be credited to “nobody” and “Jaqen”. I felt lucky to have you in my life. And to myself – thank you for persevering.

IV | Acknowledgements

This PhD has been more than an academic journey; it has shaped my resilience, clarified my beliefs, and taught me the value of persistence. It is only the beginning.

The Dutch summer wind (and gloomy winter days as well) will always carry these memories, and I will carry them with me, always.

Publications

Wang, J., Schiller, N. O., & Verdonschot, R. G. (2024). Morphological encoding in language production: Electrophysiological evidence from Mandarin Chinese compound words. *PLoS ONE*, 19(10), e0310816. https://doi.org/10.1371/journal.pone.0310816

Wang, J., Schiller, N. O., & Verdonschot, R. G. (2024). Word and morpheme frequency effects in naming Mandarin Chinese compounds: More than a replication. *Brain and Language*, 259, 105496. https://doi.org/10.1016/j.bandl.2024.105496.

Chapter 1 General introduction

1.1 Background

Human communication relies on various representations essential for effectively conveying thoughts. These representations are thought to reside in our mental lexicon - a cognitive "dictionary" that stores the forms of words we know along with their associated meanings. The nature of word representations within the mental lexicon has been the subject of much debate and research. This thesis focuses particularly on the mechanisms of compounding due to the fact that - as an essential method of word formation - compounding is a fundamental mechanism in both language processing and physical contexts.

Compounds combine morphemes (word constituents), enabling the creation of new words from existing words/morphemes by following an internal structure. Compounding plays a significant role in Mandarin Chinese word formation. For example, disyllabic compounds make up approximately 73.6% of word types and 34.3% of word tokens in the large corpus the Modern Chinese Frequency Dictionary (MCFD) is based on (Language Teaching and Research Institute of Beijing Language Institute, 1986), while monomorphemic words represent 12.0% by type and 64.3% by token (Zhou & Marslen-Wilson, 1995). Since compounding is the dominant morphological process in Mandarin - where derivational and inflectional morphology are relatively rare-it reveals unique properties of morphological processes in Mandarin, distinguishing it from Indo-European languages.

The central theme of the present thesis is the morphological encoding of Mandarin disyllabic noun compounds in language production. Key questions include: How are Mandarin disyllabic noun compounds encoded and processed in language production? What are the underlying behavioral and neural mechanisms involved in producing these compounds? Do speakers first retrieve whole lexical entries from their mental lexicon and then decompose them into their constituent morphemes, or do they process them as complete units? At which level does decomposition occur? These questions will be explored throughout the thesis.

1.2 Language production

Language production is the process of transforming thoughts into spoken words (Schiller, 2020). Speech production models vary in their overall architecture and commonly involve assumptions about how words are organized and represented in the mental lexicon. These language production theories hold the view that language production is a complex process that can be divided into several stages, including conceptual preparation, lexical access, phonological processing, and articulation (Caramazza, 1997; Dell, 1986; Levelt et al., 1999). For instance, when we name an object, the conceptual representation of the object becomes active, activating associated lexical representations. After a lexical item has been selected, the corresponding phonological information needs to be retrieved, and the lexical item is phonologically encoded. Finally, the abstract phonological information is phonetically encoded to yield motor programs used for articulation through the execution of the corresponding gestural scores.

Models of speech production often propose distinct sequential lexical levels, i.e., the lemma and the word-form level (Dell, 1986; Dell & O'Seaghdha, 1992; Garrett, 1980; Levelt, 1993; Levelt et al., 1999; Roelofs, 1996a, 1996b; Roelofs & Meyer, 1998; Roelofs et al., 1996). Lemmas capture a word's syntactic properties, while word forms convey segmental and metrical or supra-segmental details, including constituent morphemes in polymorphemic words (Roelofs, 1996a, 1996b; Roelofs & Meyer, 1998). In the two-stage model proposed by Levelt et al. (1999), producing compound words begins by activating the relevant semantic concept (i.e., dish washer), which, through spreading activation, co-activates semantically related concepts (e.g., washing machine, dryer, microwave, oven, etc.). This co-activation leads to the simultaneous activation of multiple, semantically related lexical entries at the lemma level, where they compete for selection until a single target lemma is chosen (Abdel Rahman & Aristei, 2010; Abdel Rahman & Melinger, 2009; Damian & Bowers, 2003).

Morphological structure influences speech production, although for a long time, language production models did not assign a distinct role to morphological processing (Schiller & Verdonschot, 2019). Evidence supporting the role of morpheme comes from some speech planning experiments and from speech errors. For instance, Roelofs

(1996) compared the naming latencies of word sets including an overlapping morpheme to a set of words with the same amount of phonological overlap, found a significantly greater facilitation effect for the former group compared to the latter when both were compared to a set of words without phonological overlap. This led him to conclude that morphemes serve as planning units in speech production. Besides, evidence from speech errors, such as "a floor full of holes" becoming "a hole full of floors" or "I carved a pumpkin" turning into "I pumped a carven" (Fromkin, 1973) further supports the crucial role of morphology in language production.

The question of how words and morphemes are structured and represented during Mandarin compound production remains unclear. To answer this question, the current thesis adopts these production models, focusing on the lexical selection stage and the morphological encoding processes of Mandarin compounds in the mental lexicon. In the following sections, we review relevant models of the process of compounds during language production.

1.3 The representations of compounds

Numerous studies emphasize the importance of morphology in speech production (Caramazza, 1997; Dell, 1986; Levelt et al., 1999). For instance, in Levelt et al.'s (1999) framework, morphology is considered the initial stage of word-form encoding. To investigate morphological encoding in compound production, two primary hypotheses are often discussed: the decompositional view, which posits that compounds are represented at one or more levels based on their constituent morphemes (Levelt et al., 1999), and the full-listing model, which proposes that only whole-word forms are stored in the lexicon (Butterworth, 1983; Caramazza, 1997; Dell, 1986).

A less common hybrid model combines elements of both approaches, suggesting that transparent compounds (those with meanings closely related to their constituents) are processed by decomposing them, allowing for simultaneous access to whole-form and constituent representations. In contrast, opaque compounds (i.e., meaning not related to their parts), especially high-frequency ones, are processed through whole-form access (MacGregor & Shtyrov, 2013). Levelt et al.'s (1999) model may also be seen as somewhat "hybrid," as it suggests that some compound words are "degenerate" in production, meaning they are not decomposed at the form level. The current thesis focuses on the two hypotheses of compound representation, with details on these hypotheses presented in the following two sections.

1.3.1 The full-listing hypothesis

The full-listing hypothesis suggests that only whole-word forms are represented in the lexicon (Butterworth, 1983; Caramazza, 1997; Dell, 1986). There are studies showing support for this account by manipulating the frequency of compound words and their morphemes. For instance, Janssen et al. (2008) conducted a study where participants named pictures by manipulating the frequency of the compound word and its constituents using three conditions: H(h) (high compound and high morpheme frequency), L(h) (low compound and high morpheme frequency), and L(l) (low compound and low morpheme frequency). For instance, 山羊 /shan1yang2/ “goat” was a low-frequency word, but its constituents 山 /shan1/ “mountain” and 羊 /yang2/ “sheep” were high-frequency morphemes. Their results showed that only the frequency of the whole compound word influenced naming latencies, supporting the full-listing hypothesis.

Further support for the full-listing model comes from Bi et al. (2007), who investigated two aphasic patients with difficulties in lexical access for oral and written production. Their study demonstrated that the frequency of the compound word itself, not the frequency of the constituent morphemes, affected production performance in both patients.

Chen and Chen (2006) used the implicit priming task (Meyer, 1991) to investigate whether morphological encoding played a role in producing disyllabic transparent compound words in Mandarin Chinese. Participants named compound words in a response-association task. Contrary to findings from the Dutch, their results showed that naming latencies were not sensitive to morpheme frequency, which supported a single-stage model of lexical access (Caramazza, 1997). These findings lent support to models suggesting that compound words are stored as unique lexical units, challenging theories that promote morphological decomposition during word production.

1.3.2 The decompositional hypothesis

Several studies provide evidence supporting the decomposition model, which posits that compounds are accessed through their constituent morphemes. Roelofs (1996), for instance, investigated whether the form lexicon used in speech production included morphologically decomposed entries by manipulating word frequency. His findings suggested that morphemes, as components of compound words, served as planning units in speech production, reinforcing the decomposition hypothesis. Both high- and low- frequency morpheme constituents showed a facilitatory effect - an effect that would not be expected if compounds were stored holistically, as a full-listing model would imply that constituent morpheme frequency should not impact production.

Similarly, Bien and colleagues (2005) conducted four experiments to examine the role of frequency in compound production by independently varying the frequencies of each constituent and the compound as a whole. Their results showed that compound production was sensitive to the cumulative frequency of morphemes, further supporting the decomposition model.

Chen and Chen (2015) investigated a word-onset-first planning strategy in Mandarin Chinese monomorphemic and bimorphemic compound words. Using the first character as a cue for the second, they observed a significant preparation effect, indicating that the second morpheme in a compound could be pre-planned in Mandarin. This finding aligned with the notion of morphological decomposition in compound production.

Most studies have employed paradigms without inherent lag, such as implicit priming or form preparation paradigms, which may involve both semantic and phonological activation. In contrast, the long-lag priming paradigm has proven to be a more robust method for examining morphological representation, consistently producing reliable results across various experimental techniques - behavioral, electrophysiological, and hemodynamic - and across languages (Schiller & Verdonschot, 2018). For example, Zwitserlood and colleagues (2000) used this paradigm to investigate morphological priming, where the prime was morphologically related to the target, finding that morphological priming remained significant even after multiple intervening trials. By contrast, semantic and phonological priming effects diminished at longer lags, suggesting that morphological priming operated at a distinct level from semantic or phonological priming. This conclusion was further supported by additional studies (Zwitserlood, 2000, 2002; Kaczer et al., 2015; Koester & Schiller, 2008, 2011; Verdonschot et al., 2012; Lensink et al., 2014).

The decompositional model for compounds proposes that decomposition occurs at least at one level of lexical selection, potentially at the lemma level or at the lexeme level, or both. Therefore, another central focus in the following section of this thesis is the lemma representation of Mandarin compounds.

1.3.3 The lemma representation of compounds

According to prevailing theories of speech production, naming an object (e.g., a bird) involves activating conceptual information, retrieving the relevant lexical entry, and initiating phonological and phonetic encoding prior to articulation (Caramazza, 1997; Dell, 1986; Levelt et al., 1999). While substantial empirical evidence supports the role of morpheme-based representations in speech production (Jacobs & Dell, 2014; Janssen et al., 2008; Lüttmann et al., 2011; Roelofs, 1996b), it remains unclear at which specific level these representations operate - whether at the lemma level (Levelt et al., 1999), the lexical node level (Dell, 1986), or the word-form level. The decompositional model for Mandarin compounds suggests that decomposition occurs at least at one level of lexical selection, indicating that decomposition may occur at both the lemma and lexeme levels.

Therefore, the present thesis included two experiments to explore this question: one employed distractor that was synonymous to the first constituent of compound targets in a picture-word interference task; the other examined the concreteness effect of the constituents of compound targets in a picture-naming task. These two experiments aimed to determine whether compounds were stored solely as holistic units at the lemma level, supporting a single-lemma representation account (Levelt et al., 1999), or if they were stored in a decomposed format alongside a single lemma, supporting hybrid lemma representation accounts (Sprenger et al., 2006; Kuiper et al., 2007).

Below, we discussed two models regarding the lemma representation of compounds.

Levelt and colleagues (1999) proposed that compound words were stored holistically at the lemma level, allowing access to multiple morphemes at the word-form level. After selecting a single lemma, the constituent morphemes were retrieved at the word-form level, followed by morpho-phonological and phonetic encoding before articulation. Evidence from sequential processing suggested that a compound's morphemes were encoded one after another, with the first morpheme being processed before the second (Roelofs, 1996b). Taken together, two-stage models advocated for single, holistic compound lemmas with decomposed (morpheme-based) form representations (Dell, 1986; Levelt et al., 1999).

Alternatively, a model proposed by Sprenger et al. (2006) suggested a hybrid approach, positing that idiom representations existed alongside individual constituent representations at the lemma level. They conceptualized idiomatic forms as "super lemmas" - distinct units that encompassed information about the syntactic constraints linked to the idiom, thus outlining the syntactic properties of the individual lemmas involved. Given the parallels between idioms and compounds in mapping onto a single conceptual representation while consisting of multiple morphemes, this thesis proposed that compounds might share a similar lemma-level representation as idioms, involving a holistic lemma supported by constituent lemmas.

1.4 Electroencephalographical evidence of compounds production

This thesis combined behavioral measures, such as reaction times and naming latencies, with electrophysiological measures using electroencephalography. While most previous studies on compound word production have primarily relied on behavioral data, the precise neural correlates of morphological processing remain unclear. As a non-invasive technique for measuring event-related brain potentials, EEG could provide valuable insights by tracking brain activity that is time-locked to specific events, such as stimulus presentation during a task (Woodman, 2010). Unlike reaction times, ERPs offer high temporal resolution, enabling real-time observation of cognitive processes before an explicit response occurs (Kutas & Van Petten, 1994). For example, ERPs have effectively explored morphological decomposition processes during word comprehension in visual and auditory modalities (Krott et al., 2006; Fiorentino & Poeppel, 2007; Koester et al., 2007).

ERPs include various components that are either negative or positive in polarity, each associated with specific linguistic processing responses. In the present thesis, we focused on the N400 components and the N400 effect to investigate the morphological encoding of Mandarin compounds in language production.

The N400 is a negative voltage peak that reaches its maximum amplitude approximately 400 ms after a stimulus word appeared (Kutas

& Hillyard, 1984). Although every word elicits an N400 component, the difference in N400 amplitude between contextually appropriate and inappropriate words is known as the N400 effect. Initially linked to sensitivity about lexical semantics, the N400 is now understood to reflect the ease of integrating words into context (Schiller & Verdonschot, 2019). In the process of morphological encoding, ERPs can serve as a valuable tool for investigating whether morphological priming occurs at the word form level, offering detailed insights into the temporal dynamics of this process. For instance, one ERPs study on word reading found that the N400 component was sensitive to both lexical status and morphological decomposition (McKinnon et al., 2003).

The process of morphological decomposition affects the N400 component in ERP studies, because it is associated with the ease of integrating words at the lexico-semantic level. In the present thesis, the N400 is particularly responsive to priming and word frequency effects. First, when a target item is primed, the brain is better prepared for the upcoming stimulus, resulting in a smaller N400 amplitude. In the context of language production, Koester and Schiller (2008) examined the temporal aspects of morphological encoding using a picture naming task with Dutch compound words and a long-lag priming paradigm. Their findings highlighted a connection between the N400 component and morphological encoding. Additionally, word frequency influences N400 amplitude, with high-frequency words eliciting smaller N400 responses than low-frequency words (Kutas & Federmeier, 2009, 2011;

Rugg, 1990; Van Petten & Kutas, 1990). High-frequency words, which occur more often, are easier to integrate than low-frequency words, requiring less cognitive effort for retrieval. The present thesis investigates explicitly N400 effects of morphological priming and word frequency to provide a detailed examination of the representation of Mandarin compounds in language production.

1.5 The current study

As previously outlined, the central issue of this thesis concerned the representation of Mandarin compounds, with two dominant hypotheses under debate: the decompositional and full-listing hypotheses. Given the lack of consensus in the existing literature regarding the production of Mandarin compounds, this thesis aimed to explore this question by conducting four studies employing behavioral and electrophysiological approaches. Each study focused on different aspects of Mandarin compound representation to comprehensively investigate the issue.

Chapter 2 of this thesis examined the role of morphology in speech planning in Mandarin Chinese through a long-lag priming experiment. Thirty-two native Mandarin speakers were asked to name target pictures (e.g., 山 /shan1/ “mountain”). The study employed two types of compound primes: morpheme-related (e.g., 山羊 /shan1yang2/ “goat”) and morpheme-unrelated (e.g., 飞机 /fei1ji1/ “airplane”) primes for monomorphemic targets. The target could overlap with the related prime in either the first constituent (e.g., target 山 /shan1/ "mountain" with prime 山羊 /shan1yang2/ "goat") or the second constituent (e.g., target 包 /bao1/ "bag" with prime 面包 /mian4bao1/ "bread"). Both behavioral and electrophysiological data were collected using the long-lag priming paradigm. The results supported the decompositional encoding of Mandarin compounds, although the behavioral findings contradicted earlier results from Indo-European languages. Behaviorally, naming latencies for target pictures were not facilitated by morphologically related primes. However, ERP analyses revealed that morpheme-related targets elicited a reduced N400 response compared to morpheme-unrelated targets, but only when the overlap occurred in the first constituent, not the second. The discrepancy between the behavioral and EEG findings led to additional experimental designs and methods to explore this issue further in subsequent chapters.

In **Chapter 3**, we attempted to replicate the study of Janssen et al. (2008) and added the collection of EEG data. Whether compound words are stored in our mental lexicon in a decomposed or full-listing way prompted Janssen and colleagues (2008) to investigate the representation of compounds using word and morpheme frequency manipulations. As we deemed the approach to exploring the question of Mandarin compound representation and production presented in the study of Janssen and colleagues suitable for the current study, we opted to replicate their experimental design but extend their study by introducing a new set of stimuli based on the SUBTLEX-CH Mandarin speech production corpus (Cai & Brysbaert, 2010). Additionally, we utilized EEG methodology to examine the temporal dynamics of compound production in greater detail. Regarding the current issue, ERPs can serve as a direct method to investigate the frequency effect for constituents and the whole word because words with a higher frequency tend to trigger N400s of reduced amplitude compared to words with a lower frequency. In the current study, despite ERPs analyses revealing no word frequency or morpheme frequency effects across conditions, behavioral outcomes indicated that Mandarin compounds are not sensitive to word frequency. Instead, the response times highlighted a morpheme frequency effect in naming Mandarin compounds, which contrasted with the findings of Janssen and colleagues. These findings challenged the full-listing model and instead supported the decompositional model.

Chapter 4 investigated the lexical representation of Mandarin compounds through a picture-naming task using the picture-word interference (PWI) paradigm. Thirty-nine native Mandarin speakers named 45 pictures of disyllabic noun compounds, such as 公鸡 /gong1ji1/ "rooster" (lit. "male chicken"), with three types of distractors: synonymous distractors of the first morpheme (e.g., 雄 /xiong2/ "male"), distractor words semantically related to the compound (e.g., 鹅 /e2/ "goose"), and unrelated distractors (e.g., 车 /che1/ "car"). The reactions times were recorded in this experiment. Results in the present study showed that synonym distractors produced significantly faster naming times than both semantically related and unrelated distractors, while semantically related distractors slowed responses compared to unrelated ones. These findings support a decompositional representation and a hybrid lemma representation of Mandarin compounds in the mental lexicon.

In **Chapter 5**, we explored the concreteness effect of morpheme constituents to examine how Mandarin compounds were represented in the mental lexicon, specifically whether abstract and concrete morphemes influenced the retrieval of Mandarin compounds during production. Our study addressed this question using a picture naming task with 41 participants. Two conditions were tested: the “aa” condition, representing compounds with two abstract constituents, and the “cc” condition, representing compounds with two concrete constituents. Behavioral results revealed a concreteness effect in Mandarin compounds, as naming latencies were significantly faster for compounds with two concrete constituents (“cc” condition) than those with two abstract constituents (“aa” condition). These findings supported the decompositional model, highlighting constituents’ vital role in producing Mandarin compounds.

Chapter 6 integrated the findings from each of the studies presented in this thesis. In this chapter, we synthesized our research results, discussed their theoretical implications, and proposed future directions. The goal was to answer the central question posed at the beginning of this thesis: How are Mandarin compounds represented and encoded in the mental lexicon during language production?

Chapter 2 Morphological encoding in language production: Electrophysiological evidence from Mandarin compound words

This article is published as: Wang, J., Schiller, N. O., & Verdonschot, R. G. (2024). Morphological encoding in language production: Electrophysiological evidence from Mandarin Chinese compound words. PLoS ONE, 19(10), e0310816. https://doi.org/10.1371/journal.pone.0310816

Abstract:

This study investigates the role of morphology during speech planning in Mandarin Chinese. In a long-lag priming experiment, thirty-two Mandarin Chinese native speakers were asked to name target pictures (e.g., 山 /shan1/ "mountain"). The design involved pictures referring to morpheme-related compound words (e.g., 山羊 /shan1yang2/ "goat") sharing a morpheme with the first (e.g., 山 /shan1/ "mountain") or the second position of the targets (e.g., 脑 /nao3/ "brain" with prime 电脑 /dian4nao3/ "computer"), as well as unrelated control items. Behavioral and electrophysiological data were collected. Interestingly, the behavioral results went against earlier findings in

Indo-European languages, showing that the target picture naming was not facilitated by morphologically related primes. This suggests no morphological priming for individual constituents in producing Mandarin Chinese disyllabic compound words. However, targets in the morpheme-related word condition did elicit a reduced N400 compared with targets in the morpheme-unrelated condition for the first position overlap in the ERP analyses but not for the second, suggesting automatic activation of the first individual constituent in noun compound production. Implications of these findings are discussed.

Keywords*: Overt language production, Morphology, Chinese Compound words, Long-lag paradigm, ERP*

2.1 Introduction

Human communication involves a multitude of distinct representations to convey our thoughts effectively. These representations are thought to reside within our mental lexicon. The mental lexicon is often described as a cognitive dictionary and refers to the part of our language system that hosts the lexical forms we know and their corresponding meanings.

Language processing theories contain assumptions about the organization and representation of words in the mental lexicon. For example, language production theories typically describe a sequence of cognitive processes encompassing various information types, such as conceptual preparation, lexical access, phonological processing, and articulation (Caramazza, 1997; Dell, 1986; Levelt et al., 1999). The intended object's conceptual representation becomes active when engaged in speech production, for instance, when naming a picture. This activation then extends to the lexical representations associated with these concepts. Subsequently, phonological information is retrieved, encompassing word form encoding, which is ultimately utilized for articulation by triggering the relevant gestural sequences. In speech production, morphological encoding concerns activating the words' constituents at the lexical-morphological level. Morphemes are the minimal units that convey meaning, and many studies have substantiated the involvement of morphology in speech production (Caramazza, 1997; Dell, 1986; Levelt et al., 1999). For instance, in the word production theory by Levelt and colleagues (Levelt et al., 1999), morphology is the first stage of word-form encoding.

In Chinese, compounds are morphemes (constituents) combinations with compounding, the most common method for constructing new words. Unlike classical Chinese, which is primarily monosyllabic, modern Chinese, specifically Mandarin, has increased the abundance of compound words, where multiple characters (morphemes) come together to form a word with a specific meaning (Zhou & Marslen-Wilson, 1995; Zhou et al., 1999; Zhou & Marslen-Wilson, 2014). The morphological processes of Mandarin Chinese are characterized by the dominance of compounding and the effective lack of derivational and inflectional morphology, which are abundantly present in most Indo-European languages (Zhou & Marslen-Wilson, 1995; Zhou et al., 1999; Zhou & Marslen-Wilson, 2014). For example, disyllabic compounds comprise about 73.6% by type and 34.3% by token in a large text corpus of Modern Chinese Frequency Dictionary (MCFD) (Language Teaching and Research Institute of Beijing Language Institute, 1986). Some constituents cannot meaningfully stand alone in these compounds and could be considered bound morphemes. For example, in 骆驼 /luo4tuo2/ “camel,” the individual morphemes (e.g., 骆 and 驼) are meaningless when presented in isolation, although they can be part of other compounds (e.g., 驼背 /tuo2/bei4 “hunchback”). However, for a word like 山羊 /shan1yang2/ “goat,” the individual constituents are free-standing morphemes (e.g., 山 /shan1/ “mountain” and 羊 /yang2/ “sheep”) and can be used in a sentence (e.g., 这个山很高 /zhe4ge4/ /shan1/ /hen3/ /gao1/ “this mountain is very tall”). Note that some morphemes have become “special cases,” such as 子 /zi3/, which, when free-standing, means “child” but has also become a commonly used suffix to form (concrete) nouns, such as 椅子 /yi3zi5/ “chair” or 帽子 /mao4zi5/ “hat.”

One other potentially important aspect to consider for (literate) Chinese is that due to the make-up of the Chinese writing system, literate Chinese may have some additional awareness of the meaning of the individual constituents. As English readers need to convert sound into meaning (e.g., sound → meaning), the opposite route most likely takes place for Chinese (e.g., meaning → sound), and each morpheme is typically one character (i.e., they are visually marked). This does not mean that English speakers do not realize the meaning of individual constituents, especially for transparent compounds (e.g., they would certainly be aware of "birdhouse" = bird + house). Still, a difference between the languages becomes apparent at an infrequent English (old loan) word such as "xylophone." This would likely not be parsed in "wood" + "sound" as many might not know that the sounds "xylo" and "phone" (which originally stem from Greek) bear those specific meanings. However, in Chinese, "xylophone" 木琴 /mu4qin2/, despite being infrequent as well, due to the logographic nature of the script, would still clearly signify the free morpheme 木 /mu4/ "wood." Less obvious transparent compounds might still elicit their individual constituents more strongly in Chinese; for example, "microwave" (微波 /wei1bo1/ "microwave," which is "tiny" + "wave") as the script visually marks the individual constituents (e.g., 波 /bo1/ "wave").

Therefore, the question arises of how Mandarin compound words are stored in the mental lexicon as Mandarin Chinese's morphological makeup differs from that of Indo-European languages. There are two alternative hypotheses. One is the *full-form hypothesis* (or full-listing hypothesis), which assumes that only whole-word forms are represented in the lexicon and that a word's derivational history plays no role in lexical access (Butterworth, 1983; Laudanna et al., 1992; Janssen et al., 2008). The alternative view, the *decompositional hypothesis*, assumes at least one level of lexical representation at which compounds are represented in terms of their constituent lexical morphemes (Levelt et al., 1999).

2.1.1 Full-listing hypothesis

The issue of compound representation has previously been studied by manipulating the frequency of compound words and their morphemes. The word frequency effect is one of the most robust findings in picture naming: producing an infrequent word (such as “eclipse”) is substantially slower than producing a frequent word (such as “table”) (Levelt et al., 1999). Janssen and colleagues (2008) asked people to name pictures (referring to compound words). They manipulated the frequency of the compound word and its constituents using three conditions: *HH*, *LH*, and *LL* (where the first letter refers to high (H) or low (L) whole word frequency and the second letter to high (H) or low (L) constituent frequency). For example, 山羊 /shan1yang2/ “goat” would be a low-frequency word, but its constituents 山 /shan1/ “mountain” and 羊 /yang2/ “sheep” are both high-frequency morphemes. They found that only whole word frequency influenced naming latencies, and hence, they proposed that their findings support the full-listing hypothesis.

Additional support comes from Bi and colleagues (2007), who investigated two aphasic patients with difficulty in lexical access for oral and written production, respectively. They showed that the frequencies of the compound words themselves, but not the frequency of the constituent morphemes, affected the production performance of both patients.

In another study by Chen and Chen (2006), using the implicit priming task (Mayer, 1991). Mandarin Chinese participants named compound words in the context of a response-association task to investigate whether morphological encoding is involved in producing Chinese disyllabic transparent compound words. Their results revealed, contrasting with findings in Dutch, that naming latencies were not sensitive to the compound’s morpheme frequency. This supports the assumption that word frequency is related to word forms and the interpretation of a single-stage model of lexical access. These findings align with the hypotheses proposing the existence of a single lexical level during word production and positing that compounds possess distinct lexical nodes within this level (Caramazza, 1997). They pose a challenge to the viewpoint presented by theories assuming morphological decomposition (Levelt et al., 1999).

2.1.2 Decompositional hypothesis

Some studies also present results in favor of the decomposition model. In Roelofs’s paper (1996), the issue of whether the form lexicon underlying speech production contains morphologically decomposed entries was addressed by manipulating word frequency. Only decomposed form entries would allow morphemes to be planning units in speech production. The outcome supported the idea that component morphemes of Dutch compound words were planning units in speech production and confirmed the decomposition assumption. Both high-frequency and low-frequency morphemic constituents yielded a facilitatory effect. Similarly, in a study by Bien and colleagues (2005), four experiments were conducted to investigate the role of frequency information in compound production by independently varying the frequencies of the first and second constituents and the compound itself. The results of these experiments revealed that compound word production was sensitive to cumulative morpheme frequency, which supports the decomposition hypothesis.

Furthermore, Chen & Chen (2015) investigated phonological planning in Mandarin spoken production, focusing on Mandarin words with two characters (e.g., 玻璃 /bo1li2/ “glass”). All their experiments used the form preparation task with a slight twist. Instead of learning and producing word pairs (e.g., seeing the word 梳子 /shu1zi5/ “comb” then responding with 头发 /tou2fa3/ “hair”) which is typical in this paradigm, participants had to respond with the second constituent of a word (i.e., when encountering 玻 (/bo1/) they had to respond with 璃 (/li2/). Their idea was that when the domain of phonological planning is the entire word, the planning must always begin at the start of the word. However, phonological planning might also be influenced by the overlapping beginning of a non-initial morpheme when the domain is the morpheme. Indeed, Chen & Chen observed significant preparation effects when the targets (i.e., the second part of the disyllabic words) shared the same atonal syllable in homogenous groups (e.g., 玻璃 /bo1li2/ “glass” – 茉莉 /mo4li4/ “jasmine” – 霹雳 /pi1li4/ “thunderbolt” and 淅瀝 /xi1li4/ “pattering”). In other words, the data of Chen & Chen suggested that Mandarin speakers could plan directly from the second morpheme. Although this task is, admittingly, influenced by the experimental environment, it does support the decomposition hypothesis, as the overlap in the second constituent provided the facilitatory effect (something the full-listing hypothesis would have trouble explaining). Therefore, there still seems to be an ongoing debate on whether (Chinese) compound words are stored in our mental lexicon in a decomposed or full-listing manner.

2.1.3 Long-lag paradigm

The studies investigating the representation of compounds mostly use paradigms without inherent lag, such as the implicit priming or form preparation paradigm, which may also involve activation from semantic or phonological encoding. Conversely, the long-lag priming paradigm is established and robust, especially focusing on the morphological level of representation that has proven to yield consistent and replicable results across experimental methods (i.e., behavioral, electrophysiological, and hemodynamic) and languages (Schiller & Verdonschot, 2018). For instance, Zwitserlood and colleagues (2000) used the long-lag paradigm to investigate the morphological priming effect. They demonstrated that morphological priming, where the prime is morphologically related to the target, remains effective despite many intervening trials. However, effects related to semantic and phonological priming were no longer observable at large lags. These findings imply that priming in those instances does not occur at a phonological or semantic level but at a distinct morphological level. This was subsequently corroborated by further investigations (Zwitserlood et al., 2002; Kaczer et al., 2015; Koester & Schiller, 2008; Koester & Schiller, 2011; Verdonschot et al., 2012; Lensink et al., 2014).

2.1.4 Electrophysiological evidence

Most of the previous investigations on compound production used behavioral measures (with some notable exceptions) as mentioned above; it is therefore still unclear how the precise neuro correlates of morphological processing are manifested. Unlike reaction times (RTs), event-related potentials (ERPs) offer a higher temporal resolution that enables more direct observation of cognitive processes, even before or without an explicit response (Kutas & Van Petten, 1994). ERPs have proven useful in testing or confirming decomposition processes of compound comprehension in the visual and auditory modalities (Krott et al., 2006; Fiorentino & Poeppel, 2007; Koester et al., 2007). An ERP study investigated the effects of morphological decomposition in word reading and concluded that the N400 was sensitive to lexical status and morphological decomposition (McKinnon et al., 2003). Therefore, ERPs can serve as a direct means to investigate whether morphological priming originates at the word form level, allowing for a detailed exploration of the temporal dynamics of morphological priming.

The process of morphological decomposition significantly impacts the N400 component within the context of event-related potentials (ERPs), which is closely linked to the ease of integrating words at the lexical-semantic level. This concept was initially proposed and subsequently confirmed in the field of language comprehension. Numerous studies have since then investigated this phenomenon (Barber & Domingues, 2002; Dominguez et al., 2004; Morris et al., 2007). A pronounced N400 component becomes apparent when participants are presented with unexpected stimuli (Kutas & Federmeier, 2011). When a target item is primed, it subconsciously prepares participants for upcoming content, leading to fewer surprises than an unprimed target. Consequently, an anticipated reduction in the N400 peak's amplitude occurs. In language production, Koester and Schiller (2008) examined the temporal aspects of morphological encoding in a picture naming task with Dutch compound words as stimuli. Their findings indicated a connection between the N400 component and morphological encoding. Moreover, a comprehensive meta-analytic review of the existing literature about twenty years ago proposed the existence of a morphological encoding process approximately 330 ms after the onset of a picture, assuming an average latency of 600 ms for picture naming (Indefrey & Levelt, 2004). Note that the time window of word form encoding could range between 217 and 530 ms (and likely even beyond that) due to uncertainties in their temporal estimates, as Indefrey and Levelt (2004) pointed out.

2.1.5 The position effect

Chinese disyllabic compounds have two constituents, and each constituent is either in the first or the second position. The question of whether the position will influence the decomposition process has been offered. There are two related claims. The first claim is based on the serial planning assumption, which suggests that the noninitial morphemes of a word cannot be planned before initial ones. This might lead to the expectation of larger priming effects for initial constituents (Roelofs, 1996; Roelofs & Baayen, 2002). Additionally, the fact that modifier and head constituents (first and second constituents) are not processed alike could also affect the priming effect for two different positions. For intense, Bien et al. (2005), investigated how different variables of the modifier and the head constituents affected the naming latencies of Dutch compounds. The results showed that all relevant modifier variables had facilitative effects, whereas facilitative and inhibitory effects were found for the head constituents.

Another claim can be found in Zwitserlood and colleagues' work (2000). Their study showed that the position of the target morpheme does not influence morphological priming-both the first and second constituent of compounds cause faster-naming latencies in morphologically related targets. In addition, in Koester and Schiller's paper (2008), the authors proposed that position may play a role, but a position effect could not be straightened out from their experiment, as they used two sets of stimuli with unbalanced overlap rates. Further research is therefore warranted to clear up the existence of any potential position effect. Hence, the present study controlled the first and second positions to investigate this matter further.

2.1.6 The goal of the present study

Given the situation laid out in the previous paragraphs, the present study designed a morpheme-related compound prime (e.g., 山羊 /shan1yang2/ “goat”) and a morpheme-unrelated compound prime (e.g., 飞机 /fei2ji1/ “airplane”) for a monomorphemic target. The target could overlap with the related prime in either the first (e.g., target 山 /shan1/ “mountain” with prime 山羊 /shan1yang2/ “goat”) or second (e.g., 包 /bao1/ “bag” with prime 面包 /mian4bao1/ “bread”) position. Both behavioral and electrophysiological data were collected using a long-lag priming paradigm. Faster RTs were expected in the morphologically related priming conditions (Dohmes et al., 2004). A reduced N400 amplitude is expected for the morphologically related priming conditions (McKinnon et al., 2003). The reduced N400 amplitude (here reflecting morphological encoding) is expected to begin around 330 ms after picture onset, assuming an average naming latency of 600 ms (Indefrey & Levelt, 2004). For the variables of position, we expected that the priming effect for the first position should be larger than the second position, given the serial planning assumption and the modifier facilitation effect in naming latencies (Bien et al., 2005; Roelofs, 1996; Roelofs & Baayen, 2002) contrasting with Zwitserlood and colleague's work (2000). The N400 voltage is reduced for the first position in the same time window as less cognitive effort is exerted for retrieving the first constituents.

2.2 Methodology

2.2.1 Participants

We recruited thirty-six right-handed native Chinese participants (twenty-one females) from Leiden University (the Netherlands), who all graduated at least with a bachelor's degree in China. All participants received monetary compensation for their participation. The mean age of participants was 26.6 ± 3.2 years. This study was approved by the Faculties of Humanities and Archaeology ethics committee at Leiden University (acceptance number: 2022/09). Recruitment took place between September 2022 and November 2022.

Thirty-four participants were from regions where Mandarin is spoken, including provinces of north-eastern China, north-western China, and south-western China. Two participants were native Mandarin/Cantonese bilinguals.

At the time of testing, none of the participants reported color blindness, learning disorders, hearing or visual impairments, or psychological or neurological impairments. Participants read the information sheet and signed an informed consent form before participating.

2.2.2 Materials

Forty target pictures (all monosyllabic Chinese words) were chosen, and each was paired with one morpheme-related disyllabic Chinese compound word (picture) and one morpheme-unrelated disyllabic Chinese compound word (picture) as primes. Therefore, 120 black-and-white line drawings corresponding to target pictures, morpheme-related, and morpheme-unrelated prime pictures were presented as stimuli. All compounds in the stimuli list were nouns.

Seven or eight pictures, including fillers, targets, and primes, were put between target pictures and prime pictures for each pair (one target picture was paired with either one morpheme-related prime picture or one morpheme-unrelated prime picture) as intervening trials.

Table 2.1: Example of a target picture with four prime types.

Prime type	Example (Prime)	Example (Target)
Morpheme-related/first position	山 羊 /shan1yang2/ goat	山 /shan1/ mountain
Morpheme-unrelated/first position	飞机 /fei1ji1/ airplane	
Morpheme-related/second position	风 车 /feng1che1/ windmill	车 /che1/ vehicle
Morpheme-unrelated/second position	羽 毛 /yu3mao2/ feather	

We included 124 Chinese disyllabic filler pictures and 28 monosyllabic filler pictures to allow for the precise creation of the intervening trials. Four conditions were created in the present experiment (see Table 2.1).

Each target picture was presented twice: once with a morpheme-related prime picture pair and another time with a morpheme-unrelated prime picture pair; therefore, 464 trials were administered during the experiment. The word frequency for the four conditions was controlled based on the Zipf values from the SUBTLEX-CH corpus (Cai & Brysbaert, 2010) (F (3,76) = 1.392, p > 0.05) as well as the number of strokes (F (3,76) < 1, p > 0.05) (see Table 2.2).

Table 2.2: Mean and SD of the number of strokes and word frequency (Zipf values) for prime types.

Prime type	Strokes		Word Frequency (Zipf)	
	Mean	SD	Mean	SD
Morpheme-related/first position	16.55	4.174	2.242	0.771
Morpheme-related/second position	16.40	4.358	2.181	0.785
Morpheme-unrelated/first position	18.60	5.519	2.476	0.652
Morpheme-unrelated/second position	15.10	4.866	2.346	0.427

Before the experiment, we tested whether stimuli have similar visual complexity (F (3,76) < 1, p > 0.05). Prime and target pairs were phonologically and orthographically unrelated. *Appendix 2.A* presents an overview of all prime-target combinations used in the experiment.

2.2.3 Design

A two-factorial within-subject design was adopted in this experiment, with Morpheme Relatedness (Related/Unrelated) and Position (First/Second) as the two main factors. Four conditions were created.

Two sets of main stimulus lists were prepared, and within each set were two sub-lists. Each sub-list had 232 different trials, which contained 20 primes in the related condition and 20 primes in the unrelated condition, along with 40 targets and 152 fillers. The first sub-list includes one-half of the related and unrelated primes, while the second contains the other half. Both sub-lists share the same 40 targets and 152 fillers. Each participant saw all the target and filler pictures twice and all prime pictures once, resulting in 464 (232*2) trials per person in the present study. Furthermore, intervening trials contained no phonologically or semantically related items to the target pictures.

Each sub-list was divided into four sub-block groups; hence, there were 58 trials in each sub-block. In total, eight of these sub-blocks were created for the present experiment. There were breaks among sub-blocks. Half of the participants were given the first stimulus list set during the experiment, while the other half received the second set. This approach was taken to balance the experiment and eliminate potential biases.

Due to Chinese people tending to use disyllabic words to name monosyllabic words, we used red and green frames to indicate how many words should be named in our experiment. Red-framed pictures are for monosyllabic targets and fillers, and green-framed pictures are for disyllabic primes and fillers. Because all primes (in both morphological-related and morphological-unrelated conditions) were two-character words, all prime pictures would be in green frames. Therefore, the potential effect that might be caused by different cues was counterbalanced across two prime conditions. The same goes for the target, where all target pictures were in red frames. Therefore, for two different positions, the color effect was counterbalanced as well.

2.2.4 Procedure

The experiment was designed and controlled using *E-prime 3.0* (Psychology Software Tools) and was conducted in a soundproof booth. The procedure was similar to Koester and Schiller's (2008), adapted from Dohmes et al. (2004). First, participants were given 10-15 minutes to familiarize themselves with the target and prime picture names by studying a booklet. Subsequently, the experimenter assessed whether participants correctly remembered the picture names by conducting a practice session through E-prime. In this session, participants were asked to name all the target and prime pictures as soon as possible, and the experimenter would correct them if they named those pictures erroneously.

In the experiment session, each trial began with a fixation cross displayed for 250 ms, followed by a blank screen for another 250 ms. Then, the stimulus appeared in the center of the screen for 2,000 ms. Therefore, each trial lasted 2,500 ms, which was applied to all pictures, including primes, targets, and fillers. Because participants were given color cues while naming the pictures, the stimulus presentation time was longer compared to several earlier studies (Koeater & Schiller, 2008; Verdonschot et al., 2012).

Subsequently, the experimenter recorded the validity of each trial by noting target language errors, word errors, and voice-key errors. No feedback was provided during the experiment. Throughout all three sessions, target pictures and monosyllabic filler pictures were displayed in a red frame, while prime pictures and disyllabic filler pictures were shown in a green frame. We used a long-lag paradigm for the actual experiment session. Figure 2.1 illustrates how stimuli were presented in this long-lag paradigm.

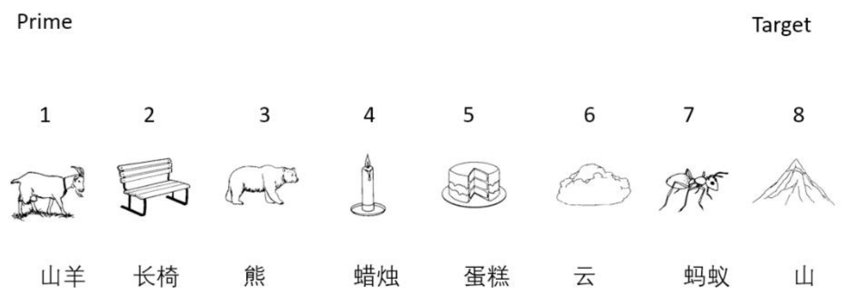


Figure 2.1: Example sequence of a prime-target combination in morpheme-related condition.

Additionally, we presented two sub-lists to each participant during data collection. To counterbalance the order effect, we alternated the sub-list presentation order: Participant 1 saw sub-list 1 first and then sub-list 2, Participant 2 saw sub-list 2 first and then sub-list 1, and so on. Moreover, we had two versions of the main stimuli. Half of the participants used the first version, and the other half used the second version to minimize order and repetition effects.

2.2.5 Electrophysiological recording and data processing

EEG data were collected using *BrainVision Recorder* software (*Version 1.23.0001*) by *Brain Products GmbH*, with an *EasyCap* electrode cap configured according to the standard 10/20 montage (see *Appendix 2.B*). The recordings utilized 32 electrodes (*BioSemi ActiveTwo*) positioned on the scalp according to the standards of the American Electroencephalographic Society (1991). Additionally, the vertical electrooculogram (*VEOG*) was recorded using two external electrodes placed above and below the participant's left eye. In contrast, the horizontal electrooculogram (*HEOG*) was recorded with electrodes positioned at the outer canthus of each eye. Two flat electrodes were attached to the mastoids. CMS and DRL electrodes served as ground references. The EEG signals were subsequently re-referenced offline to the mean of the two mastoids. Data were sampled at 512 Hz, and an offline band-pass filter ranging from 0.01 to 30 Hz was applied, following established procedures from previous studies (Koester & Schiller, 2008; Lensink et al., 2014).

2.3 Data analysis

2.3.1 Behavioral and EEG data exclusion

The two bilingual participants' naming latencies and EEG data were excluded because their native languages were Cantonese and Mandarin. The naming latencies of the other two participants were excluded because of their low accuracy rate. The EEG data for the four other participants were lost due to EEG recording failures and insufficient signal quality. Therefore, thirty-two participants remained for naming latencies analysis, and twenty-eight remained for EEG data analysis in the present paper. For the remaining data, 8.32% of data trials were excluded from further RT analysis, including error trials and trials with reaction times, which deviated more than 2.5 SDs from the mean per participant per condition. For EEG data, 23.15% of data trials were eliminated from the ERP data analysis, including error trials (7.99%) and epochs removed during artifact rejection (15.16%).

2.3.2 Behavioral data analysis

Behavioral data were analyzed using *RStudio Version 4.2.2*. Initially, we calculated descriptive statistics for naming latencies for each condition (see Table 2.3). Subsequently, a single-trial modeling approach was implemented using the lme4 package (Bates & Maechler, 2020). Specifically, we employed a generalized linear mixed-effects model (*GLMM*) with the *glmer ()* function, utilizing a gamma distribution to appropriately model the positively skewed reaction time (RT) data.

Table 2.3: Mean naming latencies (only correct trials included) for each condition.

Condition	Naming latency (ms)	
	Mean	SD
Morpheme-related/first position	779	213
Morpheme-related/second position	809	212
Morpheme-unrelated/first position	776	202
Morpheme-unrelated/second position	798	203

Random effects were chosen in such a way as to avoid over-parameterization and to balance Type-I error and power (Matuschek et al., 2017). We followed a modeling approach where we maintained the simplest possible model structure in light of our main manipulations (Bates & Maechler, 2020; Matuschek et al., 2017; Winter, 2019). Our model included *relatedness* and *position* as fixed factors, *prime semantic category*, *target semantic category*, *prime animacy*, and *target animacy* as co-variates. As participants saw the targets twice, we also included the *sequence* of targets as a co-variate. *Subject* and *item* were the two random effects in the model.

We performed model fit checks by plotting the model residuals against predicted values. We used the *avova ()* function to perform model comparisons and likelihood ratio tests based on Akaike's Information Criterion, AIC (Akaike, 1974), BIC (Neath & Cavanaugh, 2012), and the log-likelihood to establish the best-fitting model for our data. When applicable, we performed *Tukey* corrected post-hoc contrasts to estimate effect sizes using the *emmeans* () function (Lenth et al., 2019).

For naming latencies, the model of the best fit was *RT ~ relatedness + position + sequence + (1|subject) + (1|item)* (see Table 2.C.1 in *Appendix 2.C*). The Subject and Item's random slope could not make the model converge, so they were not included in the current mode. When doing the data analysis, we first introduced the interaction of the relatedness and position in our behavioral data analysis. However, the interaction between relatedness and position did not significantly improve the model after the ANOVA test (with $p > 0.05$) compared to the model without interaction, so we decided to exclude it from the behavioral model. Contrary to findings in other studies, participants were not significantly faster in naming morpheme-related items compared to morphologically unrelated items with $\beta = 0.012$, $SE = 0.010$, $t = 1.159$, $p = 0.247$ (Figure 2.2). Position yielded no significant effect either ($\beta = - 0.051$, $SE = 0.045$, $t = - 1.140$, $p = 0.254$). The final model included the factor sequence, showing a significant effect ($\beta = - 0.059$, $SE = 0.010$, $t = 5.69$, $p < 0.001$). The model did not include the target animacy, prime animacy, target categories, and prime category co-variates due to non-convergence or singular fit or because they did not fit the model significantly.

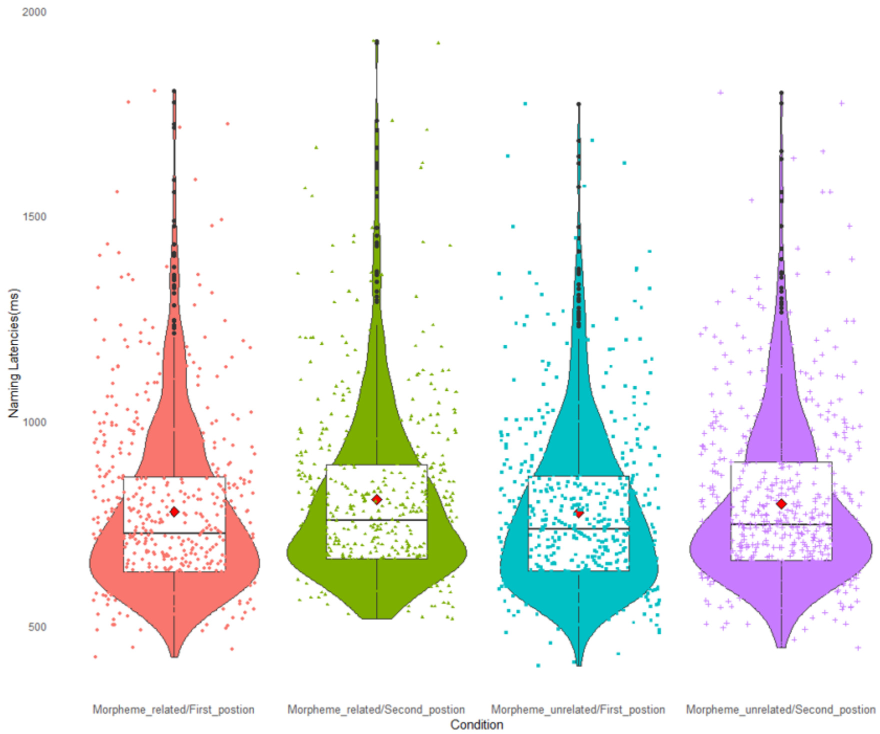


Figure 2.2: Mean naming latencies by condition for the picture naming task (n = 32).

2.3.3 EEG data analysis

EEG data were pre-processed using *Brain Vision Analyzer 2.1 (Brain Products GmbH).* We followed the EEG data preprocessing procedure laid out by Von Grebmer zu Wolfsthurn and colleagues (2012a; 2012b). The process included visual inspection of the signal, re-referencing, and linear derivation for the HEOG and VEOG electrodes, followed by a band-pass filter with a low-pass cutoff at 0.01

Hz and a high-pass cutoff at 30 Hz. Ocular correction and artifact rejection were then performed. Offline re-referencing was conducted to the average of the left and right mastoid electrodes. Subsequently, epochs of corrected trials were generated around stimulus onsets, and baseline correction for each segment was performed using the average EEG activity from the 200ms preceding stimulus onset (Zu Wolfsthurn Sarah von Grebmer et al., 2021a).

After pre-processing and exporting our data, we conducted a cluster-based analysis to tentatively explore regions of interest and potential time windows associated with significant modulations of the EEG signal. We used a permutation test computed with the *permutes* package (Voeten 2019), which included the voltage amplitudes for all data electrodes across the time window from 0ms to 800ms across conditions. This analysis was performed before considering the mean naming latencies of four conditions to minimize the influence of artifacts during data analysis (Von Grebmer zu Wolfsthurn et al., 2021a; 2012b). As shown in Figure 2.3, the output of this test indicated potentially significant modulations of the EEG signal within the time window between 415ms and 575ms. This time window has been previously associated with the N400 component.

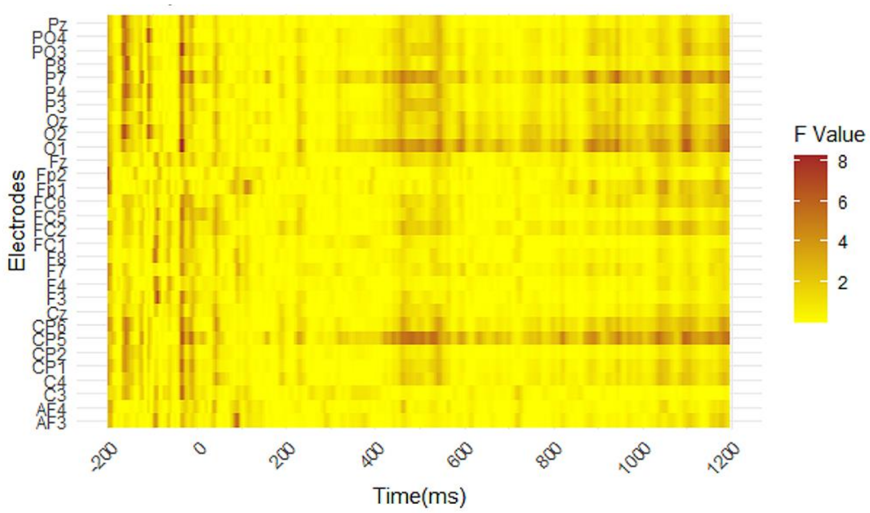


Figure 2.3: Permutation tests across all electrodes. The time window is between 0 and 1,200ms post-stimulus onset. Larger F-values are depicted in darker colors, indicating an increased likelihood of a statistically significant effect of our manipulations on voltage amplitudes.

Next, we divided electrodes into six areas of interest, i.e., anterior-left, anterior-right, central-left, central-right, posterior-left, and posterior-right regions (Verdonschot e al., 2012). Based on the output from the permutation test and previous literature associating central and posterior regions to the N400, we defined our ROIs as the following electrodes: *P3*, *P7*, *CP5*, *O1*, *C4*, *FC2*, *FC6*, and *Fz* (see Figure 2.B.1 in *Appendix 2.B*).

Finally, we employed a single-trial linear mixed models (LMM) approach (Frömer et al., 2018) to enhance the traditional average-type analysis. The traditional method has been criticized for its limitations, including equally weighted observations by condition, participant, and independent factor levels (Von Grebmer zu Wolfsthurn et al., 2021a;

2012b). These assumptions are often compromised due to design factors and during the EEG data pre-processing stages (Von Grebmer zu Wolfsthurn et al., 2021a; 2012b). An alternative method is single-trial linear LMM, which has been endorsed by many researchers since its first application to EEG data in 2011 (Amsel, 2011). This model includes fixed effects and estimates the random variance between subjects and items, known as random effects. It can be applied to datasets with variable effect sizes and unbalanced designs (Baayen et al., 2008; Fröber et al., 2017).

For the single-trial LMM approach, we incorporated all available voltage values for each epoch within the designated time window of 415ms to 575ms, without averaging across segments from the same condition, to preserve by-subject and by-item variance. Fixed effects included relatedness and position, while ROI and sequence were co-variates. Participants and Items were included as random effects in this single-trial analysis. The model-fitting procedure resembled the behavioral analyses. Figure 2.4 displays the mean voltage amplitudes for the entire epoch of 800ms for each condition and for two factors (Relatedness and Position) in the central posterior regions within the selected N400 time window.

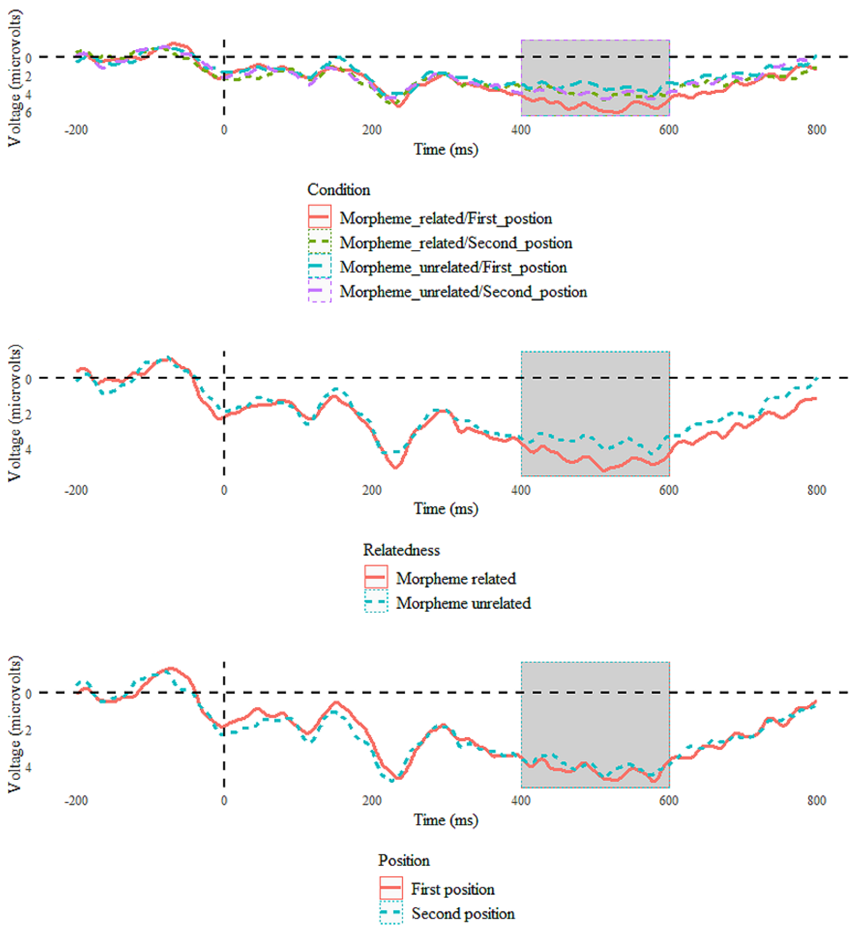


Figure 2.4: Voltage amplitudes by Condition, Relatedness, and Position over time. For channels C4, FC2, FC6, Fz, P3, P7, CP5, and O1 for the picture naming task (n = 28), the time window of the interest is from 415ms - 575ms and is highlighted in grey. Negativity is plotted up.

The model with the best fit was *amplitude ~ relatedness * position + ROI + sequence + (1 | participant) + (1 | item)* (see Table 2.D.1 in *Appendix 2.D*). The *Participant* and *Item*'s random slope could not make the model converge, so they were not included in the current mode. This model yielded a main effect for *relatedness* with by-subject and by-item random intercept. The difference in amplitude between targets in the morphologically related and unrelated conditions was significant (β = - 2.22, *SE* = 0.698, *t* = - 3.185, *p* = 0.00285).

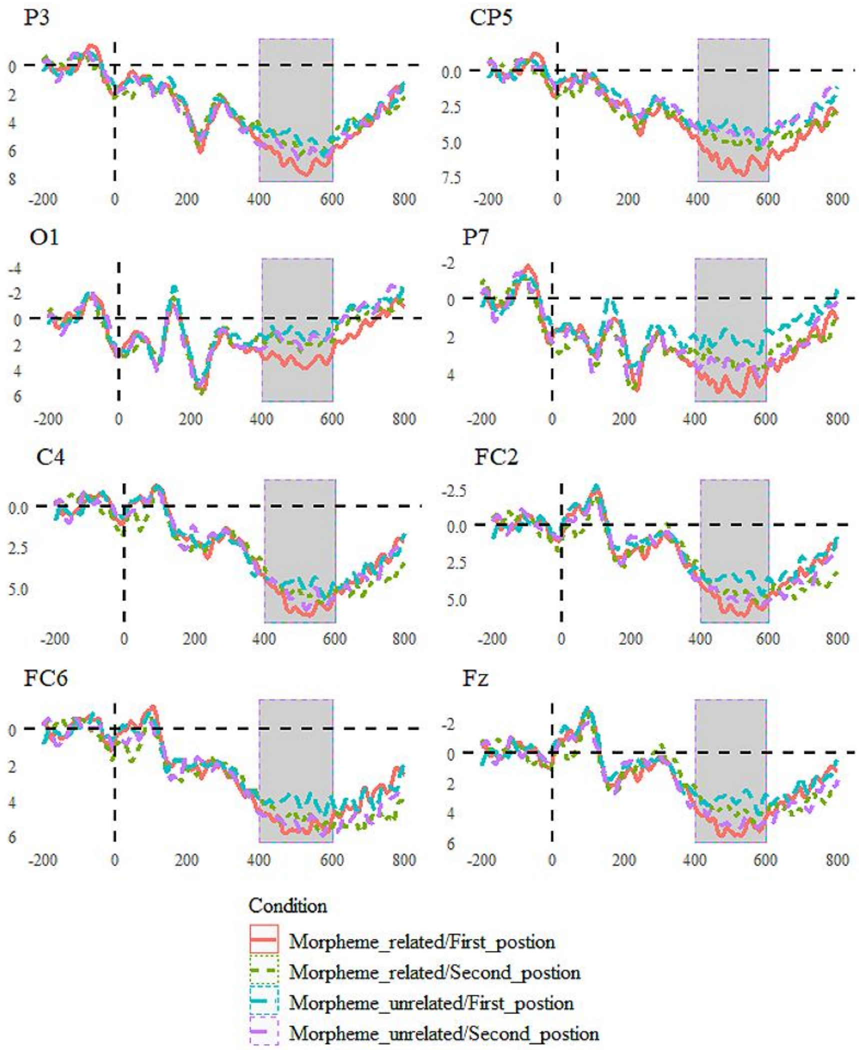


Figure 2.5: Voltage amplitudes by condition over time for each individual electrode. ROIs were included in the analysis of the long-lag picture naming task (n = 28). The time window of interest from 415 ms to 575ms is highlighted in grey. Negativity is plotted up.

The difference in amplitude between the first and second position conditions was also significant (β = - 1.497, *SE* = 0.093, *t* = -16.107, *p* < 0.001). There was an interaction between *relatedness* and *position* (β = - 2.09, *SE* = 0.1356, *t* = - 15.419, *p* < 0.001). The sequence effect was significant with β = 0.414, *SE* = 0.455, *t* = 9.106, and *p* < 0.001. Grand-average ERPs for the four conditions are shown in Figure 2.5 for visualization of the individual channels included in this analysis.

2.4 Discussion

This study investigated morphological encoding in Chinese disyllabic compound words and their storage in our mental lexicon. In a long-lag priming experiment, which included a substantial delay between primes and targets, thirty-six native Mandarin Chinese speakers were asked to name pictures on a screen. Behavioral data and electrophysiological data were collected and then analyzed. The behavioral results were surprising and went against earlier findings in Indo-European languages, showing that the target picture naming was *not* facilitated by morphologically related primes in both first and second positions. Note that the first-position condition did elicit shorter RTs than the second-position condition. This suggests that morphological priming for individual constituents in producing Mandarin Chinese disyllabic compound words at the behavioral level does not occur.

The results reported here have implications for the two models of lexical representation discussed in the Introduction. According to

Levelt and colleagues' (1999) model, compounds are stored in their decomposed form at the lexeme level. Consequently, the targets should be able to retrieve their lexeme representations faster if they had already encountered them in primed pictures (in contrast to unrelated pictures). Contrastingly, according to the single-stage model of lexical representation (1997), compounds are stored in their full listing format at the lexical level. In other words, the whole word form is retrieved when naming the pictures instead of the individual constituents. Thus, when naming individual constituents, participants would have to retrieve these again, which predicts that the naming latencies should be the same when naming morpheme-related prime pictures in contrast to unrelated condition prime pictures. In line with this prediction, the results from the present experiment revealed that the naming latencies between morpheme-related and morpheme-unrelated conditions make no significant difference. Additionally, our results seem to support the results shown by others (Janssen et al.,2008; Bi et al., 2007; Chen & Chen 2006).

However, in the ERP analyses, morpheme-related prime stimuli did elicit a reduced N400 at the central posterior regions when compared to morpheme-unrelated prime stimuli in the 415 ms - 575 ms time window, which is in line with previous results (Koester & Schiller, 2008; Dohmes et al., 2004) at both the temporal and spatial levels. The reduced negativity can be interpreted as an N400 effect, as the N400 is sensitive to morphological processing (Koester & Schiller, 2008; McKinnon et al., 2003). This result suggests that the morphological priming effect occurs at the morphological encoding level and supports a decomposed model. Previous research (Koester & Schiller, 2008; McKinnon et al., 2003) suggested that the outcomes stemmed from a genuine morphological process. Similarly, we claim that the current effects cannot be explained by a semantic or phonological relation between primes and targets. This is because semantic and phonological effects usually do not last beyond the temporal gap between prime and target in a long-lag paradigm (Zwitserlood et al., 2000; Feldman, 2000).

Furthermore, the timing of our ERP priming effects aligns well with the timeline for morphological encoding, as opposed to other cognitive phases like conceptual preparation or lemma retrieval (Indefrey & Levelt, 2004). Indefrey and Levelt proposed that morphological encoding, the primary phase in word form encoding, starts around 330ms post-picture presentation. The initiation of the N400 effect in our study closely mirrors this predicted onset (Indefrey & Levelt, 2004). These projections are based on a response latency of 600ms. Adjusting the onset of morphological encoding to correspond with a response latency of 750ms-800ms in the present study, our observed average reaction time should yield a rough estimate of 413ms-440ms for the initiation of morphological encoding. Interestingly, this estimation is almost identical to the onset of the N400 effects in the present study. Consequently, the N400 effects support the hypothesis that our morphological priming effect occurs and originates at the word-form level.

One question remains, however: how can we account for the divergence between the ERP findings and the absence of RT effects in the present study? In previous work (Janssen et al., 2008; Chen & Chen 2006), a potential explanation was proposed, namely that the absence of a morpheme priming effect in Mandarin Chinese might be due to cross-linguistic differences. The involvement of morphological representations may not be significant in languages characterized by straightforward morphology, such as Mandarin Chinese. Conversely, these representations will likely be more prominent in languages featuring intricate morphology, like Dutch. Another possible explanation might relate to the long-lag paradigm, which may have lacked the power to pick up on any fine-grained modulation in the Chinese morphological encoding process. Potential future research could focus on investigating the implicit RT results. However, EEG data are sensitive enough to capture these differences because ERPs offer a higher temporal resolution that enables more direct observation of cognitive processes, even before or without an explicit response (Kutas & Petten, 1994). EEG can provide more information on the current research questions regardless of the implicit outcome of naming latencies.

In addition, we did not control for transparency in the present study. On the one hand, opaque words are not common in Mandarin; on the other hand, many opaque words are mono-morphemic compounds (like 沙发 /sha1fa1/ "sofa") in Mandarin. However, according to

Koester and Schiller (2008), Zwitserlood et al. (2000; 2002), and Verdonschot et al. (2012), the opaque and transparent words elicited no significant difference in morphological priming. We believe more research could be done regarding this aspect of Mandarin. The results in the present study suggest that morphological priming for constituents is present in the morphological encoding process of Chinese disyllabic compound words, even though the naming latencies between morphologically related and unrelated conditions yielded no difference. However, the morphological involvement is not as strong as in Dutch. Chinese disyllabic compounds are stored with two separate constituents instead of the whole word representation. This suggestion aligns with a decomposed storage model (Levelt et al., 1999; Bien et al., 2005; Roelofs, 1996) rather than a single-stage model (Caramazza, 1997; Janssen, 2008; Chen & Chen 2006). Then, given the presence of morphological encoding, the morphological priming effect originates in the word form level according to the temporal course of the N400 effect.

The repeated picture presentation across two blocks lead to reduced picture naming latencies in this experiment, according to the statistical results in both naming latencies and ERP analyses. This effect is consistent and comparable to previous findings (Zwitserlood et al., 2000; 2002). Importantly, the block effect did not interact with manipulating the variable of primary interest, i.e., Prime type. Therefore, it is suggested that the general facilitation across blocks is independent of linguistics processes. This repetition effect may, for example, reflect the more efficient visual processing or recognition of the pictures.

Regarding the position effect, the present results from RTs showed that the naming latencies for the first position were shorter than the second, which aligns with our expectations, even though this difference is not significant. In the ERP results, the significant effect of position and interaction between relatedness and position showed that the first and second constituents were processed differently. A larger reduced N400 was found for the first than the second position. These findings support the claim that both positions have priming effects, but the priming effect is larger for the first position. The significant difference between the first and the second constituents supported the serial planning assumption and the modifier constituent facilitation effect. However, the challenge of understanding how the head constituent influences compound production, as observed in (Bien et al., 2005), remains unsolved. This influence could be either facilitatory or inhibitory in nature. Further research is necessary to shed light on this issue.

In addition, there were both free and bound morphemes in the stimuli of the present study. Bound and free morphemes could cause differences in the processing of Chinese compounds. There might be a possibility that compounds with different morpheme-type combinations are processed differently, which is worth exploring. This could also be why our behavioral data did not have an effect and only the first position morphological priming effect was found in the ERP analysis. We did a post-hoc analysis regarding two types of morphemes, but no significant statistics results were found ($p > 0.05$). This might be because we had balanced the items for two morpheme types. More research can be done on this aspect in the future. To conclude, in the present long-lag priming experiment, picture naming was behaviorally not facilitated by the production of morphologically related compound words. However, morphological priming for individual constituents was present from an N400 effect in the central posterior region in the ERP analysis, especially concerning the priming of the first constituent. The N400 effect presented in the process appears to originate at a word form level corresponding to the temporal estimate of morphological encoding, and two individual constituents are both activated and processed during morphological encoding.

Acknowledgments

We thank Sarah Von Grebmer zu Wolfsthurn for her help during data collection and analysis.

Data Availability

All anonymized data will be held in a public repository at https://osf.io/nqhy4/

Funding

JQW: China Scholarship Council (CSC) full PhD programme (48months) scholarship (nr:202008130142). The funder did not play any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors have declared that no competing interests exist.

Appendix

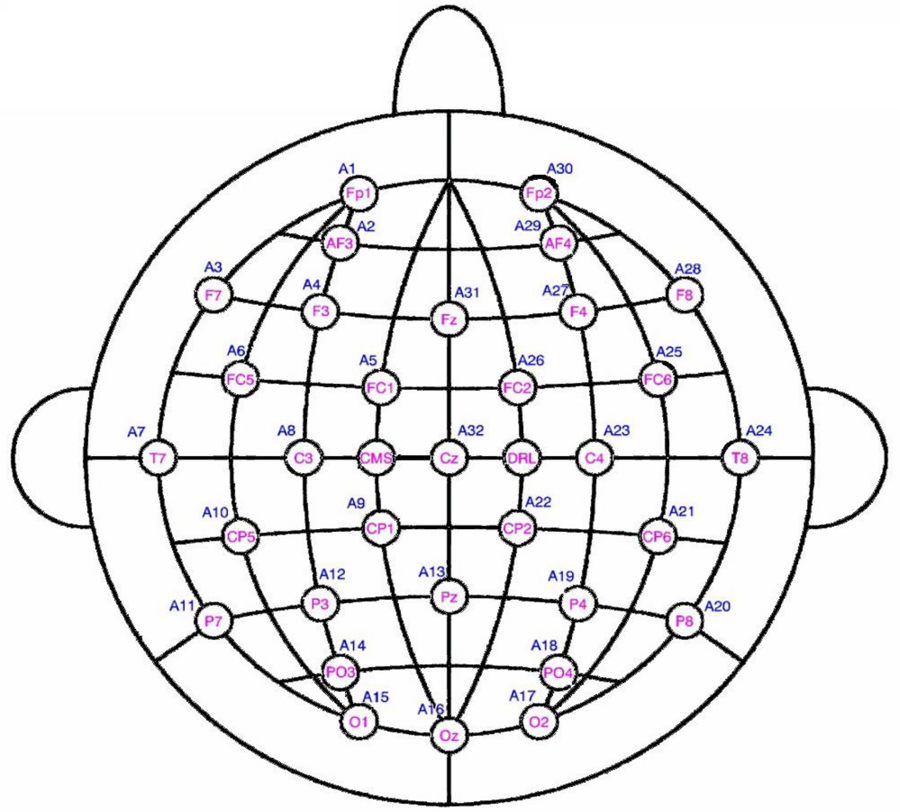
2.A Experimental stimuli.

Target	Morpheme-related distractor	Morpheme-unrelated distractor
山	山羊	飞机
鸟	鸟巢	蜡烛
脚	脚印	熊猫
牙	牙刷	南瓜
竹	竹筒	蜂蜜
手	手机	拼图
眼	眼罩	钢琴
鱼	鱼竿	树桩
鹿	鹿角	领带
灯	灯塔	孔雀
信	信封	香蕉
雨	雨伞	皇冠
雪	雪人	青蛙
月	月饼	翅膀
旗	旗袍	秋千
书	书架	蘑菇
轮	轮胎	彩虹
鼠	鼠标	钮扣
石	石榴	樱桃
草	草莓	电话
车	风车	羽毛
刀	剪刀	玉米
针	别针	围巾

脑	电脑	钥匙
杯	奖杯	毛巾
盘	键盘	苹果
线	电线	邮票
球	气球	老虎
船	飞船	骆驼
花	烟花	喷泉
包	面包	天鹅
梯	滑梯	洋葱
绳	跳绳	戒指
扇	电扇	磁带
箭	火箭	菠萝
镜	泳镜	乌龟
龙	恐龙	口红
桶	马桶	面具
钉	耳钉	烟斗
牛	蜗牛	卡车

2.B EEG electrode montage

Figure 2.B.1: EEG montage. 10/20 system - 32 channel montage from BioSemi, including CMS and DRL but excluding external channels (www.biosemi.com/headcap.htm).



2.C Model parameters: response times

Table 2.C.1: Specification of best fit model for RTs (ms) for n = 32.

Formula: RT ~ Relatedness + Position + Sequence + (1 | Subject) + (1 | Item)				
Fixed effects	Estimate	95%CI	t-value	p-value
(Intercept)	1.32	[1.075, 1.605]	25.92	<0.001 ***
Relatedness: Morpheme unrelated	0.012	[-0.203, 0.224]	1.16	0.247
Position: Second position	-0.051	[-0.28, 0.235]	-1.14	0.254
Sequence	-0.059	[-0.138, 0.2988]	5.69	<0.001 ***
Random effects				
σ^2	0.050			
τ00Item	0.005			
τ00Subject	0.006			
ICC	0.098			
NSubject	32			
NItem	40			
Observations	2560			
Marginal R^2	0.025			
Conditional R^2	0.200			

2.D Model parameters: N400

Table 2.D.1: Specification of model of best fit for Voltage Amplitudes (microvolts) for n = 28.

Formula: Amplitude ~ Relatedness * Position + ROI + Sequence + (1 | Participant) + (1 | Item)				
Fixed effects	Estimate	95%CI	t-value	p-value
(Intercept)	6.05	[3.696, 8.179]	5.143	< 0.001***
Relatedness: morpheme unrelated	-2.224	[-3.570, -0.805]	-3.185	0.000285
Position: second position	-1.497	[-1.689, -1.309]	-16.107	< 0.001***
ROI: left posterior	-1.209	[-1.307, -1.109]	-23.150	<0.001***
Sequence	0.414	[0.325, 0.501]	9.106	<0.001 ***
Relatedness Morpheme unrelated *Position Second	2.092	[1.840, 2.360]	15.419	<0.001 ***
Random effects				
σ^2	311.798			
τ00Item	4.809			
τ00Subject	31.862			
ICC	0.014			
NSubject	28			
NItem	40			
Observations	610080			
Marginal R^2	0.003			
Conditional R^2	0.108			

Chapter 3 Word and morpheme frequency effects in naming Mandarin Chinese compounds: More than a replication

This article is published as: Wang, J., Schiller, N. O., & Verdonschot, R. G. (2024). Word and morpheme frequency effects in naming Mandarin Chinese compounds: More than a replication. Brain and Language, 259, 105496. https://doi.org/10.1016/j.bandl.2024.105496.

Abstract:

The question whether compound words are stored in our mental lexicon in a decomposed or full-listing way prompted Janssen and colleagues (2008) to investigate the representation of compounds using word and morpheme frequencies manipulations. Our study replicated their study using a new set of stimuli from a spoken corpus and incorporating EEG data for a more detailed investigation. In the current study, despite ERP analyses revealing no word frequency or morpheme frequency effects across conditions, behavioral outcomes indicated that Mandarin compounds are not sensitive to word frequency. Instead, the response times highlighted a morpheme frequency effect in naming Mandarin compounds, which contrasted with the findings of Janssen and colleagues. These findings challenge the full-listing model and instead support the decompositional model.

Keywords: *Language production; Word frequency; Morpheme frequency; Compound representation; Picture naming; Mandarin Chinese*

3.1 Introduction

Theories of language production (Caramazza, 1997; Dell, 1986; Levelt et al., 1999) have specific assumptions about how words are organized and represented in the mental lexicon and suggest that word production occurs in several stages: conceptual preparation, lexical access, phonological encoding, and articulation. During speech production, such as when naming a picture, the conceptual representation of the intended object becomes active. This activation then extends to the lexical representations associated with this concept. Subsequently, phonological information is retrieved, which involves the encoding of word forms and is ultimately used for articulation by initiating the relevant speech gestures.

Many speech production models include a role for morphology (Koester & Schiller, 2008, 2011; Levelt et al., 1999; Roelofs, 1996; Zwitserlood et al., 2000, 2002). For example, in Levelt and colleagues' word production theory (Levelt et al., 1999), which largely draws on results from the Dutch language, morphological encoding is the initial stage of word-form encoding. This stage involves the construction of words and defining their internal structures at the word form level. This has raised the question of how morphologically complex words, including derived words (e.g., "happiness"), inflected words (e.g., "running"), and compound words (e.g., "birdhouse"), are represented in our mental lexicon and how they are comprehended and produced. The present study aimed to investigate the representation of Mandarin compound production by reviewing the previous literature on complex words in both language production and comprehension.

3.1.1 The representation of morphologically complex words

3.1.1.1 Production of complex words

For language *production* of complex words typically a distinction is made between decomposition and full-listing models (Caramazza, 1997; Janssen et al., 2008; Janssen et al., 2014) with *hybrid* models (i.e., some complex words have full-listing storage, but other complex words are decomposed or follow a rule-based system) being somewhat scarcer. Levelt et al.'s model (1999; p 25) states that compound words (e.g., "birdhouse") have a single lemma which then in turn activates two lexemes (<BIRD> + <HOUSE>). Note that Levelt et al.'s model (1999; p 27) may be considered "hybrid" in the sense that it suggests that certain compound words are "degenerate" in production, meaning they are not decomposed at the form level. For example, the Dutch word *aardappel* "potato" appears semantically composed of *aard* "earth" and *appel* "apple," but it is produced as a single unit: *aar-dap-pel* (not *aard-ap-pel*). In contrast, a non-degenerate opaque compound like *oogappel* which is *oog* "eye" + *appel* "apple," meaning "apple of my eye" (a term of endearment for children), is decomposed at the form level, as evidenced by its pronunciation as *oog-ap-pel,* rather than *oo-gap-pel*. Inflected words (e.g., "escorted") in Levelt et al.'s model (1999) have a marking at the lemma level indicating tense (e.g., escort + PAST will activate the morphemes <ESCORT>+<ED>). Note that other theories, such as the words and rules theory by Pinker and Ullman (2002), similarly state that regular verb forms can be generated by a rule (i.e., a unification operation applied to a specific morpheme), just as how phrases and sentences are formed. When an irregular verb form happens to be stored (e.g., "drank"), it prevents a rule from applying (e.g., blocking *drinked), but anywhere else (by default) the rule applies. In Levelt et al.'s model (1999), the case is somewhat more difficult for complex derivational morphology especially when words would change syntactic class (e.g., the adjective "weak" + NESS forming the noun "weakness") for which Levelt et al. (1999) propose that these most likely are lemmas in their own right (i.e. "weakly" and "weakness" are separate lemmas).

3.1.1.2 Comprehension of complex words

Regarding language comprehension, which is more amply investigated, a similar division between decomposition can be made

(Longtin & Meunier, 2005; Rastle & Davis, 2008; Koester et al., 2004, 2009), full-listing (e.g., Butterworth 1983; Norris & McQueen, 2008) and hybrid models (Caramazza et al., 1988; Frauenfelder & Schreuder, 1991). For example, inflectional and derivational processes have received ample attention in the comprehension literature (Bozic & Marslen-Wilson, 2010; Marslen-Wilson et al., 1994; Penke et al., 1997; Rodriguez-Fornells et al., 2001; Smolka et al., 2015). For inflectional processes, the "morphological violation paradigm" in which a verb takes an incorrect form (e.g., the past participle **getanz-en* meaning "danced" in German, which should be *getanz-t*) has shown that differential EEG patterns occur for regular vs. irregular verbs supporting a hybrid model in which regular/irregular verbs are processed differently (Penke et al., 1997). Koester and colleagues (2009) investigated the time course of semantic integration in auditory compound word processing and found that the lexical-semantic integration of compound constituents occurs incrementally, supporting the decompositional hypothesis.

Others, for instance, Winther Balling and Baayen (2008), have suggested that there are two key moments during the auditory recognition of a complex word where its recognition likelihood changes significantly. First, when unrelated words are ruled out at the (typical) "uniqueness point" in which only one option remains in the competition for word recognition. For instance, for a simple word such as "candle," the uniqueness point might occur at the sound /d/ because up to /kænd/ "cand," it could be confused with words like "candy," but once the /l/

is added, the word becomes clearly distinguishable as “candle.” However, for complex words such as derivations, there might be a continuation after the base word is recognized. For example, the word “hope” might receive its first uniqueness point at “p” where it is distinguished from words such as “hole” or “home.” However, when acting as a part of a complex word, it could receive another uniqueness point, for example, at the “f” for “hopeful” where it is distinguished from other morphological stem-related words such as “hopeless.”

Others working on visual word recognition have posited that there might be an early level of representation where (seemingly) complex words are broken down based on their morpho-orthographic features (Rastle et al., 2004). For example, a semantically transparent morphological relationship of a prime with the target (e.g., cleaner-CLEAN) would give rise to facilitation. However, Rastle et al. (2004) also found that when encountering a word like “corn,” a prime such as “corner” would facilitate lexical decision times as well, even though the two words were not morphologically related, but see Baayen et al. (2011) for different views. This suggested a form of morphological decomposition that functioned differently from the semantic-based decomposition involved in the early stages of visual (complex) word recognition.

3.1.2 The representation of compound words

Returning to the issue of the processing of compound words in language production, as stated earlier, the most prominent theoretical models are either decompositional, full-listing, or hybrid. The decompositional model holds the view that compounds are represented in terms of their constituents unless they are degenerate (Levelt et al., 1999), while the full-listing hypothesis suggests that the whole-word forms are only fully listed in our mental lexicon (Butterworth, 1983; Caramazza, 1997; Dell, 1986). Dual-route models combine elements of both approaches, proposing that transparent compounds are processed by breaking them down into their components while also allowing for parallel access to the whole-form representation. However, for opaque compounds, particularly those with high frequency, the whole-form access is assumed to be the dominant mechanism (MacGregor & Shtyrov, 2013). The present study focused on the decompositional and full listing hypotheses and hence summarized the relevant literature of these two models in the following sections.

3.1.2.1 The decompositional model

The presence of morphologically decomposed entries in the form of lexicon underlying speech production was addressed by manipulating word frequency by Roelofs (1996). He demonstrated the effect of constituent frequency on lexical access during speech production planning by using Dutch compounds. Implicit priming experiments were conducted, including homogeneous blocks where the stimuli shared a common form and heterogeneous conditions where they did not, to investigate whether the speech production system could plan non-initial constituents of a word before the initial ones. These experiments revealed the task's sensitivity to morphological planning. A more pronounced facilitatory effect was observed when the initial syllable, constituting a morpheme, was shared (e.g., *nasmaak* "after-taste"-*nagalm* "reverberation"-*najaar* "autumn"), compared to producing disyllabic simple words in which the same overlapping part "na" constituted a syllable but not a morpheme (e.g., *nagel* "nail"-*natie* "nation"-*nader* "further"). These results supported the idea that component morphemes of compound words served as planning units in speech production, confirming the decomposition assumption.

Likewise, in a study conducted by Bien and colleagues (2005), four experiments employing a position-response association task were carried out to explore the influence of frequency information on Dutch compound word production. They independently manipulated the frequencies of the first and second constituents and the frequency of the compound itself. Compound production latencies demonstrated notable variability based on factorial contrasts in the frequencies of both constituent morphemes rather than being influenced by a factorial contrast in compound frequency, providing further reinforcement for decompositional models of speech production.

Other evidence came from the studies (Kaczer et al., 2015; Koester & Schiller, 2008, 2011; Lensink et al., 2014; Verdonschot et al., 2012; Zwitserlood et al., 2000, 2002; Wang et al., 2024) that employed the long-lag priming paradigm to investigate morphological processing, showing that morphological priming remained effective even with many intervening trials. These studies above suggested that priming in those instances occurred at a distinct morphological level rather than at a phonological or semantic level. This finding was corroborated by further investigations. For example, in a study by Koester and Schiller (2008), they found that mere form overlap, as in prime-target pairs like *jasmijn* “jasmine”- *jas* “coat,” did not facilitate picture naming, indicating that morphological priming represents a distinct form of priming separate from form-identity priming. While these mentioned findings align with decompositional hypotheses in compound production, it is essential to note that there are other studies that favor a full-listing model.

3.1.2.2 The full-listing model

The evidence of the full-listing model comes from Janssen and colleagues (2008). In their study, they tried to answer the question of the representation of compound words in our mental lexicon by manipulating the compound and constituent frequencies of Mandarin compound words. Native Mandarin speakers were asked to name objects in a picture naming task. There were three conditions: H(h), L(h), and L(l). High (H) or low (L) compound word frequency was denoted by the first upper-case letter, while high (h) or low (l) constituent morpheme frequency was indicated by the second lower-case letter in parenthesis. Their analysis revealed that only the compound frequency influenced naming latencies in Mandarin compound production, providing support for full-listing hypothesis. Another study by the same lab (Janssen et al., 2014) showed that this pattern (effect of compound but not constituent frequency) extended to naming pictures using English (i.e., “oillamp” is a low-frequency compound, but its constituents are highly frequent; “bobsled” is a low-frequency compound with low-frequency constituents). However, when a lexical decision task (LDT) was used (Experiment 2), constituent frequency effects arose. Janssen et al. stated that when a semantic input representation drove word retrieval, constituent effects were absent, but when the signal in the LDT (i.e., a visually or auditorily presented word) contained the constituents, they would be accessed separately, and in that case constituent frequency had an effect.

Additional support for the full-listing hypothesis comes from a study by Bi and colleagues (2007), which investigated two Chinese aphasic patients with lexical access difficulties in oral and written production in naming disyllabic compound pictures. Their findings indicated that the frequency of the compound word, rather than the constituent, affected the production performance of both patients in Mandarin compound production.

In another investigation by Chen and Chen (2006), Mandarin Chinese speakers participated in an implicit priming task (Meyer, 1991), naming compound words in a response-association task. Their goal was to explore whether morphological encoding played a role in producing Chinese disyllabic transparent compound words. Their results showed that naming latencies were not sensitive to the compound’s constituent frequency.

The studies above provided inconsistent results and the debate on how compound words are organized and produced remains unclear. Additionally, there may be some methodological issues with earlier studies. For instance, Janssen and colleagues' study (2008) based their stimuli on the information found in the Modern Chinese Frequency Dictionary (MCFD) (Language Teaching and Research Institute of Beijing Language Institute, 1986). However, the MCFD was published forty years ago and primarily based on written texts rather than speech production, which could potentially influence the evaluation of stimuli's frequency. For example, 水池 /shui3chi4/ "basin" was deemed a high-frequency compound word in the MCFD, but it was not in the SUBTLEX-CH corpus (Cai & Brysbaert, 2010), which was based on speech (e.g., movie subtitles).

3.1.3 The current study

As we deemed the approach to exploring the question of Mandarin compounds representation and production presented in the study of Janssen and colleagues suitable for the purpose of the current study, we opted to replicate their design framework and extend their study by introducing a new set of stimuli based on the SUBTLEX-CH Mandarin speech production corpus (Cai & Brysbaert, 2010). Additionally, we utilized EEG methodology to examine the temporal dynamics of compound production in greater detail. Event-related potentials (ERPs) offer a higher temporal resolution in contrast to reaction times, allowing for more direct observation of cognitive processes, even before an explicit response is made (Kutas & Federmeier, 2011; Kutas & Van Petten, 1994). Regarding the current issue, ERPs can serve as a direct method to investigate the frequency effect for constituents and the whole word due to words with a higher frequency tend to trigger N400s of reduced amplitude compared to words with a lower frequency when all other factors remain consistent (Kutas & Federmeier, 2009, 2011; Rugg, 1990; Van Petten & Kutas, 1990).

The prediction of the present study is to have similar reaction times when comparing H(h) to L(h) condition and shorter naming latencies for L(h) condition than for L(l) condition. Moreover, a reduced N400 amplitude was predicted in L(h) and L(l) conditions; no reduced N400 amplitude was predicted in H(h) and L(h) conditions.

3.2 Methodology

3.2.1 Participants

Thirty-two native Mandarin Chinese speakers, aged 20 to 32 years (mean age: 26.42, SD: ± 2.71), including six males, were recruited from Leiden University. All participants were from Mandarin-speaking provinces in China and spoke Mandarin as their mother tongue. Participants who had been living in the Netherlands for less than two years were included, while those who had been residing in the Netherlands for longer were excluded due to potentially higher proficiency in English and Dutch. All participants had normal or corrected-to-normal vision and received monetary compensation for their participation. At the time of testing, none reported color blindness, learning disorders, hearing or visual impairments, or psychological or neurological conditions. Participants read an information sheet and provided informed consent by signing a consent form before the study began.

3.2.2 Materials

In the process of material design for the present replication, our approach followed the methodology outlined by Janssen and colleagues (2008). The design closely mirrored the structure proposed by Jescheniak and Levelt (1994). We curated three sets of images: (a) L(l) pictures featuring names comprised of low-frequency compound words and composed of low-frequency constituents; (b) L(h) pictures, including names formed by low-frequency compound words but consisting of high-frequency constituents; and (c) H(h) pictures, characterized by names with high-frequency compound words and composed of high-frequency constituents. Three conditions of three different frequency distributions were created for the present experiment. Although most of the compound's constituents were noun-noun pairs, there were occasional instances where a verb formed a constituent (e.g., 扫帚 /sao4zhou3/, meaning "broom," where 扫 means "to sweep"). However, these cases were rare and not expected to influence the results.

The average cumulative frequency of each constituent was calculated in the present study. For example, the Chinese compound word 山羊 /shan1yang2/ "goat" was composed of two constituents 山 /shan1/ "mountain" and 羊 /yang2/ "sheep." The compound word frequency referred to the occurrence of 山羊 /shan1yang2/ "goat." Cumulative morpheme frequency was the combined frequency of all homophonic constituents, disregarding their written form. This involved summing the frequencies of all homophones for each constituent and then averaging them to obtain the mean cumulative frequency in this study. For instance, when calculating the cumulative frequency of 羊 /yang2/ "sheep" within 山羊 /shan1yang2/, all its homophones like 阳 /yang2/ "sun" and 洋 /yang2/ "foreign," etc., were considered. The resulting morpheme frequency was derived from averaging the cumulative frequencies of both constituents, namely 山 /shan2/ "mountain" and 羊 /yang2/ "sheep." Based on the rationale outlined by Janssen and colleagues (2008), the adoption of cumulative frequency was justified by the fact that homophonic morphemes are condensed into a single lexeme node within Levelt's model (1999) under examination in this study. The decision to utilize average cumulative frequency stemmed from its strong correlation with individual constituent frequencies. This approach offered a reliable estimation of the impact of constituent frequency on compound production.

We calculated the word and morpheme frequencies in the present study based on the SUBTLEX-CH corpus (Cai & Brysbaert, 2010). We selected 28 disyllabic noun compounds for each condition in the stimuli set and incorporated 36 filler pictures, which were also disyllabic noun compounds. Twenty-eight black-and-white line pictures labeled as L(h) were carefully chosen, and each was paired with an L(l) picture, ensuring a match on average whole-word frequency (t (54) = -1.22, p = 0.23). Additionally, each L(h) picture was paired with an H(h) picture, ensuring a match in average morpheme frequency (t (54) < 1). The visual complexity of all target pictures was controlled (t (54) < 1) in the present study. There was a significant difference in the average morpheme frequencies between the L(h) and L(l) pictures (t (54) = -2.02, p = 0.04). The word frequencies of L(h) pictures were found to be lower than the whole-word frequency of H(h) pictures with t (54) = -10.43, p < .0001, and the word frequencies of L(l) pictures were found to be lower than the whole-word frequency of H(h) pictures as well, with t (54) = - 9.30, p < .0001. See Table 3.1 below for the detailed information.

Table 3.1: Mean frequency distribution of the picture names in the experiment.

Condition	Example	English translation	Mean compound frequency (Zipf)	Mean constituent frequency (Zipf)	Mean left constituent frequency (Zipf)	Mean right constituent frequency (Zipf)

H(h)	电话 (dian4hua4)	telephone (electricity+speech)	3.62	2.52	2.39	2.65
L(h)	长椅 (chang2yi2)	bench (long+chair)	2.12	2.54	2.56	2.51
L(l)	蜡烛 (la4zhu2)	candle (wax+candle)	1.93	2.43	2.43	2.42

The fillers were used both as warm-up stimuli at the beginning of each block and as fillers throughout the experiment. We did not control the frequencies of fillers because we did not analyze the reaction time to filler pictures. The complete list of stimuli is listed in *Appendix 3.A*.

3.2.3 Design

A 3 by 3 factorial within-subject design was adopted in this experiment, with frequency distribution and repetition level as fixed factors. A picture-naming task was employed and a pseudo-randomized design per participant was implemented. Pictures with the same category and the same phonological onset would not be subsequently presented to avoid priming effects of the same category and phonological overlap.

Participants were presented with 120 pictures, comprising 84 experimental pictures and 36 fillers. The proper experiment consisted of three blocks and each picture appeared once per block. We introduced repetition level as an experimental factor (three repetitions), and each participant encountered each picture three times during the experiment. Consequently, there were 360 trials in total for each participant in the entire real experimental session.

3.2.4 Procedure

The experiment was designed and controlled using *E-prime 3.0* (Psychology Software Tools) and was conducted in a soundproof booth. Participants were seated in front of a computer in a dimly lit room. A microphone connected to a Chronos response device containing a voice key was used to record naming latencies.

The experiment comprised three phases. The initial familiarization phase involved familiarizing participants with the pictures for the subsequent experiment. Each trial began with a 500 ms presentation of a fixation cross. Subsequently, the picture appeared for 2,000 ms, followed by the display of its name underneath it after 1,000 ms. Upon seeing its name, participants were instructed to name the picture verbally, and a new trial commenced after a delay of 1,000 ms.

The following practical part involved participants practicing the experimental task for all pictures, while the third part constituted the actual experiment. The trial structure for the second and third parts was identical. Participants were first presented with a 700 ms fixation cross, followed by the picture display for 1,500 ms or until the participant made a vocal response. Subsequently, there was a blank of 2,000 ms before the start of the subsequent trial (see Figure 3.1).

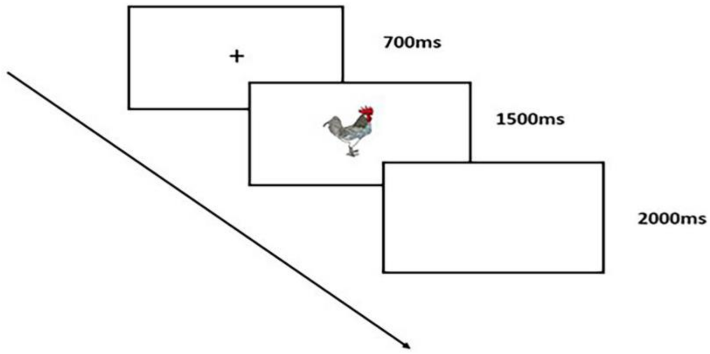


Figure 3.1: A trial sequence for the picture-naming task.

The experimental session consisted of three repetitions with 120 trials in each repetition. There was a break within each repetition. Therefore, the entire experimental session was partitioned into six blocks. The experimenter documented the validity of each trial by noting target language errors, word errors, and voice-key errors. Participants did not receive feedback during the experiment.

3.2.5 Electrophysiological recording and data processing

EEG data were collected using *Brain Vision Recorder* software (Version 1.23.0001) by Brain Products GmbH. An EasyCap electrode cap was employed following the standard 10/20 montage (see Figure 3.B.1 in *Appendix 3.B*). We recorded EEG data from 32 electrodes (*BioSemi Active Two*) placed on the scalp according to the American Electroencephalographic Society standards (1991). The vertical electrooculogram (*VEOG*) was recorded from two external facial electrodes placed above and below the participant's left eye, and the horizontal electrooculogram (*HEOG*) was recorded from two electrodes at the outer canthus of each eye. Additionally, two flat electrodes were positioned at the mastoids. *CMS* and *DRL* electrodes served as ground reference. The EEG signal was later re-referenced offline using the mean of the two mastoids. Data were sampled at a rate of 512 Hz from DC to 102.4Hz (analogue anti-aliasing filter frequency at 1/5th of the sampling rate). The voltage amplitudes were measured approximately every 1.96 ms. A band-pass filter of 0.01-30 Hz was applied offline, following procedures outlined in previous studies (Koester & Schiller, 2008; Lensink et al., 2014).

3.3 Data analysis and results

3.3.1 Behavioral and EEG data exclusion

The data from three participants were excluded from the analysis due to their high error rates in naming pictures and the high rate of artifacts in their EEG data. Error trials (3.06%) and outlier trials (2.75%) with reaction times that deviated more than 2.5 SDs from the mean per participant per condition were eliminated. Eleven items (12.74%) were removed because participants consistently made mistakes in naming these pictures during the experiment. In total, 27.65% of data trials were excluded from further RT analysis. For EEG data, the data for three participants and eleven items were removed. Error trials and outliers were also eliminated from the ERP data analysis based on behavioral data. Artifact rejection (22.65%) was administered during the processing stage.

3.3.2 Behavioral data analysis

Behavioral data were analyzed using *RStudio* Version 4.2.2. We first calculated descriptive statistics for naming latencies for each condition (see Table 3.2). Then, we used a single-trial modeling approach applying the *lme4* package. We employed a generalized linear mixed effect model (GLMM) using the *glmer()* function with a gamma distribution to model positively skewed RT data.

Table 3.2: Mean naming latencies (only correct trials included) for each condition and each repetition level (n = 29).

Condition	Naming latencies (ms) per repetition				
	First	Second	Third	Mean	SD
H(h)	660	647	656	655	129
L(h)	667	635	648	650	129
L(l)	686	654	659	666	135

To prevent over-parameterization and strike a balance between Type-I error and power, we employed a strategy for selecting random effects that prioritized simplicity in our model structure given the primary manipulation (Von Grebmer zu Wolfsthurn et al., 2021a, 2021b). In our data analysis, we followed the approach advocated by Matuschek and colleagues (Matuschek et al., 2017), which emphasizes that model selection should be guided by the underlying data. This approach involves incremental model building, with selection taking place at each step. Model comparisons and likelihood ratio tests were performed using the *anova()* function based on Akaike's Information Criterion (*AIC*) (Akaike, 1974), Bayesian Information Criterion (*BIC*) (Neath & Cavanaugh, 2012), and log-likelihood for model comparisons on each step to see whether the newly added factor had improved the model significantly. Where applicable, Tukey-corrected post-hoc contrasts were executed using the *emmeans()* function. In the model, the fixed effects included *Condition* and *Repetition*. *Subject* and *Item* were introduced as random effects in this single-trial analysis. Additionally, we incorporated an interaction effect between *Condition* and *Repetition*.

For naming latencies, the model with the best fit was *RT ~ Condition + Repetition + (1 + Repetition | Subject) + (1 + Condition | Subject) + (1 | Item)* (see Table 3.C.1 in *Appendix 3.C*). The results showed that the naming latencies of H(h) and L(h) conditions were not significantly different with β = 4.85, *SE* = 4.96, *z* = 0.98, *p* = 0.59; the conditions of H(h) and L(l) yielded a significant difference in naming latencies with β = - 12.38, *SE* = 4.33, *z* = -2,86, *p* = 0.01; L(h) and L(l) conditions elicited significantly different RTs with β = - 17.23, *SE* = 5.88, *z* = - 2.93, *p* = 0.01. No interaction effect between *Condition* and *Repetition* was observed (*F* = 1.53, *p* = 0.19), and therefore the interaction was removed from this model.

3.3.3 EEG data analysis

EEG data were preprocessed using *Brain Vision Analyzer* 2.1 (Brain Products GmbH), following the guidelines on its website (https://www.brainproducts.com/downloads/analyzer). The preprocessing pipeline involved several steps: initial visual inspection of the signal, re-referencing, and linear derivation for HEOG and VEOG electrodes. The offline recordings were re-referenced to the average of the left and right mastoid electrodes, followed by filtering with a low-pass filter set at 0.01 Hz and a high-pass filter set at 30 Hz. Subsequently, ocular correction and artifact rejection procedures were applied. Signal segmentations were explicitly applied to correct trials, creating epochs centered around stimulus onsets to investigate voltage amplitudes for the targeted event-related potential (ERP) component. Segments flagged as problematic during artifact rejection were omitted from further analysis. Baseline correction was implemented for each segment by utilizing the average EEG activity in the 200 ms preceding stimulus onset (Von Grebmer zu Wolfsthurn et al., 2021a, 2021b).

After preprocessing, the data were exported into *RStudio* for statistical analyses. We first selected the Regions of Interest (ROIs) and defined the time windows for analysis (Von Grebmer zu Wolfsthurn et al., 2021a, 2021b). To tentatively explore the locus of the effect of frequency, we conducted a permutation test using the permutes package (Voeten, 2019). This test analyzed voltage amplitudes from all electrodes within a time window between 0 and 1,200 ms across three conditions and three repetition levels. Larger F-values, indicated by darker colors, suggest a higher likelihood of a statistically significant effect of the manipulations on voltage amplitudes. As shown in Figure 3.2, the test results highlighted potential modulations in frontal-central areas between 300 and 500 ms post-stimulus onset.

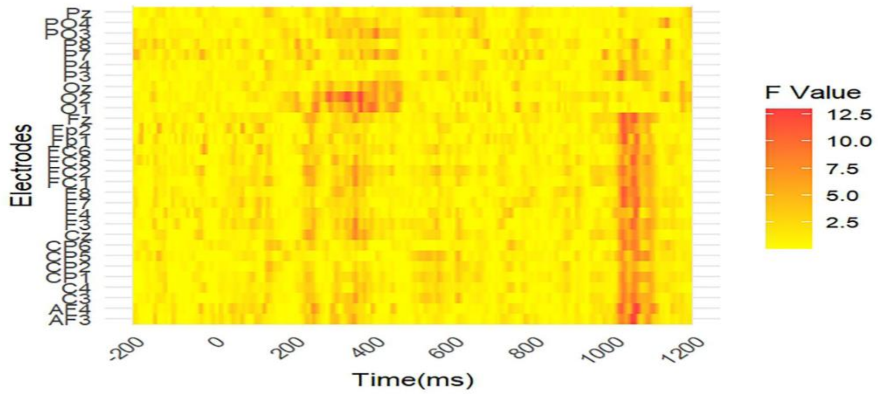


Figure 3.2: Permutation tests for three conditions and three repetition levels across all data electrodes between 0 and 1,200 ms post-stimulus onset. Larger F-values are shown in darker colors and give an increased likelihood of a statistically relevant effect of our manipulations on voltage amplitudes (n = 29).

As for the selection of electrodes in the ROIs, the permutation test highlighted ten relevant electrodes: F3, F4, Fz, FC1, FC2, Cz, C4, C3, P7, and PO3. However, not all the highlighted electrodes were relevant to the N400 component, as signal noise could also have contributed to the observed highlights. Based on previous literature (Brown & Hagoort, 1993; Lau et al., 2008; Šoškić et al., 2022), nine typical N400 electrodes, including F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 were commonly reported as ROIs. We chose the electrodes that overlapped the previous literature and the highlighted channels identified in our permutation test. We defined our ROIs as the following frontal and central electrodes: Cz, C4, C3, Fz, F3, and F4. Figure 3.3 illustrates the mean voltage amplitudes for the epoch of 1,200 ms for three conditions at three repetition levels. The mean amplitudes for each channel in the selected ROIs are shown in Figure 3.4.

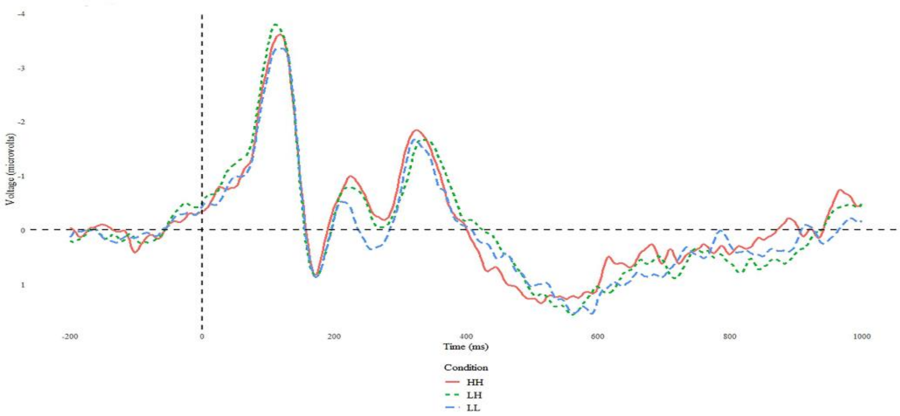


Figure 3.3: Voltage amplitudes for three conditions across three repetition levels over time for channels Fz, F3, F4, Cz, C3, and C4 in the picture-naming task (n = 29). The time window of interest is from 300 to 500 ms. Negativity is plotted up.

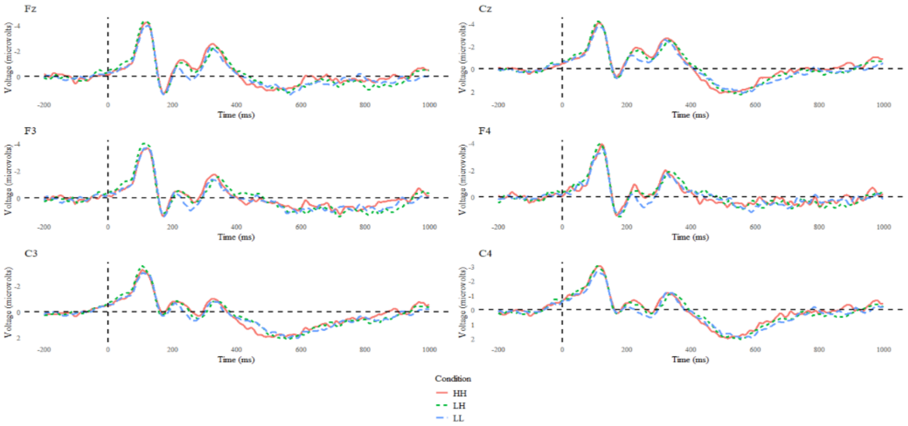


Figure 3.4: Voltage amplitudes for three conditions across three repetition levels over time for each channel of Fz, F3, F4, Cz, C3, and C4 in the picture-naming task (n = 29). The time window of interest is from 300 to 500 ms. Negativity is plotted up.

To surpass the limitations of conventional average-type analysis, we adopted a single-trial Linear Mixed Models (LMM) approach (Frömer et al., 2018; Spinnato et al., 2015; Von Grebmer zu Wolfsthurn et al., 2021a, 2021b). The traditional average-type analysis assumes equal weight for observations across conditions and participants. It assumes independence of factor levels, which often become problematic due to the intricacies of experimental designs. During EEG data preprocessing stages (Von Grebmer zu Wolfsthurn et al., 2021a, 2021b), the LMM approach addresses these limitations by adding random effects and is suitable for datasets with varying effect sizes and unbalanced designs (Baayen et al., 2008; Fröber et al., 2017). In the single-trial Linear Mixed Model (LMM) approach, we included all individual voltage values for each epoch within the selected time window of interest (300-500 ms). Rather than averaging the voltage values across segments from the same condition, we preserved the distinct voltage values to maintain by-subject and by-item variance. The model fitting procedure followed the same steps as in the behavioral data analysis. In this model, the fixed effects included *Condition* and *Repetition*, with *Region* as a covariate to investigate differences among various brain regions. Based on previous studies (Lensink et al., 2014; Von Grebmer zu Wolfsthurn et al., 2021a, 2021b), we divided the brain into seven regions: left-anterior, right-anterior, left-medial, right-medial, left-posterior, right-posterior, and midline. *Subject* and *Item* were included as random effects in this single-trial analysis. We incorporated an interaction effect between *Condition* and *Repetition*. Model comparisons were conducted using the *anova()* function, and factors that did not significantly improve the model fit were excluded during the final model selection procedure.

The model with the best fit was *Amplitude ~ Condition + Repetition + Region+ (*1 | *Subject) + (*1 | *Item)* (see Table 3.D.1 in *Appendix 3.D*). The differences in amplitude between the H(h) and the L(h) conditions were not significant (β = 0.30, *SE* = 0.26, t = 1.13, p = 0.26). The differences in amplitude between the conditions L(h) and L(l) were not significant (β = 0.16, *SE* = 0.25, t = 0.67, p = 0.51). The effect of *Region* was significant in the right medial region (β = 0.61, *SE* = 0.02, t = 38.22, p < 0.001), right anterior region (β = - 0.18, *SE* = 0.02, t = - 11.24, p < 0.001), left medial region (β = 0.81, *SE* = 0.02, t = 50.81, p < 0.001) and mid-line channels (β = - 0.05, *SE* = 0.01, t = - 3.51, p < 0.001). These effects showed that there were significant differences among different brain regions, indicating different processes related to word frequency. Different repetitions yielded significant differences between the first and second repetition levels (β = - 0.12, SE = 0.01, t = - 11.06, $p < 0.001$) and between the first and third repetition levels (β = - 0.35, SE = 0.01, t = - 31.07, $p < 0.001$). No interaction effect between *Condition* and *Repetition* was observed because after the model comparison of the ANOVA test, no significant improvement was found. Therefore, the interaction was not to be retained in this model.

Furthermore, based on the average amplitude of the selected channels, an N2 component was observed in Figure 3.3. To further investigate this, we conducted an exploratory analysis to test for the presence of an N2 effect within a time window between 200ms and 300ms. The results were β = - 0.07, SE = 0.33, t = - 0.24, p = 0.81 for H(h) and L(h) conditions and β = 0.04, SE = 0.31, t = 1.20, p = 0.23 for L(h) and L(l) conditions. Although no significant differences were found within the initially selected time window of 300-500 ms, we conducted an exploratory analysis using a narrower time window of 400-500 ms to assess the N400 effect. In this analysis, no significant differences were found between the H(h) and L(h) conditions with β = 0.53, SE = 0.35, t = 1.51, p = 0.14 or between the L(h) and L(l) conditions with β = 0.14, SE = 0.33, t = 0.43, p = 0.67.

3.4 Discussion

The ongoing discussion about the representation of compound words in our mental lexicon has prompted numerous studies to examine their representation. Some studies (Bien et al., 2005; Koester & Schiller, 2008, 2011; Levelt et al., 1999; Verdonschot et al., 2012) advocate the decomposed representation of compounds, while others (Chen & Chen, 2006; Janssen et al., 2008) support the full-listing approach. To further explore this topic, we replicated the design framework established by Janssen and colleagues (2008), using a new set of stimuli and integrating EEG into the design in order to investigate the research question whether the production of Mandarin compound words is influenced by morpheme or compound frequency in the present study.

Interestingly, the reaction times across conditions in our study did not align with the earlier findings of Janssen and colleagues (2008). Specifically, in our study, the L(h) condition, where word frequency was low but morpheme frequency was high, exhibited similar reaction times to the H(h) condition, where both word and morpheme frequencies were high, in picture naming tasks. In contrast, the L(h) condition was notably faster than the L(l) condition, where both word and morpheme frequencies were low. Due to the reported morpheme frequency effect, our current results aligned with other studies (Bien et al., 2005; Chen & Chen, 2015; Koester & Schiller, 2008; Roelofs, 1996) and lent support to decompositional model (Levelt et al., 1999).

The differences between two sets of stimuli may help explain the discrepancies observed between the two studies. Different corpora used could lead to different frequency distributions. Janssen and colleagues' study (2008) based their stimuli on the information found in the Modern Chinese Frequency Dictionary (Language Teaching and Research Institute of Beijing Language Institute, 1986). Our present study was based on SUBTLEX-CH corpus (Cai & Brysbaert, 2010) from movie subtitles. The differences between written and spoken corpora could potentially influence the evaluation of stimuli's frequency distributions.

The ERP results in the present study did not align with the behavioral data observed in our study. The current ERP findings revealed no significant differences in word frequency between the H(h) and L(h) conditions, nor were there morpheme frequency effects between the L(h) and L(l) conditions. While the early N2 component was noted for the L(h) and L(l) conditions and showed consistent tendency in the frontal and central regions, statistical analysis did not reveal significant differences. The time course of N2 component observed in this study was consistent with previous research (Strijkers et al., 2010), which identified an early component with approximately 180 ms after picture presentation and provided electrophysiological evidence for an early influence of frequency on speech production. Additionally, while regional effects were observed in our study, the main effect of frequency was not significant, so we did not explore this aspect further. Future research could aim to refine these areas to gain a deeper understanding of the underlying mechanisms.

Additional factors might have accounted for the current results. For instance, this study did not control for semantic transparency, which could impact the mechanism of compound word composition. Transparent compound words are easily understood because their meaning can be inferred from the meanings of their constituents. In contrast, opaque compound words do not allow for such straightforward inference, making it necessary to store their meanings as whole units in the lexicon (Schiller, 2020; Schiller & Verdonschot, 2019). Though some studies (MacGregor & Shtyrov, 2013; Tsang et al., 2022) have reported the effects of semantic transparency, other studies also showed that opaque and transparent compound words often do not differ significantly regarding morphological priming (Koester & Schiller, 2008; Verdonschot et al., 2012; Zwitserlood et al., 2000, 2002). For example, Koester and Schiller's paper (2008) found that the production of morphologically related and complex words facilitated subsequent picture naming and resulted in a reduced N400 compared to unrelated prime words, with no significant difference observed between transparent and opaque relations. Future research could account for the semantic transparency of stimuli to better understand its influence on morphological processing.

In addition, although we recruited native Mandarin speakers and controlled the duration of their stay in other countries to mitigate the effects of their multilingual background, this factor could still potentially influence the results. Furthermore, we did not account for the age of acquisition during the stimulus design. Despite a post-hoc analysis showing no significant difference ($t < 1$) in the effect of age of acquisition among three conditions in the present study, these speaker-related variables should be considered in future research.

3.5 Conclusion

In summary, the present study revealed that the picture naming latencies of Mandarin compound words at the behavioral level were influenced by the frequency of constituent morphemes rather than the frequency of the whole word. The behavioral data supported the predictions of a two-stage model of lexical access (Levelt et al., 1999) over a single-stage model (Caramazza, 1997). However, the ERP results did not show frequency effects at either the whole-word or morpheme-constituent levels. This discrepancy between the behavioral and ERP findings underscored the need for further research to explore and understand these deviations. The present results emphasized the importance of additional studies on morphological representations in Mandarin Chinese, which could contribute to our understanding of linguistic processes, even though they may differ from those observed in languages with more complex morphology, such as Dutch.

CRediT author contribution statement

Jiaqi Wang: Conceptualization, Methodology, Validation, Investigation, Formal analysis, Data curation, Funding acquisition, Writing-original Draft, Writing-review and editing, Visualization. **Niels O. Schiller:** Conceptualization, Writing-review and editing, Funding acquisition, Supervision. **Rinus G. Verdonschot:** Conceptualization, Methodology, Writing-review and editing, Supervision.

Acknowledgments

We sincerely thank Sarah von Grebmer zu Wolfsthurn and other group members for their invaluable support throughout the stimuli selection, data collection, and analysis phases of this study. We also extend our gratitude to the lab technicians for their assistance. We extend our gratitude to Claartje Levelt for her valuable feedback on this manuscript. Finally, we are grateful to all our participants for their participation and our anonymous reviewers for their comments and feedback.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the China Scholarship Council (No. 202008130142).

Data availability

Data will be made available on request.

Appendix

3.A Experimental stimuli

Chinese	English	Pinyin	English for constituent	Condition
电话	telephone	diànhuà	electricity + speech	HH
警察	police	jǐngchá	police + check	HH
医生	doctor	yīshēng	medical + scholar	HH
照片*	photo	zhàopiàn	photo + slice	HH
衣服	clothing	yīfú	clothes + clothes	HH
学校*	school	xuéxiào	learn + school	HH
眼睛	eye	yǎnjīng	eye + eye	HH
医院*	hospital	yīyuàn	medical + yard	HH
飞机	airplane	fēijī	fly + machine	HH
电视	television	diànshì	electricity +vision	HH
监狱	prison	jiānyù	supervise + prison	HH
礼物	gift	lǐwù	gift + object	HH
头发	hair	tóufǎ	head + hair	HH
老师*	teacher	lǎoshī	old + master	HH
钥匙	key	yàoshi	key + key	HH

手机	cell phone	shǒujī	hand + machine	HH
城市*	city	chéngshì	city + city	HH
电脑	computer	diànnǎo	electricity + brain	HH
酒吧*	bar	jiǔbā	wine + bar	HH
汽车	car	qìchē	gasoline + car	HH
啤酒	beer	píjiǔ	beer + wine	HH
地球	earth	dìqiú	earth + ball	HH
炸弹	bomb	zhàdàn	bomb + bullet	HH
音乐	music	yīnyuè	sound + music	HH
蛋糕	cake	dàngāo	egg + cake	HH
乐队*	band	yuèduì	music + team	HH
法官	judge	fǎguān	law + officer	HH
教堂*	church	jiàotáng	teach + hall	HH
长椅	bench	chángyǐ	long + chair	LH
海马	seahorse	hǎimǎ	sea + horse	LH
水井*	well	shuǐjǐng	water + well	LH
弹弓	slingshot	dàngōng	bullet + bow	LH
手电	flashlight	shǒudiàn	hand + electricity	LH

风车	windmill	fēngchē	wind + car	LH
树枝	branches	shùzhī	tree + branch	LH
日历	calendar	rìlì	day + history	LH
钱包	wallet	qiánbāo	money + bag	LH
电车	tram	diànchē	electricity + car	LH
书包	bag	shūbāo	book + bag	LH
背心	vest	bèixīn	back + heart	LH
旗帜	banner	qízhì	flag + banner	LH
别针	pin	biézhēn	pin + needle	LH
鹿角	antlers	lùjiǎo	deer + horn	LH
蚂蚁	ant	mǎyǐ	ant + ant	LH
长城	great wall	chángchéng	long + wall	LH
轮椅	wheelchair	lúnyǐ	wheel + chair	LH
水杯	water cup	shuǐbēi	water + cup	LH
试管	test tube	shìguǎn	test + tube	LH
海盗	pirate	hǎidào	sea + robber	LH
奖杯	trophy	jiǎngbēi	prize + cup	LH
毛巾	towel	máojīn	hair + towel	LH

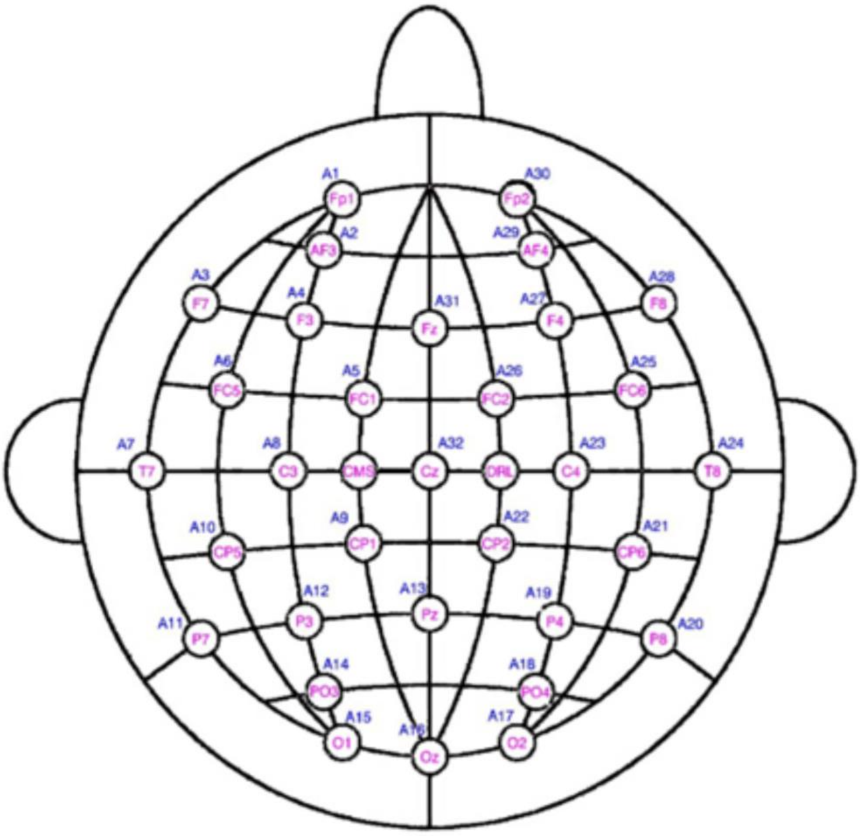
箭头	arrow	jiàntóu	arrow + head	LH
河马	hippo	hémǎ	river + horse	LH
火箭	rocket	huǒjiàn	fire + arrow	LH
面具	mask	miànjù	face + tool	LH
手铐	handcuffs	shǒukào	hand + shackle	LH
钮扣	buttons	niǔkòu	button + buckle	LL
插头	plug	chātóu	insert + head	LL
卷尺	tape measure	juǎnchǐ	roll + ruler	LL
衬衫	shirt	chènshān	lining + shirt	LL
菜板	chopping board	càibǎn	vegetable +board	LL
肩膀	shoulder	jiānbǎng	shoulder + arm	LL
仓库	storehouse	cāngkù	warehouse + storage	LL
翅膀	wing	chìbǎng	wing + arm	LL
豌豆	pea	wāndòu	pea + bean	LL
盾牌	shield	dùnpái	shield + card	LL
拱桥	arch bridge	gǒngqiáo	arch + bridge	LL
黑板	blackboard	hēibǎn	black + board	LL
胶囊*	capsule	jiāonáng	glue + bag	LL

孔雀	peacock	kǒngquè	hole + sparrow	LL
恐龙	dinosaur	kǒnglóng	fear + dragon	LL
蜻蜓	dragonfly	qīngtíng	dragonfly + dragonfly	LL
奶酪	cheese	nǎilào	milk + cheese	LL
鞭炮	firecracker	biānpào	whip + cannon	LL
竹简*	bamboo slips	zhújiǎn	bamboo + slips	LL
琵琶	lute	pípá	lute + lute	LL
漏斗	funnel	lòudǒu	leak + cane	LL
墓碑	tombstone	mùbēi	grave + monument	LL
葡萄	grape	pútáo	grape + grape	LL
拐杖	crutch	guǎizhàng	turn + cane	LL
扫帚	broom	sàozhǒu	sweep + broom	LL
雕塑	sculpture	diāosù	carve + sculpture	LL
熨斗	iron	yùndǒu	iron + cane	LL
萝卜	radish	luóbo	radish + radish	LL

*: deleted items

3.B EEG electrode montage

Figure 3.B.1: 10/20 32-channel montage from BioSemi including CMS and DRL but excluding external channels.



3.C Model parameters: response times

Table 3.C.1: Specification of best-fit model for response times (RTs) for picture naming task (n = 29). Note that estimates are reported in milliseconds.

Formula: RT ~ Condition + Repetition + (1 + Repetition | Subject) + (1 + Condition | Subject) + (1 | Item)				
Term	Estimate	[95%CI]	t-value	p-value
(Intercept)	679.03	[665.55, 685.64]	144.1	<0.001***
Condition: LH	-4.85	[-13.27, 2.93]	-0.98	0.328
Condition: LL	12.38	[2.77, 18.08]	2.86	0.004**
Repetition: 2	-28.34	[-30.02, -18.32]	-7.38	<0.001***
Repetition: 3	-17.96	[-21.96, -10.29]	-3.36	<0.001***
Random effects				
σ^2	0.03			
τ00Item	144.63			
τ00Subject [Repetition]	562.46			
τ00Subject [Condition]	509.57			
τ11Subject [Repetition2]	331.71			
τ11Subject [Repetition3]	810.72			
τ11Subject [ConditionLH]	219.55			
τ11Subject [ConditionLL]	441.29			
ICC	0.84			
NSubject	29			
NItem	73			
Observations	6033			
Marginal R^2	0.16			
Conditional R^2	1.00			

Note: σ^2: Residual variance, representing the unexplained variability in the model. τ_{00}: Variance of random intercepts, indicating how much baseline levels vary across groups (e.g., subjects or items). τ_{11}: Variance of random slopes, reflecting how much the effect of a predictor varies across groups. R^2: Proportion of variance explained by the model, indicating overall model fit.

3.D Model parameters: N400 component

Table 3.D.1: Specification of the model of best fit for Voltage Amplitudes (microvolts) (n = 29).

Formula : Amplitude ~ Condition + Repetition + Region + (1 | Subject) + (1 | Item)				
Fixed effects	Estimate	95%CI	t-value	p-value
(Intercept)	0.45	[-1.23, 0.32]	-1.15	0.26
Condition: HH	0.30	[-0.22, 0.81]	1.13	0.26
Condition: LL	0.16	[-0.32, 0.65]	0.67	0.51
Repetition: 2	-0.12	[-0.15, -0.10]	-11.06	<0.001***
Repetition: 3	-0.36	[-0.37, -0.33]	-31.07	<0.001***
Region: anterior right	-0.18	[-0.21, -0.15]	-11.24	<0.001***
Region: central left	0.81	[0.77, 0.84]	50.81	<0.001***
Region: central right	0.61	[0.58, 0.64]	38.22	<0.001***
Region: midline	-0.05	[-0.08, -0,02]	-3.51	<0.001***
Random effects				
σ^2	71.09			
τ00Item	0.80			
ICC	0.03			
NSubject	29			
NItem	73			
Observations	3389256			
Marginal R^2	0.002			
Conditional R^2	0.06			

Note: σ^2: Residual variance, representing the unexplained variability in the model. τ_{00}: Variance of random intercepts, indicating how much baseline levels vary across groups (e.g., subjects or items). τ_{11}: Variance of random slopes, reflecting how much the effect of a predictor varies across groups. R^2: Proportion of variance explained by the model, indicating overall model fit.

Chapter 4 The effect of synonym distractors on naming Mandarin compound words

4.1 Introduction

Compound words (e.g., “birdhouse” or “sunflower”) are omnipresent in most languages, such as English and Chinese, and are effortlessly spoken and comprehended. The meaning of a compound is shaped by the significance of its constituents. These meaning relations can range from highly transparent, as exemplified by Mandarin 公鸡 /gong1ji1/ “rooster” (lit. “male chicken”) or English “birdhouse” (a house for birds), to relatively opaque, as observed in Mandarin 背心 /bei4xin1/ “vest” (lit. “back heart”) or English “butterfly” (Badecker, 2001; Libben et al., 2003; Zwitserlood, 2014). However, the interaction between compounds and their constituents during the compound production process remains debatable, particularly regarding how much their morphological structure influences processing within the lexical system.

According to prevailing speech-production theories, the process of production, for instance, the act of naming an object (e.g., a bird), entails the activation of conceptual information, retrieval of the corresponding lexical entry, and the initiation of phonological and phonetic encoding preceding articulation (Caramazza, 1997; Dell, 1986; Levelt et al., 1999). In this production process, morphemes, the smallest units carrying meaning (Lieber, 2021), play a crucial role at the word-form level. While there is substantial empirical evidence supporting the idea of morpheme-based representations playing a role in speech production (Jacobs & Dell, 2014; Lüttmann et al., 2011; Roelofs, 1996a; Köster & Schiller, 2008; 2011; Wang et al., 2024), they do not conclusively pinpoint to the specific level at which these representations come into play, whether at the lexical-syntactic level (referred to as the lemma) or the word-form level or both. For example, according to two-stage models of speech production (Levelt et al., 1999), compounds are represented at least at one level of lexical representation in terms of their constituent lexical morphemes. The present study focuses on the representation of Mandarin compounds at the lemma level by first introducing two dominant representation models.

4.1.1 Two-stage model

The two-stage model (see Figure 4.1) proposes distinct sequential lexical levels: a lemma and a word-form level (Dell, 1986; Dell & O'Seaghdha, 1992; Garrett, 1980; Levelt, 1993; Levelt et al., 1999; Roelofs, 1996a, 1996b; Roelofs & Meyer, 1998; Roelofs et al., 1996). Lemmas represent a word's syntactic properties, while word forms specify segmental and metrical information and constituent morphemes in polymorphemic words (Roelofs, 1996a, 1996b; Roelofs

& Meyer, 1998). In two-stage models, the production of compound words unfolds as follows. Naming an object (e.g., 公鸡 /gong1ji1/ "rooster") initiates with the activation of the relevant semantic concept, co-activating semantically related concepts (e.g., 鸭子 /ya1zi/ "duck," 母鸡 /mu3ji/ "hen," 鸡蛋 /ji1dan4/ "egg," and 农场 /nong2chang3/ "farm") through spreading activation. This results in the activation of multiple semantically related lexical entries at the lemma level, competing for lexical selection until one target lemma is chosen (Abdel Rahman and Aristei, 2010; Abdel Rahman & Melinger, 2009; Damian & Bowers, 2003; see Bürki et al., 2020 for a meta-analysis). Levelt and colleagues (1999) propose that compound words are stored holistically at the lemma level. This compound lemma accesses multiple morphemes at the word-form level. After selecting a single lemma, the constituent morphemes are retrieved at the word-form level. Morpho-phonological and phonetic encoding follows before articulation occurs (Indefrey, 2011; Indefrey & Levelt, 2004). Sequential processing evidence suggests that a compound's morphemes are processed sequentially, with the first morpheme being encoded before the second, and so on (Roelofs, 1996b). Two-stage models posit single, holistic compound lemmas but decomposed (morpheme-based) form representations (Dell, 1986; Levelt et al., 1999).

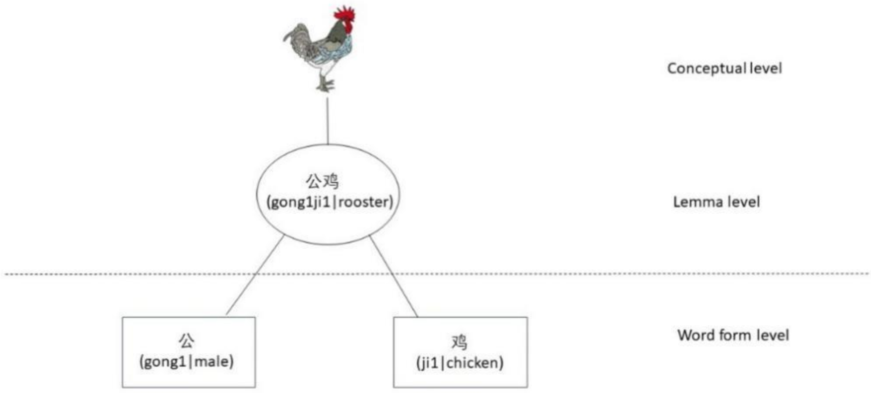


Figure 4.1: The two-stage model (holistic lemma model).

Given that a compound also possesses linguistic properties, such as grammatical class and lexical gender, which may not be directly linked to its individual constituents, one plausible approach is conceptualizing the compound structure as an autonomous lemma node. Mondini and colleagues (Mondini et al., 2004) investigated the mental representation of Verb-Noun (VN) nominal compounds utilizing neuropsychological methods. Their study assessed the lexical retrieval of compound nouns in 30 aphasic patients through a visual confrontation naming task. The targeted names include VN compounds, Noun-Noun (NN) compounds, and long morphologically simple nouns (LSN). To evaluate participants' ability to produce simple nouns and verbs, another visual confrontation naming task involving objects and actions is conducted. The study's findings confirm that individuals with a disproportionate verb deficit also struggle to name VN compounds. The data support the notion of (de)compositional processing of compound words. Another subgroup of patients demonstrates a specific impairment in naming compound nouns compared to similarly lengthy simple nouns, regardless of their VN or NN morphological structure. This suggests a specific disorder related to retrieving two distinct lexemes with a single lexical entry. The independent lemma node can activate constituent representations stored separately at the lexeme level. In this way, the compound is considered to have a distinct representation that can engage and draw upon individual constituent elements when needed. Overall, the literature supports the two-stage model of language production, though the precise locus of decomposition varies across different studies.

4.1.2 Hybrid lemma account

On the other hand, according to the super lemma model (Sprenger et al., 2006; Kuiper et al., 2007; see Figure 4.2), multi-word lexical items, such as idiomatic forms and phrases, are conceptualized as "super lemmas” at the lemma level. A superlemma is a separate representation of multi-word lexical items at the lexical-syntactic processing level. It encodes the syntactic properties of the idiom while being linked to its individual components, the simple lemmas. Given the similarities between compounds and idioms, i.e., they both map onto a single conceptual representation and consist of multiple morphemes; it is proposed in the present study that Mandarin compounds might share a similar hybrid representation at the lemma level as idioms, involving a holistic lemma complemented by constituent-specific lemmas.

In this hybrid representation approach, the selection and processing of an idiomatic expression is very similar to that of a single word. For example, because producing the multi-lexical item "hit the road" requires the selection of its super lemma "hit the road," producing the Mandarin compound 公鸡 ("rooster") needs the selection of its super lemma 公鸡 in a holistic form as well. The probability of selecting the target superlemma from the mental lexicon depends on the ratio of the superlemma's activation level to the total activation of all lemmas (both superlemmas and simple lemmas), a concept known as Luce's ratio (1959; 1986). Once the superlemma is selected, the syntactic constraints associated with that item become accessible to the production system, influencing or restricting the syntactic properties of the simple lemmas involved. Additionally, selecting the target superlemma determines the simple lemmas to be chosen in subsequent processing steps. Superlemma selection is a prerequisite for activation spreading to the dependent simple lemmas, and the selection of these lemmas also follows the principles of Luce's ratio. The target lemma competes with all other active superlemmas and simple lemmas throughout the process (Sprenger et al, 2006).

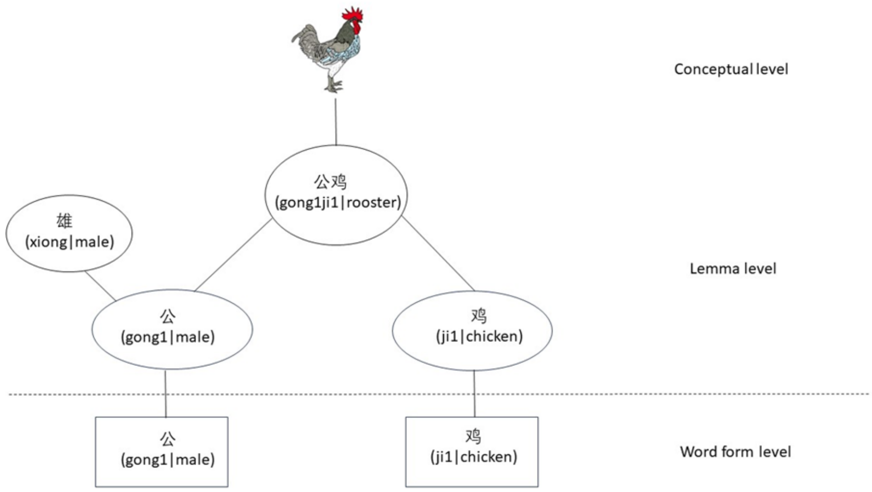


Figure 4.2: The super lemma model (hybrid lemma account).

In summary, it is unclear how exactly compound words are encoded in Mandarin, particularly regarding the representation of the constituent morphemes at different lexical levels - whether at the word-form level, lemma level, or both. Due to the limited and inconsistent evidence available, the precise nature of Mandarin compound representation is still unclear. The present study aimed to further explore this issue by examining the effect of synonym distractors on compound naming latencies, providing new insights into the representation of Mandarin compounds.

4.1.3 Effects of synonym distractors on lexical processing

The connection between a word and a synonym is one of the intrinsic relations connected based on meaning because words are organized and stored based on their meaning relations (Collins & Loftus, 1975; Collins & Quillian, 1969). Languages often contain multiple words with the same meaning due to historical and etymological factors. Clark (1987) argued that "any difference in form in a language marks a difference in meaning," implying that "there are no true synonyms" (p. 3). Social and linguistic processes further distinguish synonyms by associating them with specific contexts or subtly different referents. Although the synonyms used in the present study may not be interchangeable in all contexts, they were acceptable alternative (nearly synonymous) names for the objects as stated in a study by Dylman and Barry (2018).

Several studies have investigated the effects of synonyms, providing evidence for the activation of target word synonyms during the language production process. For instance, research by Peterson and Savoy (Peterson & Savoy, 1998) demonstrated that words tend to activate their synonyms. Notably, during processing, the word "couch" activated its synonym "sofa," indicating that such activation extends even to phonological encoding. This finding challenges the earlier notion proposed by Levelt et al. (1999) that phonological activation of synonyms was not feasible. Additionally, Levelt et al. (1999) acknowledged that simultaneously selecting multiple lexical representations can occur under time pressure. Roelofs (2008) argued that the language production model is indeed "weakly cascading," providing further insights into the activation of synonyms in lexical processing.

Furthermore, Dylman and Barry (2018) explored a synonymous distractor facilitation effect in language production. They presented five experiments using the picture-word task to investigate lexical selection by comparing the effects of translation distractors in bilinguals and synonym distractors in monolinguals. In their study, the effects of synonyms in monolinguals seem to "mirror" the effects observed for translations in bilinguals, suggesting shared mechanisms in lexical selection for both monolinguals and bilinguals. They proposed that for all speakers, multiple candidate words were activated by the semantic representation required to convey an intended idea in speech or to name an object in a laboratory setting. The selection of a specific word depended on the activation levels of lexical representations from various sources-most notably, from a language control system that responds to language choice, task demands, and, more broadly, the speaker's assessment of what was appropriate for the communicative context. Their study provided valuable insights for the present study, suggesting a facilitation effect of synonym distractors and providing implications for the lexical selection.

In the present study, the rationale behind selecting synonyms as distractors was driven by the aim to replicate the substantial overlap in semantic features expected to exist among morphologically related words. Therefore, the effects of synonym distractors could "simulate" the morphologically related words. For instance, by using synonymous word pairs (e.g., 公 /gong1/ and 雄 /xiong2/ both mean "male"), the present study sought to capture and mirror the considerable similarity in meaning that typically characterizes words with a common morphological root. This choice allows us to investigate and isolate the effects of constituent relatedness on semantic processing and feature overlap (Marslen-Wilson et al., 1994). Utilizing synonyms in the present study is to examine their impact on constituents, assuming they are stored in a decomposed form at the lemma level. For instance, the influence of 雄 /xiong2 / "male" on 鸡 /ji1/ "chicken" should be minimal if not decomposed in the lexical retrial of 公鸡 /gong1ji1/ "rooster."

4.1.4 The characteristics of Mandarin

Regarding the characteristics of Mandarin, it could make Mandarin worth exploring and provide different insights regarding its representation in the mental lexicon because Mandarin is classified as a logographic system, with characters as the fundamental writing units, and is composed of numerous strokes organized into a square shape (Tan et al., 2001). Chinese speakers may have heightened awareness of lexical items due to the presence of Chinese characters, which often provide immediate visual cues to their meanings. The configuration of a Chinese character often imparts contextual clues about its meaning. Additionally, Chinese characters commonly integrate semantic radicals, constituent elements that offer insights into the character's general meaning, further enriching the context and aiding comprehension. For instance, the character 虫 /chong2/, meaning "insect," serves as a visual indicator in characters like 蚁 /yi3/ for "ant," 蛾 /e2/ for "moth," and

蝴蝶 /hu2die2/ for “butterfly.” Even without knowing the specific character, recognizing 虫 suggests an association with insects, except 虹 /hong2/ meaning “rainbow.” This underscores the significance of the logographic writing system.

In addition, unlike alphabetic languages, Chinese characters do not correspond directly to phonemes but instead map onto meaningful morphemes in the spoken language. The pronunciation of Chinese characters is established at the syllable/morpheme level. It necessitates learning through repetitive rote memorization of the connection between the visual form of characters and their corresponding sounds. Occasionally, this learning process is facilitated by including sub-character units, usually authentic characters in their own right. This implies that the regular or quasi-regular grapheme-phoneme conversions commonly found in alphabetic language are not feasible in Chinese (Plaut, 1996; Tan & Perfetti, 1998).

In summary, the semantic and phonological features in Mandarin may enhance speakers’ lexical awareness through the visual cues provided by characters and radicals, emphasizing a high awareness of individual constituents. This suggests a potential difference in the lemma representation of constituents in Mandarin compound production compared to Indo-European languages. The synonym distractor condition in the present study offers insights for further exploration of this topic.

4.1.5 The present study

The present study aims to provide further evidence regarding the representation of compounds at the lemma level. Using the picture-word interference (PWI) paradigm, native Mandarin speakers overtly named pictures of disyllabic compounds. This paradigm traces back to the original Stroop color-naming paradigm (Glaser & Glaser, 1989; La Heij et al., 1998). The experimental setup involved the simultaneous presentation of distractor words alongside target pictures, with participants instructed to name the pictures while disregarding the distractors rapidly. By comparing naming latencies in the presence of experimentally manipulated distractor words, researchers can observe facilitation or interference depending on the relation between targets and distractors (Glaser & Düngelhoff, 1984).

In the present study, three conditions were introduced to investigate the representation of Mandarin compound words at the lemma level (e.g., 公鸡 /gong1ji1/ "rooster"). The first condition includes distractors synonymous with the initial morpheme, i.e., having the same meaning but being phonologically and orthographically dissimilar to the compound's initial morpheme (i.e., 雄 /xiong2/ "male"). This was done to explore whether the first constituent could be individually affected by its synonym. The second condition includes distractors (e.g., 鹅 /e2/ "goose") semantically related to the entire compound. This condition was used to assess semantic processing and access to the holistic lemma of the whole compound. The last condition includes unrelated distractors, which served as baseline condition (e.g., 车 /che1/ “car”). We expected that the naming latencies for the synonym distractor condition would be faster than for the semantically related distractor condition and the unrelated control condition due to the synonym facilitation effect (Dylman and Barry, 2018); the semantically related distractor condition is expected to be the slowest due to a semantic interference effect (Rosinski, 1977; Lupker, 1979).

4.2 Methodology

4.2.1 Participants

Thirty-nine right-handed native Chinese college students were recruited from Beijing Normal University and China University of Petroleum (Beijing) with normal or corrected-to-normal vision. Three participants were excluded due to technical problems. The final sample included 36 adult speakers (13 males, mean age 20.55 ± 1.60 years, range 18-24 years).

At the time of testing, none of the participants reported color blindness, learning disorders, auditory or visual impairments, or psychological or neurological impairments. Participants read the information sheet and signed an informed consent form before participating. They received monetary compensation for their participation.

4.2.2 Materials

Forty-five transparent Chinese disyllabic compound words represented by 45 colored pictures were selected. A group of native Mandarin speakers evaluated the semantic transparency of the stimuli. The images are sourced from the MultiPic corpus (Duñabeitia et al., 2018), a colored picture corpus. Additionally, an illustrator created extra images that were not available in this corpus. For instance, the stimulus 首都 /shou3du1/ “capital city” was presented alongside a map of China, with the location of Beijing highlighted by a red star.

The Picture-Word Interference (PWI) paradigm was used in the present study with three conditions. In the first condition, the distractor words were monosyllabic synonyms with the same meaning as the first constituent but phonologically and orthographically different from it. The distractor words in the second condition were semantically related to the whole compound. The third condition's distractors were unrelated to the whole compound or the two separate constituents. The distractors in the last two conditions were monosyllabic and disyllabic, with monosyllabic dominance. Therefore, a total of 135 trials were created for each participant. The word frequency ($F < 1$) of both targets and distractors was controlled (see Table 4.1). *Appendix 4.A* presents an overview of all items used in the experiment.

The effect of visual complexity in the PWI paradigm was consistent across the different distractor conditions, as the same picture was used in all three conditions in the present study. This allowed us to compare these conditions directly. The number of strokes for three distractor conditions was not controlled. Due to the fact that we included monosyllabic and disyllabic distractors for the semantically related and unrelated conditions, and we only included monosyllabic distractors for the synonym condition, the number of strokes was different for the three conditions.

Table 4.1: Illustration of the experimental conditions.

Condition	Target	Distractors	Word Frequency (Zipf)	
			Mean	SD
Synonymous	公 鸡 /gong2ji2/ “rooster”	雄/xiong2/ “male”	2.92	0.94
Semantically related		鹅 /e2/ “goose”	3.00	0.61
Unrelated		车 /che1/ “car”	2.97	0.66

4.2.3 Design and Procedure

A one-factorial (three conditions) within-subject design was adopted in this experiment. Two sets of pseudo-randomization stimuli lists were prepared. Participants were instructed to name the pictures during the experiments. The experiment was designed and controlled using *E-prime 3.0* (Psychology Software Tools) and was conducted in a soundproof booth. There were three sessions in this experiment, including the familiarization phase, the practice phase, and the experiment phase.

Participants were first instructed to familiarize themselves with the pictures and their names by learning the names of the target pictures. Subsequently, the experimenter assessed whether participants correctly remembered the picture names in the practice session. In this session, participants were asked to name all the target pictures as fast as possible, and the experimenter would correct any errors. In the experiment phase proper, participants were asked to name the pictures they had practiced while ignoring the distractor words on these pictures. Since we had two versions of the stimuli, half of the participants used the first version, and the other half used the second version to minimize any effect of order and repetition.

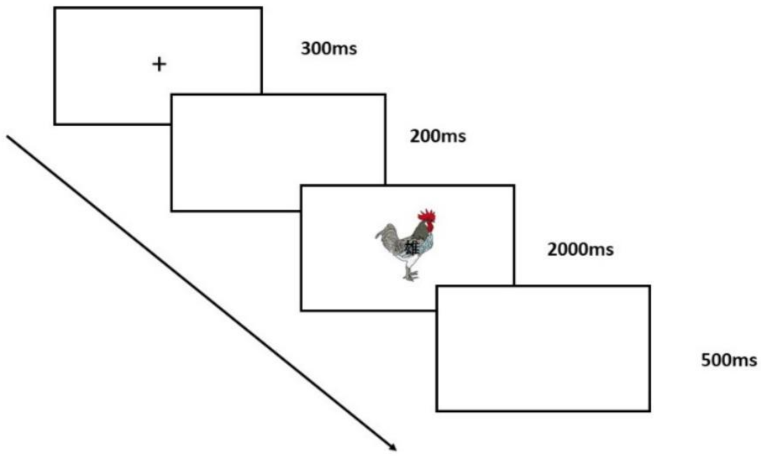


Figure 4.3: Trial sequence for the picture-naming task.

There were short breaks during the experiment. Every trial started with the display of a fixation cross lasting 300ms, succeeded by a 200ms interval with a blank screen. Subsequently, the stimulus appeared at the center of the screen for a duration of 2,000ms, followed by an additional blank screen lasting 500ms (see Figure 4.3). At the same time, the experimenter recorded the validity of each trial by noting target picture errors and voice-key errors. No feedback was provided during the experiment.

4.3 Data analysis

A proportion (6.65%) of data trials were excluded from the RT analysis due to naming errors and trials with RTs that deviated more than 3 SD from the mean per participant per condition. Behavioral data were analyzed using *R Studio* Version 4.2.2. We first calculated each condition's descriptive statistics for naming latencies (see Table 4.2). Then, we employed a generalized linear mixed effect model (GLMM) using the *glmer()* function with a gamma distribution to model skewed RT data.

Random effects were chosen in such a way as to avoid over-parameterization and to balance Type-I error and power (Matuschek et al., 2017). We followed a modeling approach where we maintained the simplest possible model structure considering our central manipulations (Bates & Maechler, 2020; Matuschek et al., 2017; Von Grebmer zu Wolfsthurn et al., 2021). Our model included Conditions as fixed factors and Trials as co-variates. In the present PWI paradigm, participants saw and named the targeted picture three times, which yielded a Repetition effect. Therefore, the trial order was included in the model.

Table 4.2: Mean naming latencies (in ms) separated by distractor condition.

Distractor	Naming Latency (ms)		Accuracy (%)	
	Mean	SD	Mean	SD
Synonymous	829	182	95.74	4.46
Semantically related	866	187	94.32	5.51
Unrelated	847	180	95.56	4.95

We performed model fit checks by plotting the model residuals against predicted values. We used the *avova()* function to perform model comparisons and likelihood ratio tests based on Akaike's Information Criterion, *AIC* (Akaike, 1974), *BIC* (Neath & Cavanaugh, 2012), and the log-likelihood in order to establish the best-fitting model for our data. In the PWI paradigm, the repetition problem should be considered, i.e., Trial was added to the model building. The factor of *Trial* was scaled before being added to the model. The package of *emmeans* was used to contrast three conditions, using the *tukey* method as adjustment.

For naming latencies, the model of best fit was *RT ~ Condition + Trial + (1 | Subject) + (1 | Item)* (see Table 4.B.1 in *Appendix 4.B*). The results showed that participants were significantly faster in naming target pictures when the distractors shared the meaning with the first constituent than unrelated distractors with β = - 16.6, *SE* = 4.42, *z* = -

3.75, $p < 0.001$ (see Figure 4.4). The naming latencies of the second condition where distractors were semantically related to the whole compound words were significantly longer than distractors were unrelated with $\beta = 21.1$, $SE = 5.9$, $z = 3.58$, $p = 0.001$. The factor Trial was included in the final model, showing no significant effect ($\beta = 2.911$, $SE = 2.18$, $z = 1.34$, $p = 0.181$).

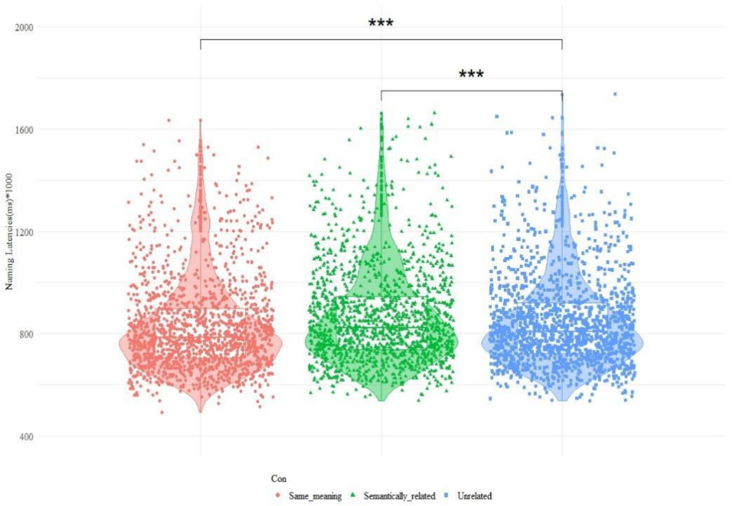


Figure 4.4: Mean naming latencies by condition for the picture naming task (n = 36).

4.4 Discussion

The current study used the PWI paradigm to address the representation of Mandarin compounds at the lemma level in the mental lexicon. Native Mandarin speakers were asked to name pictures and ignore distractors; behavioral data were collected and analyzed. The findings revealed that naming latencies for synonym distractors were significantly faster than for unrelated distractors, aligning with the facilitation effects of synonyms in a previous study (Dylman & Barry, 2018). Semantically related distractors led to significantly longer latencies compared to the control group, which was aligned with the inhibition effects in previous studies (Levelt et al., 1999; Lupker, 1979; Roelofs, 1992; Starreveld & La Heij, 1995, 1996), showing that semantically related distractors are inhibitory in the present study.

As for the semantically related distractor condition, we failed to control the semantically-related distractor condition. In our original research design, we should have selected semantic-categorically related distractors, but in the experiment conduction, two types of semantically related distractors were involved in the present study: categorically related distractors and associatively related distractors. Both two types of semantically related distractors elicited interference effects in the present study even though the inhibitory effect was usually only found for semantic categorically related distractors. The interference effect caused by the categorically related distractors aligned with the inhibition effects in previous studies (Levelt et al., 1999; Lupker, 1979; Roelofs, 1992; Starreveld & La Heij, 1995, 1996).

As for the inhibitory effect caused by semantic associative distractors, earlier findings on associatively related items are more mixed, typically showing either no effect or facilitation (Costa et al., 2005; Kuipers & La Heij, 2008; Mahon et al., 2007; but see Sailor & Brooks, 2014; Hantsch et al., 2005). The inhibitory effect of associatively related distractors in the current experiment contrasted with previous studies. However, our findings aligned with those of Lupker (1979), who investigated the effects of categorical and associative relationships in picture-word interference. Lupker assessed whether these relationships produced additive effects by comparing distractors that were solely categorically related with those that were both categorically and associatively related. His study found that both types of distractors caused the same level of interference.

In addition, the current results also have implications for lexical competition theories (Levelt et al., 1999). The picture-word interference tasks are often explained through lexical competition theories, where response selection is slowed due to competition between activated related words (e.g., Levelt et al., 1999). When a distractor activates a related non-target word, it becomes a competitor, slowing the selection of the target name. Furthermore, Finkbeiner, Gollan, and Caramazza (2006) highlighted that "the closer two lexical representations are in meaning, the more difficult it will be to select the correct one" (p. 153). This "hard problem" of lexical selection affected synonyms retrieval in a language. Jescheniak and Schriefers (1998) found that distractors related to near-synonym names slowed target naming times, showing the influence of closely related meanings which supported the above-mentioned "hard problem". In our study, however, synonym distractors appeared to be co-activated without evidence of interference, challenging theories of lexical selection that rely on competition. The selection-by-competition model would predict interference from closely related synonyms of constituents, which was not observed in the present study.

As for the representation of compounds, according to Levelt and colleagues' (1999) model, compounds are initially stored as a single lemma at the lemma level and subsequently decomposed at the ensuing lexeme level, where they then undergo phonological encoding. The observed facilitation effect from constituent synonyms in this study supported a separate lemma account for each constituent in Mandarin compound production. This effect indicated that the activation occurred at the conceptual level for each constituent, followed by lemma selection, alongside the lemma of the synonyms of these constituents. The lemmas of these synonym distractors facilitated the retrieval of compounds because they shared a semantic relationship with the first constituent of the compounds. These findings carry implications for the super-lemma account (Sprenger et al., 2006; Kuiper et al., 2007) of lexical representation for multi-lexical items outlined in the Introduction.

The current findings suggested that the lemma representation of Mandarin compounds differs from that of Dutch or German compounds, as anticipated by the two-stage model. As outlined in the Introduction, these observed differences in Mandarin compounds may result from cross-linguistic distinctions. A key difference lies in Mandarin's logographic writing system, where characters-the basic writing units-are composed of multiple strokes arranged within a square form (Tan et al., 2001). Mandarin speakers may develop heightened lexical awareness due to the presence of characters and radicals, visual elements that instantly evoke word meanings, potentially enhancing lexical awareness compared to speakers of languages lacking such visual cues. Additionally, Mandarin characters are pronounced at the syllable or morpheme level, requiring rote memorization to associate visual forms with sounds. Learning is often supported by sub-character units, which are independent characters themselves. Unlike alphabetic languages, Mandarin characters correspond not to phonemes but to meaningful morphemes in spoken language, making the regular grapheme-phoneme conversions seen in alphabetic languages irrelevant in Mandarin (Plaut, 1996; Tan & Perfetti, 1998).

In addition, manipulating synonyms of constituents to address the current research question is uncommon. Limited research has focused on synonym distractors, particularly within the context of compound components, due to the potentially dense semantic overlap between synonyms and compounds. In this study, synonyms serve as distractors that share the same meaning as the first constituent but differ in phonology and orthography. This setup mirrors the substantial semantic overlap typical of morphologically related words. The activation and retrieval of the first constituent's meaning in the mental lexicon contribute to the facilitation effect observed with synonyms. In Mandarin compounds, which are incredibly transparent, the synonym of a constituent often encompasses the semantic features of the entire compound, which provides a challenge to explore how to control for potential confounds when designing experiments on the language production process. There is significant potential for further investigation of synonyms, especially in languages like Mandarin, where many synonyms exist.

It is essential to highlight that the present study examined transparent words, leaving aside the investigation of opaque compounds. Two characters are combined when forming transparent disyllabic compounds in Mandarin, each playing a clear and immediately understandable role in conveying the compound's overall meaning. This process highlights a greater emphasis on meaning over sound than single words, as characters are chosen and combined based on their semantic contributions to the intended word. The unique features mentioned indicate potential variations in the cognitive mechanisms of Mandarin logographic processing. For instance, consider the compound word "xylophone," composed of the Greek words "xylo" (meaning wood) and "phone" (related to sound). However, this etymological origin may go unnoticed by those who do not speak Greek. In contrast, in Mandarin, the word for xylophone, 木琴 /mu4qin2/, clearly indicates its association with wood, thanks to the character 木 /mu4/. This immediate connection between characters and meanings in Chinese facilitates a more direct and intuitive understanding, enhancing lexical awareness compared to languages without such visual representations. In addition to the above characters of Mandarin transparent compounds, the distinction between opacity and transparency seems to have less impact in European languages because both opaque and transparent primes exhibit priming effects in compound production, as observed in the works of Koester & Schiller (2008) and Zwitserlood et al. (2000, 2002) (but also Zwitserlood, 2014). Nevertheless, this choice raises the possibility that the obtained results may differ from those seen in semantically transparent compounds, suggesting that future research endeavors could also consider investigating opaque compound words.

Furthermore, specific modifiers within the compounds examined in this study are abstract. It is worth noting that the present study did not implement control over the concreteness of these modifiers, emphasizing the need for future research to delve into the potential impact of concreteness on the mental lexicon's representation and processing. Additionally, the study specifically concentrated on the modifiers of compounds, manipulating only the first constituent. This opens avenues for further exploration into the distinctions between modifiers and heads within compounds.

Regarding the types of free and bound morphemes, we did not control for this factor, and a post-hoc analysis was conducted for synonym distractors on the two morpheme types, revealing a significant difference with $p = 0.017$. It is important to note that the stimuli for free synonym distractors and bound synonym distractors were not balanced in the present experiment, with free synonyms outnumbered by bound synonym distractors. This finding suggests implications for future studies further to investigate the potential differences between free and bound morphemes. Besides, there are orthographically radical overlaps between synonym distractors and target names in the present experiment, for example, 日 /ri4/ “sun” and 阳台 /yang2tai2/ “balcony.” A previous study (Wang et al., 2021) indicated no orthographic effect at the conceptual level of speech planning in Mandarin Chinese, and we also conducted a post-hoc analysis concerning orthographic radical overlap for synonym distractors, which revealed no significant difference ($t < 1$) for this potentially confounding factor. Another implication for future studies is related to the logographic writing system of Mandarin, where characters are the basic units of writing, which could cause potentially different results for illiterate speakers.

4.5 Conclusion

This study focused on the representation of Mandarin compounds in language production at the lemma level. The results indicated that synonyms as distractors yielded a facilitation effect, while semantically related distractors resulted in significantly longer latency. These findings imply that the separate lemmas of constituents are represented in our mental lexicon alongside the holistic single lemma of the compounds, supporting the concept of a hybrid lemma representation of Mandarin compounds and further supporting the decompositional model.

Appendix

4.A Experimental stimuli

Target		Distractors		
Stimuli	Constituents meaning	Synonym	Semantical-related	Unrelated
公鸡 rooster	male+chicken	雄 male	鹅 goose	风 wind
斑马 zebra	point+horse	点 point	熊 bear	云 cloud
口哨 whistle	mouth+whistle	嘴 mouth	绳 rope	雨 rain
乌龟 turtle	black+turtle	黑 black	蛇 snake	雪 snow
音箱 speaker	voice+box	声 voice	鼓 drum	猫 cat
内脏 viscera	internal+organ	里 inside	肺 lung	镜 mirror
牙医 dentist	tooth+doctor	齿 tooth	医院 hospital	火 fire
目镜 eyepiece	eye+galss	眼 eye	标本 specimen	杯 cup
毛笔 writing	hair+pen	发 hair	纸 paper	勺 spoon
屋顶 roof	room+top	室 room	门 door	鱼 fish
洋葱 onion	foreign+onion	外 outside	蒜 garlic	鞋 shoe
皇冠 crown	royal+cap	帝 emperor	权力 power	碗 bowl
阳台 balcony	sun+platform	日 day	窗 window	滑板 skateboard

面具 mask	face+tool	脸 face	戏剧 drama	草 grass
奶嘴 nipple	milk+mouse	乳 milk	婴儿 baby	药 medicine
相框 photo	photo+frame	容 allow	墙 wall	猪 pig
袋鼠 kangaroo	bag+rat	包 bag	考拉 koala	裙 skirt
衣架 coat hanger	cloth+hanger	服 clothes	裤 pants	磁铁 magnet
恶魔 demon	evil+devil	坏 bad	鬼 ghost	锁 lock
脚注 footnote	foot+note	足 foot	书 book	狗 dog
河马 hippo	river+horse	川 river	鹿 deer	钉 nail
画架 easel	paint+shelf	图 picture	笔 pen	假发 wig
城堡 castle	city+fort	市 city	公主 princess	石头 stone
菜单 menu	dish+list	肴 dish	餐馆 restaurant	拉链 zipper
圆规 compass	round+regulation	圈 round	尺 ruler	腿 leg
话筒 microphone	speech+tube	语 language	歌手 singer	桌 table
壁虎 gecko	wall+tiger	墙 wall	蜘蛛 spider	轮 wheel
路灯 street lamp	road+lamp	道 road	桥 bridge	腿 leg
恐龙 dinosaur	fear+dragon	惧 fear	鳄鱼 crocodile	椅 chair
山羊 goat	mountain+sheep	峦 mountain	牛 ox	贝壳 shell
刺猬 hedgehog	thorn+hedgehog	芒 mango	松鼠 squirrel	房子 house

国王 king	country+king	邦 state	王后 queen	天平 balance
青蛙 frog	green+frog	绿 green	蟾蜍 toad	灯 lamp
木偶 puppet	wood+puppet	树 tree	舞台 stage	孔雀 peacock
皱纹 wrinkle	wrinkle+texture	褶 pleats	老人 elder	信 letter
闹钟 alarm clock	noisy+clock	吵 noisy	表 watch	星 star
明星 star	bright+star	亮 bright	红毯 redarpet	球场 court
软管 hose	soft+tube	柔 soft	水 water	心 heart
硬币 coin	hard+coin	坚 firm	钱 money	蚂蚁 ant
空竹 diabolo	empty+bamboo	虚 virtual	表演 performance	水壶 kettle
喜鹊 magpie	happy+bird	乐 happy	鹰 eagle	刀 knife
单杠 horizontal	single+bar	独 alone	体操 gymnastics	叉 fork
棒球 baseball	stick+ball	棍 stick	手 hand	地图 map
首都 capital	head+city	头 head	故宫 forbidden	盐 salt
快递 express	fast+delivery	迅 fast	卡车 truck	咖啡 coffee

4.B Model parameters: reaction times

Table 4.B.1: Specification of best-fit model for response times (RTs) for picture naming task (n = 36). Note that estimates are reported in milliseconds.

Formula: RT ~ Condition + Trial + (1 | Subject) + (1 | Item)				
Term	Estimate	[95% CI]	t	p-value
(Intercept)	852.757	[833.606, 871.908]	87.273	< 0.001 ***
Condition: semantically related	37.739	[29.576, 45.902]	9.061	< 0.001 ***
Condition: control	18.806	[9.790, 27.822]	4.088	0.001 ***
Trial	2.911	[-1.352, 7.174]	1.338	0.181
Random effects				
σ^2	0.033			
τ00Item	995			
τ00Subject	1548			
ICC	1.00			
NSubject	36			
NItem	45			
Observations	4537			
Marginal R^2	0.088			
Conditional R^2	1.000			

Chapter 5 Concreteness effect of constituents in Mandarin compounds

5.1 Introduction

According to the language production model proposed by Levelt et al. (1999), producing a word involves conceptual preparation, lexical access, phonological encoding, and articulation. Conceptual representations become activated when we engage in speech production, which then spreads to lexical representations linked to that concept. Before words can be articulated, phonological information is retrieved, involving word form encoding. Although this theory was initially developed based on single-word production, it has since been extended to account for the production of phrases and morphologically complex words, including compound words.

Two prominent models address the lexical representation of compound words during production: full-listing models (Butterworth, 1983; Caramazza, 1997; Dell, 1986) and decompositional models (Levelt et al., 1999), with hybrid models somewhat scarcer. According to full-listing models, compound words are stored holistically in the mental lexicon. In contrast, decompositional models suggest that compounds are represented and processed through their constituents. Both models are supported by empirical studies. For instance, to investigate the representation of compound words, Janssen et al. (2008)

found that overall word frequency, rather than constituent frequency, influenced naming latencies in Mandarin compound words, supporting the full-listing hypothesis in language production. This was further supported by Bi et al. (2007), who showed that compound word frequency, not constituent morpheme frequency, affected the production performance of Chinese patients with aphasia. Additionally, Chen and Chen (2006) found that in Mandarin, naming latencies were not sensitive to constituent frequency, which aligned with the idea of a single lexical level between semantics and phonology.

In contrast, other studies support the decompositional model of compound word production. For instance, Roelofs (1996) and Bien et al. (2005) demonstrated that morpheme frequency and the morphological structure of compounds played a significant role in speech production planning, indicating that compound words are decomposed into their morphemes during production. Further evidence for decompositional model comes from studies using long-lag priming paradigms (Kaczer et al., 2015; Koester & Schiller, 2008, 2011; Lensink et al., 2014; Verdonschot et al., 2012; Wang et al., 2024), which showed that morphological priming persisted over many trials, suggesting that morphological processing occurred at a distinct level from phonological or semantic processing.

The research summarized above explores the representation and production of compound words in the mental lexicon of various languages. However, there are no consistent findings regarding this question. Examining the concreteness and abstractness of constituents to further explore their roles in compound production could offer valuable insights. The concept of abstractness refers to ideas that are neither purely physically nor spatially constrained, making them generally more variable in their content across individuals and more challenging to associate with a single image than concrete concepts (Del Maschio et al., 2021).

Compared to abstract words, concrete words are easier to comprehend, faster to read aloud and to evaluate for meaningfulness, and easier to remember (Belmore et al., 1982; Gerhand & Barry, 2000; Lee & Federmeier, 2008; Paivio, 1991; Schwanenflugel, 1991; Schwanenflugel et al., 1988). For instance, a concrete word like “table” is processed more quickly and accurately than an abstract word like “honesty” at the behavioral level. This preferential processing of concrete over abstract words is known as the concreteness effect (Holcomb et al., 1999; Kounios & Holcomb, 1994; West & Holcomb, 2000). Extending from single word to compounds processing, the concrete/abstract constituents could influence the retrieval of the compound words. For instance, consider the concrete compound words “church bell” and “apple core.” In “church bell,” both two constituents “church” and “bell” are concrete, whereas in “apple core,” the constituent “core” is relatively abstract. Therefore, the present study aimed to use the concreteness effect of constituents to investigate their influence on the retrieval of compound in Mandarin production.

There are studies working on concreteness effect in language production. For instance, research conducted by Hanley et al. (2013) used different paradigms to investigate the concreteness effect during language production. He designed two experiments to examine the impact of concreteness on word retrieval within sentence contexts. In Experiment 1, participants were asked to generate words from dictionary definitions. Results showed that abstract words were more challenging to retrieve, leading to more omissions and alternate responses than concrete words. Participants also experienced more tip-of-the-tongue (TOT) states when retrieving abstract words, indicating more phonological retrieval difficulties. In Experiment 2, participants generated words to complete sentences describing specific events. The number of abstract words recalled was significantly higher than that of concrete words in this task. The above literature showed that concreteness effects existed during language production and could serve as a robust method to investigate the research question in the present study.

Additionally, the neural mechanisms underlying the differences between concreteness and abstractness have been explored through the performance of patients with neural system damage, who often perform better on concrete than abstract words. For instance, in a study by Catricalà et al. (2014), patients with Alzheimer’s disease (AD) and the semantic variant of primary progressive aphasia (sv-PPA) completed tasks with controlled abstract and concrete stimuli. Results indicated that sv-PPA patients performed better with abstract than concrete concepts, showing category-specific effects: emotion concepts were preserved in AD, whereas social relations were selectively impaired in sv-PPA. Occasionally, AD patients displayed a living-non-living dissociation. These findings suggested that semantic memory disorders led to distinct patterns in abstract and concrete domains, highlighting differences in the brain regions affected by each condition.

Though much research has demonstrated that concrete expressions were processed more efficiently than abstract expressions across various languages and cognitive tasks, most of these studies have been conducted in language comprehension. Work has seldomly been done in language production. One primary reason could be that it is hard in production tasks, for instance, picture naming, to present and name pictures of abstract concepts. Hanley et al. (2013), for instance, addressed this limitation using indirect methods, embedding concrete and abstract words within sentence contexts.

Given the above, the present study aimed to investigate the concreteness effect of constituents in Mandarin compound words during the production process by using two sets of concrete compounds as targets in a picture-naming task. At the same time, the concreteness of their constituents was manipulated to explore whether constituent concreteness affects Mandarin compound production while controlling for the overall concreteness of the compounds. Two conditions were created: the “aa” condition, where both constituents of the concrete compound were abstract, and the “cc” condition, where both constituents were concrete. It was predicted that naming latencies would be shorter for the “cc” condition compared to the “aa” condition due to the concreteness effect of the constituents at the behavioral level mentioned in the present study.

5.2 Methodology

5.2.1 Participants

Forty native right-handed Mandarin speakers, including six males, were recruited from Leiden University. The mean age was 24.05, with an SD of ± 2.31. All participants were from Mandarin-speaking provinces in China and spoke Mandarin as their mother tongue. Participants who had been living in the Netherlands for less than two years were included, while those who had been residing in the Netherlands for longer than two years were excluded from recruitment due to potentially higher proficiency in English and Dutch. All participants had normal or corrected-to-normal vision and received monetary compensation for their participation.

This study was approved by the Faculties of Humanities and Archaeology ethics committee at Leiden University (acceptance number: 2022/09). At the time of testing, none of the participants reported color blindness, learning disorders, hearing or visual impairments, or psychological or neurological conditions. Participants read an information sheet and provided informed consent by signing a consent form before the study began.

5.2.2 Materials

In this study, we examined two conditions: the abstract (aa) condition, where the compound words were concrete, but their two constituents were abstract, and the concrete (cc) condition, where both the compound words and their constituents were concrete. Hence, only concreteness of the constituents varied between conditions. For example, in the cc condition, 河马 (/he2ma3/ "hippo") is a concrete compound word with two concrete constituents: 河 (/he2/ "river") and 马 (/ma3/ "horse"); in the aa condition, 乐队 (/yue4dui4/ "band") is a concrete compound word with two abstract constituents: 乐 (/yue4/ "music") and 队 (/dui4/ "team"). It is important to note that all target stimuli in both conditions were concrete disyllabic noun compounds. Forty-two Mandarin disyllabic compound nouns were selected as stimuli, with twenty-one words assigned to each condition.

We first calculated the concreteness and word frequencies of our stimuli using two corpora to assess the concreteness and word frequency of the entire compound words to ensure no significant differences in concreteness or word frequency between the two conditions.

For this, we utilized the MELD-SCH corpus, which provides concreteness and abstractness ratings for 9,877 two-character Mandarin Chinese words (Xu & Li, 2020), and SUBTLEX-CH corpus (Cai &

Brysbaert, 2010), which provides word frequencies. We controlled for the concreteness of the compounds (t (40) < 1) based on MELD-SCH, as well as the compound frequency (t (40) < 1) using the SUBTLEX-CH corpus (see Table 5.1).

Table 5.1: The concreteness of the compound words.

Corpus	Condition	Compound word	
		Frequency	Concreteness
MELD-SCH	aa	2.46	1.47
	cc	2.36	1.42
	p-value	0.45	0.18

To control for the concreteness and word frequency of the constituents, we conducted a cross-corpus comparison using two corpora. One was the Hong Kong Chinese Character Psycholinguistic Norms corpus (Su et al., 2023), which provides ratings for 4,376 individual Chinese characters on various factors, including word frequency, age of acquisition, familiarity, imageability, and concreteness.

This corpus allowed for more comprehensive control of the variables related to the constituent characters. The other corpus we used was from Liu et al. (2007), which provides word naming and psycholinguistic norms in Chinese, including information on familiarity, imageability, frequency, age of acquisition, and concreteness. By utilizing both corpora, we ensured a thorough control of the concreteness and word frequency of the constituents in our study.

We then ensured that each constituent's concreteness was carefully controlled, confirming that the mean concreteness levels of constituents and the concreteness levels of each constituent differed significantly between the conditions. Furthermore, we controlled for each constituent's frequency to ensure no significant differences across conditions (see Table 5.2). In addition, we controlled for other sub-lexical variables of the constituents, including stroke count (t (40) < 1) and imageability (t (40) < 1), ensuring that these factors did not significantly differ across the two conditions in the present study. Furthermore, the concreteness of the selected compounds and their constituents were rated by a group of 20 participants online, and their ratings (t (40) < 1) confirmed the validity of these chosen stimuli. See *Appendix 5.A* for detailed stimuli list.

Table 5.2: The concreteness of the constituents.

Corpus	Condition	1st		2nd		Mean	
		Frequency	Concreteness	Frequency	Concreteness	Frequency	Concreteness
Su et al. (2023)	aa	2.39	4.04	2.36	4.27	2.38	4.16
	cc	2.42	6.35	2.08	6.01	2.25	6.18
	p	0.85	<0.001	0.12	<0.001	0.31	<0.001
Liu et al. (2007)	aa	3.76	4.75	3.86	5.33	3.81	5.03
	cc	3.92	6.58	3.61	6.43	3.76	6.50
	p	0.30	<0.001	0.10	<0.001	0.67	<0.001

Furthermore, we employed 42 colored pictures, with 21 pictures assigned to each condition in the present study. Thirty-four of these pictures were selected from the Multipic corpus (Duñabeitia et al., 2018), while eight were created by an illustrator in a similar style. The name agreement (t (40) < 1) and picture complexity (t (40) < 1) of these selected pictures were evaluated by another 20 participants online, following the method outlined by Snodgrass and Vanderwart (1980). In addition, we included ten filler pictures, which were used as warm-up trials at the beginning of each block, and as filler pictures during the experiment, with their frequencies and concreteness spanning a wide range. Participants were presented with 52 pictures, including 42 experimental and 10 filler pictures.

5.2.3 Design

The experiment in the present study employed a one-factorial within-subject design, with Concreteness as the fixed factor and Subject and Item as random factors. Participants engaged in a picture-naming task, where a pseudorandomized design was applied. Two versions of the pseudorandomized stimuli list were created during the experiment, with the first half using the first version and the other half using the second version. Pictures from the same category or those with the same phonological onset were not shown consecutively in the pseudorandomized design to avoid priming effects.

5.2.4 Procedure

The experiment was presented using *E-prime 3.0* (Psychology Software Tools) and was conducted in a soundproof booth. Participants were seated in front of a computer in a dimly lit room. They used a microphone connected to a Chronos response device containing a voice key. The experiment consisted of three phases.

First, participants were given 10-15 minutes to familiarize themselves with the target pictures and their names by studying a booklet at their own pace. Subsequently, the experimenter assessed whether participants correctly remembered the picture names by conducting a practice session through another booklet without the picture names. In this session, participants were asked to name all the target pictures as soon as possible, and the experimenter would correct them if they misnamed any pictures.

In the experiment session, each trial began with a fixation cross displayed for 250 ms, followed by a blank screen for another 250 ms. Then, the target picture appeared in the center of the screen for 2,000 ms. Therefore, each trial lasted 2,500 ms, which was applied to all target and filler pictures. See Figure 5.1 for a detailed demonstration of the experiment trial in the present study.

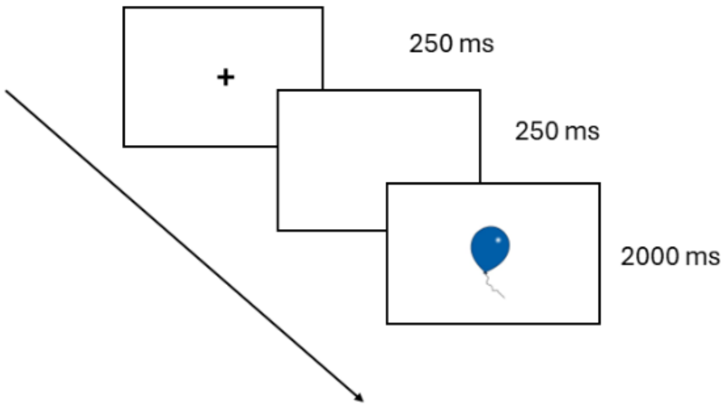


Figure 5.1: A trial sequence for the picture-naming task.

5.3 Data analysis and results

5.3.1 Behavioral data exclusion

Error trials (1.49%) and outlier trials (3.10%) with reaction times that deviated more than 2.5 SDs from the mean per participant per condition were eliminated for further RT analysis.

5.3.2 Behavioral data analysis

Behavioral data were analyzed using *RStudio* Version 4.2.2. We first calculated descriptive statistics for naming latencies for each condition (see Table 5.3 and Figure 5.2). Then, we employed a generalized linear mixed effect model (GLMM) using the *glmer ()*

function from the *lme4* package with inverse Gaussian errors to model positively skewed RT data.

Table 5.3: Mean naming latencies (only correct trials included) for each condition and each repetition level (n = 40)

Concreteness	Naming latencies (ms)	
	Mean	SD
aa	811.90	152.22
cc	770.63	123.91

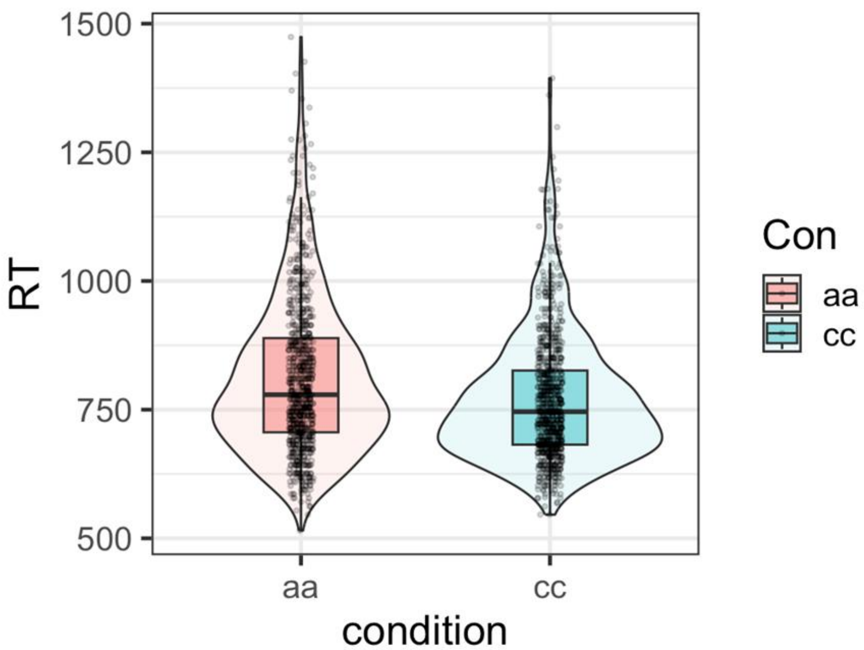


Figure 5.2: Mean naming latencies for each condition (n = 40).

To avoid over-parameterization and achieve a balance between Type-I error and statistical power, we employed a backward elimination strategy for the random effects structures of the model (Bates, 2015; Matuschek et al., 2017). Model comparisons and likelihood ratio tests were conducted at each step using the *anova()* function, evaluating Akaike's Information Criterion (AIC) (Akaike, 1974), Bayesian Information Criterion (BIC) (Neath & Cavanaugh, 2012), and log-likelihood to determine whether the addition of a new factor significantly improved the model.

In this analysis, *Concreteness* was included as a fixed effect, while *Subject* and *Item* were introduced as random effects in the study. For naming latencies, the model of the best fit was *RT* ~ *Concreteness* + (1 + *Concreteness* | *Subject*) + (1 | *Item*). The results showed that the naming latencies of the *aa* and *cc* conditions significantly differed with $\beta = -41.85$ (95% CI [-57, -26.7]), SE = 7.73, $t = -5.42$, and $p < 0.001$.

5.4 Discussion and conclusion

The ongoing debate regarding the representation of compound words in our mental lexicon has led to numerous studies exploring this issue. The present study aimed to address this question by examining the concreteness effect of word constituents in compounds. Two conditions were established: "cc" condition representing concrete compound words with two concrete constituents and "aa" condition representing concrete compound words with two abstract constituents.

The behavioral results in the present study showed significantly faster reaction times for the cc condition than the aa condition, suggesting concreteness effects of the constituents in Mandarin compound production. The results reported in the present study have implications for the two models of lexical representation discussed in the Introduction. According to the full-listing model of lexical representations, compounds are stored in their full-listing format at the lexical level. In other words, the whole word is retrieved when naming the pictures instead of the individual constituents, which predicts no constituent effects.

Contrastingly, the decompositional model predicts constituent effects due to the fact that each constituent is retrieved during compound production. Therefore, the present results support the decompositional hypothesis regarding Mandarin compound production.

The present findings offer insights into models of lemma representation (Levelt et al., 1999; Marelli et al., 2012; Sprenger et al., 2006) of compound words. According to Levelt's model (1999), compounds are initially represented as a single lemma, which is later decomposed at the lexeme level for phonological encoding. However, the concreteness effect observed in this study suggests that distinct lemma entries may exist for individual constituents during the production of Mandarin compound words. The fact that the concreteness of the constituents influences the production of compounds in the present study supports the hybrid lemma account proposed by Marelli et al. (2012) and Sprenger et al. (2006).

The concreteness effect observed at the lemma level in the present study went against previous studies in Indo-European languages supporting a two-staged model (Mondini et al., 2004) with only a single lemma for word production. This difference could stem from cross-linguistic distinctions. First, Chinese writing is classified as a logographic system, where characters, the fundamental writing units, are composed of numerous strokes organized into a square shape (Tan et al., 2001). Chinese speakers may exhibit heightened lexical awareness attributed to the presence of Chinese characters and radicals.

These visual components serve as immediate reminders of the meanings of words, potentially contributing to a more pronounced awareness of lexical items in Chinese than in languages lacking such visual cues. In addition, the pronunciation of Chinese characters is established at the syllable/morpheme level, requiring repetitive rote memorization to connect visual forms with corresponding sounds. Learning is often aided by sub-character units, authentic characters in their own right. Unlike alphabetic languages, Chinese characters do not directly correspond to phonemes but map onto meaningful morphemes in spoken language. This aspect implies that the regular or quasi-regular grapheme-phoneme conversions commonly found in alphabetic languages are not feasible in Mandarin Chinese (Plaut, 1996; Tan & Perfetti, 1998).

An alternative explanation for the concreteness effect in language production is that abstract words have weaker semantic-lexical association than concrete words (Hanley et al., 2004; 2013).

Hanley and colleagues (2004) used Foygel and Dell's (2000) speech production model to simulate imageability differences by varying semantic-lexical connection strengths. They posited that high-imageability words have stronger semantic-lexical association, while low-imageability words have weaker ones. In this model, lower association results in weaker activation of lexical units, causing more alternates and omissions for low-imageability words.

In a definition task by Hanley et al. (2013), participants rated the most common alternatives to abstract target words as more compatible with definitions than those for concrete words, supporting this semantic-lexical weights hypothesis. On the other hand, Newton and Barry (1997) argued that abstract words are harder to retrieve in word production tasks because they face stronger competition from semantically similar words. They proposed that retrieval issues for abstract words arise after activating the target word's semantic representation. Since abstract words typically share many semantic features with other words, phonological representations of competing words are more likely to reach the activation threshold, increasing semantic errors.

Although concrete words also activate competing items, their distinctive semantic features make them more likely to reach the activation threshold over competitors in tasks like word definition, reducing the likelihood of errors. Besides, Hoffman and colleagues (2011) found that abstract words are more ambiguous and have more senses than concrete words, as demonstrated through latent semantic analysis. The imageability effect observed in their study suggested that abstract words may be more difficult to process because they depend more on context for meaning, which is often lacking in experimental settings (Schwanenflugel, Harnishfeger, & Stowe, 1988).

These assumptions extend beyond the single-word level to the level of morphemes (constituents) in compound production. In the present study, the facilitation of concrete constituents showed support for the claim that it is easier for concrete concepts which have more distinctive semantic features and higher imageability to reach the activation threshold. Together with these studies above, the present study highlights the importance of concreteness in language production, suggesting that future research could explore how concreteness influences language production more broadly.

In addition, it is essential to highlight that the present study did not control for the semantic transparency. The distinction between opacity and transparency seems to have less impact in Indo-European languages based on previous literature (e.g., Koester & Schiller, 2008; but also see Zwitserlood, 2014). Both opaque and transparent primes exhibit priming effects in compound production, as observed in the works of Koester and Schiller (2008). Nevertheless, there is a possibility that the obtained results may differ when applying to transparent or opaque words, suggesting that future research endeavors could also consider investigating opaque compound words in their research design.

Furthermore, although we selected noun compounds as stimuli, we did not control for the word category (e.g., noun, verb) of their constituents. Hanley et al. (2013) highlighted the effects of word category on lexical retrieval in their two experiments. Therefore, future research could investigate this aspect further to provide more detailed evidence addressing the role of word class in compound representation.

In conclusion, the present study's results demonstrated that the concreteness of constituents probably played a role at the lemma level in Mandarin compound production and further supported the decompositional account instead of the full-listing account of Mandarin compounds.

Appendix

5.A Experimental stimuli

Stimuli	Pinyin	English	Condition	English of each constituent
信封	xìnfēng	envelope	aa	信(letter) + 封 (seal)
公主	gōngzhǔ	princess	aa	公 (public) + 主 (master)
囚犯	qiúfàn	prisoner	aa	囚 (prisoner) + 犯 (criminal)
国王	guówáng	king	aa	国 (country) + 王 (king)
士兵	shìbīng	soldier	aa	士 (warrior) + 兵 (soldier)
彩虹	cǎihóng	rainbow	aa	彩 (color) + 虹 (rainbow)
戒指	jièzhǐ	ring	aa	戒 (warn) + 指 (finger)
抽屉	chōutì	drawer	aa	抽 (pull) + 屉 (drawer)
拼图	pīntú	puzzle	aa	拼 (piece together) + 图
教堂	jiàotáng	church	aa	教 (teach) + 堂 (hall)
日历	rìlì	calendar	aa	日 (day) + 历 (calendar)
珍珠	zhēnzhū	pearl	aa	珍 (precious) + 珠 (pearl)
皇冠	huángguàn	crown	aa	黄 (emperor/royal)
算盘	suànpán	abacus	aa	算 (calculate) + 盘 (plate/tray)
衬衫	chènshān	shirt	aa	衬 (lining) + 衫 (shirt)

西装	xīzhuāng	suit	aa	西 (west) + 装 (clothes)
贺卡	hèkǎ	greeting card	aa	贺 (congratulate) + 卡 (card)
邮差	yóuchāi	postman	aa	邮 (mail) + 差 (messenger)
阳台	yángtái	balcony	aa	阳 (sun) + 台 (platform)
音响	yīnxiǎng	audio	aa	音 (sound) + 响 (loud)
领带	lǐngdài	tie	aa	领 (collar) + 带 (belt)
冰箱	bīngxiāng	refrigerator	cc	冰 (ice) + 箱 (box)
坟墓	fénmù	grave	cc	坟 (grave) + 墓 (tomb)
天鹅	tiān'é	swan	cc	天 (sky/heaven) + 鹅 (goose)
手掌	shǒuzhǎng	palm	cc	手 (hand) + 掌 (palm)
树叶	shùyè	leaves	cc	树 (tree) + 叶 (leaf)
楼梯	lóutī	stairs	cc	楼 (building) + 梯 (ladder)
水果	shuǐguǒ	fruit	cc	水 (water) + 果 (fruit)
河马	hémǎ	hippo	cc	河 (river) + 马 (horse)
海鸥	hǎi'ōu	seagull	cc	海 (sea) + 鸥 (gull)
火箭	huǒjiàn	rocket	cc	火 (fire) + 箭 (arrow)
灯塔	dēngtǎ	lighthouse	cc	灯 (lamp) + 塔 (tower)
牛奶	niúnǎi	milk	cc	牛 (cow) + 奶 (milk)

眼镜	yǎnjìng	glasses	cc	眼 (eye) + 镜 (lens)
花瓶	huāpíng	vase	cc	花 (flower) + 瓶 (bottle)
药丸	yàowán	pill	cc	药 (medicine) + 丸 (pill)
蛋糕	dàngāo	cake	cc	蛋 (egg) + 糕 (cake)
雨伞	yǔsǎn	umbrella	cc	雨 (rain) + 伞 (umbrella)
雪山	xuěshān	snow mountain	cc	雪 (snow) + 山 (mountain)
风车	fēngchē	windmill	cc	风 (wind) + 车 (vehicle)
鸟笼	niǎolóng	birdcage	cc	鸟 (bird) + 笼 (cage)
龙虾	lóngxiā	lobster	cc	龙 (dragon) + 虾 (shrimp)

Chapter 6 General discussion

This thesis examined the morphological encoding and the representation of Mandarin disyllabic noun compounds in the mental lexicon during language production. Four experiments were conducted using different experimental designs and manipulating different factors at both behavioral and electrophysiological levels. Two primary models were tested: the decompositional model and the full-listing model. In this chapter, I first summarize the main research questions and key findings from each experiment, discussing their relevance and contributions to the existing literature. Then I integrate these findings from each experiment to provide a broader perspective in terms of their theoretical and methodological implications and then propose directions for future research. Additionally, I address unresolved questions and conclude by discussing the potential limitations of this thesis.

6.1. Discussion of chapter findings

Chapter 1 provides a general introduction to the topic and research questions investigated in my thesis and discusses the theoretical models employed in the following empirical chapters.

In **Chapter 2**, a long-lag priming experiment examined the morphological encoding of Mandarin disyllabic compound words. The study recruited thirty-six native Mandarin speakers who were asked to name pictures displayed on a screen, with both behavioral and electrophysiological data being collected and analyzed. The primary aim was to investigate whether morphological priming would occur in this context, as observed in Indo-European languages. Surprisingly, the behavioral results revealed no facilitation in picture naming when the targets were preceded by morphologically related primes, which contrasted with previous research on Indo-European languages. Specifically, there was no significant difference in reaction times (RTs) for target naming between morphologically related and unrelated primes in the first or second constituent positions. However, the first-position primes did elicit shorter RTs than the second-position primes, suggesting a potential but weak position effect. These results challenged the idea that morphological priming for individual constituents occurred at the behavioral level during the production of Mandarin disyllabic compound words. However, the electrophysiological data provided further insight, with event-related potential (ERP) analysis revealing a reduced N400 response for morpheme-related primes compared to unrelated primes. This N400 reduction suggested that morphological priming occurred at the encoding level, supporting the decompositional model (Levelt et al., 1999), and aligned with similar findings in prior studies (e.g., Koester & Schiller, 2008). The morphological priming effect in Mandarin may have occurred because disyllabic compounds in Mandarin are stored in the mental lexicon in terms of separate constituent representations rather than as whole-word forms.

An additional finding of this study was the effect of picture repetition across experimental blocks, which led to faster picture naming latencies. This repetition effect was observed in both behavioral and ERP data and was consistent with previous research (Zwistserlood et al., 2000, 2002). However, the repetition effect was independent of the linguistic processes of interest, as it did not interact with the prime type. Regarding the position effect, the RTs showed that naming latencies were shorter for first-position constituents than for second-position constituents, although this difference was not statistically significant. However, the ERP data found a significant interaction between prime relatedness and position, with a larger N400 reduction for first-position constituents than for second-position constituents. This supported the idea that the first constituent of a compound word was processed differently from the second and that priming effects were stronger for the first position. These findings are consistent with the serial planning hypothesis, which means the words are processed and planned in a sequential order. However, the exact role of the head constituent in compound production needs to be further explored in future research.

In conclusion, the results of this long-lag priming experiment suggested that morphologically related primes did not facilitate the picture naming of target morphemes at the behavioral level compared with the morphologically unrelated conditions in Mandarin disyllabic compounds. However, the ERP data indicated that morphological priming occurred for the first position morphological overlaps during the morphological encoding process, as evidenced by the N400 effect. This supported a decompositional model of lexical representation, where the two constituents of a compound word are stored and represented separately. Although these findings provide insights into the processing of Mandarin compound words, further research is needed to address unresolved questions, such as the influence of morpheme type, such as bound/free morpheme and the role of the head constituent in compound word production.

Chapter 3 examined how compound words were represented in the mental lexicon, focusing on morpheme and compound frequency effects in Mandarin compound production. Previous research has produced conflicting views on whether compound words are represented in a decomposed form (Bien et al., 2005; Koester & Schiller, 2008, 2011; Levelt et al., 1999) or stored as whole units (Chen & Chen, 2006; Janssen et al., 2008). To address this issue, the present study followed the design framework of Janssen et al. (2008), incorporating EEG to explore whether the production of Mandarin compound words was influenced by morpheme or compound frequency. The results of the present study showed that the reaction times (RTs) in picture naming tasks did not align with the earlier findings of Janssen and colleagues. Specifically, in this study, the L(h) condition (low word frequency but high morpheme frequency) exhibited similar RTs to the H(h) condition (high word and high morpheme frequency). In contrast, the L(h) condition was faster than the L(l) condition (low word and low morpheme frequency). The observed morpheme frequency effect aligned with previous studies (Bien et al., 2005; Chen & Chen, 2015; Koester & Schiller, 2008; Roelofs, 1996), lending support to the decompositional model proposed by Levelt et al. (1999), where morphemes play a critical role in word production. Discrepancies between the present study and Janssen et al.'s findings might stem from different corpora: Janssen and colleagues used the Modern Chinese Frequency Dictionary (1986), which is based on written materials; the present study relied on SUBTLEX-CH (Cai & Brysbaert, 2010), which is based on spoken language data derived from movie subtitles. Differences between written and spoken corpora may have influenced the frequency distributions of the stimuli, contributing to the different outcomes.

However, EEG data did not align with behavioral results. While RTs indicated a morpheme frequency effect, ERP data showed no significant differences between the three frequency conditions, suggesting that both the morpheme frequency effect and word frequency effect were not observed at the electrophysiological level. Additionally, the N2 component, typically linked to frequency effects in the early phases of speech production planning (Strijkers et al., 2010), appeared around 180 ms for L(h) and L(l) conditions, though statistical analyses showed no significant frequency differences. This mismatch between behavioral and ERP findings highlights the complexity of morphological processing in Mandarin and the need for further research to clarify these differences.

In conclusion, the behavioral findings suggested that Mandarin compound production was affected by individual morpheme frequency instead of whole-word frequency. This supported a decompositional model where compounds were retrieved based on their morphemes rather than as single units. However, the lack of frequency effects for EEG data indicated a complex relationship between behavioral and electrophysiological findings. Future research should delve into these frequency effects, considering language-specific factors such as semantic transparency and speaker variables, such as multilingual background, to deepen our understanding of morphological representation in Mandarin compounds and potential cross-linguistic differences in morphological processing.

Chapter 4 investigated the representation of Mandarin compound words at the lemma level in the mental lexicon. Using the picture-word interference (PWI) paradigm, native Mandarin speakers were asked to name pictures while ignoring visually presented distractor words. Behavioral data were analyzed across three conditions including (1) synonym distractors, (2) semantically related distractors, and (3) unrelated distractors as control condition. The findings of this chapter showed that naming latencies for synonym distractors were significantly faster than for the control condition, indicating facilitation effects consistent with previous studies (Dylman & Barry, 2018). In contrast, semantically related distractors resulted in slower latencies compared to controls, aligning with the inhibition effects documented in prior research (Levelt et al., 1999; Lupker, 1979; Roelofs, 1992; Starreveld & La Heij, 1995, 1996).

According to Levelt et al.'s (1999) model of lexical access, compounds are initially stored as single lemmas at the lemma level, with decomposition occurring at the lexeme level during phonological encoding. However, our study found a facilitation effect from constituent synonyms, supporting the idea of separate lemmas for each constituent in Mandarin compound production. This suggests that activation begins at the conceptual level for each constituent, followed by lemma selection that includes constituent synonyms. These findings have implications for the superlemma account of lexical representation for multi-lexical items (Sprenger et al., 2006; Kuiper et al., 2007).

Moreover, our study showed that synonym distractors were co-activated without interference, challenging competition-based theories. Picture-word interference tasks are often explained through lexical competition theories, which suggest that response selection is slowed down by competition among related words (e.g., Levelt et al., 1999). In this model, a distractor that activates a non-target word acts as a competitor, delaying the selection of the target word. Additionally, Finkbeiner, Gollan, and Caramazza (2006) suggested that "the closer two lexical representations are in meaning, the more difficult it will be to select the correct one," highlighting a "hard problem" of lexical selection that complicates synonym retrieval within a language. However, our findings contradict this assumption, challenging the competition-based account (Levelt et al., 1999).

In conclusion, this study provides evidence supporting a hybrid lemma account (Sprenger et al., 2006) for Mandarin compound word representation. The facilitation effect from synonyms indicates that both whole compounds and their constituents are represented as lemmas in the mental lexicon. Our findings suggest that the lemma representation of Mandarin compounds differs from that of Dutch or German compounds, likely influenced by the unique characteristics of the Mandarin language and its writing system. Further research is needed to explore the impact of semantic transparency, concreteness, and the distinction between free and bound morphemes.

In **Chapter 5**, we investigated the representation of compound words in the mental lexicon, with a particular focus on the role of concreteness in word constituents. We established two experimental conditions: the "cc" condition, consisting of compound words with two concrete constituents, and the "aa" condition, featuring compounds with two abstract constituents. Behavioral results indicated significantly faster reaction times for the "cc" condition than the "aa" condition, highlighting a concreteness effect in compound word production.

These findings have important implications for two models of lexical representation: the full-listing and decompositional models. The full-listing model posits that compound words are stored as whole units in the mental lexicon, predicting no influence from individual constituents. In contrast, the decompositional model suggests that compounds are retrieved via their constituents, expecting constituent-based effects. The observed concreteness effect of constituents aligns with the decompositional model, suggesting that each constituent is accessed during compound word production. Our results also support theories of the hybrid lemma account, as proposed by Sprenger et al. (2006). According to Levelt et al. (1999), compounds are stored as single lemmas at the lemma level but are decomposed into individual constituents at the lexeme level for phonological encoding. However, the observed concreteness effect of constituents implies that the lemma entries for individual constituents contribute to compound word production, supporting the hybrid lemma account over the single lemma account.

Additionally, these findings underscore the concreteness effect in language production. One possible explanation for this effect is that abstract words have weaker semantic-lexical associations than concrete words (Hanley et al., 2004; 2013), a phenomenon that may extend to the morpheme (constituent) level. This aligns with Newton and Barry's (1997) assertion that abstract words are harder to retrieve due to stronger competition from semantic neighbors. Through latent semantic analysis, Hoffman and colleagues (2011) also demonstrated that abstract words possess greater ambiguity and more senses than concrete words. It is important to note that this study did not consider the impact of semantic transparency. While transparency differences have shown minimal effects in some Indo-European languages (e.g., Koester & Schiller, 2008; but see Zwitserlood, 2014), future research should incorporate opaque compounds to explore these effects further.

In summary, this study provides evidence in favor of the decompositional model and a hybrid lemma representation for Mandarin compound words in production. Additionally, these results emphasize the significance of concreteness in language production, suggesting that future research should further explore how concreteness influences language production across different linguistic contexts.

6.2 Theoretical implications

The findings from four experiments in the present thesis contribute to understanding how Mandarin compound words are processed and represented in our mental lexicon. Our thesis offers theoretical implications for both the decompositional and full-listing model of lexical representation, which are introduced in the following section. Language production theories posit that the production of compound words unfolds as follows: naming an object begins with activating the relevant semantic concept, which simultaneously activates semantically related concepts through spreading activation. This leads to the activation of multiple semantically related lexical entries at the lemma level, which compete for lexical selection until one target lemma is chosen (Dell, 1986; Dell & O'Seaghdha, 1992; Garrett, 1980; Levelt, 1993; Levelt et al., 1999; Roelofs et al., 1996). Regarding the process of compound production, Levelt et al. (1999) suggest that compound words are stored in the lexicon based on their morphemes, e.g., the decompositional model. In contrast, Caramazza's (1997) full-listing model argues that compound words are stored holistically, requiring participants to retrieve the entire word form when naming a compound target.

In this thesis, the N400 effect found in Chapter 2 and the morpheme frequency effects observed in Chapter 3 support the decompositional model. In Chapter 2, an N400 effect was observed for the morphologically related prime condition compared to the unrelated prime condition. Given that the N400 component is sensitive to morphological processing, this finding strengthens the argument that the observed priming effects reflect morphological encoding in language production. In Chapter 3, morpheme frequency and compound frequency were manipulated, revealing that Mandarin compounds were influenced by morpheme frequency rather than compound frequency. This sensitivity to morpheme frequency highlights the role of constituents in the production of Mandarin compounds. Furthermore, Chapter 4 explored holistic and hybrid lemma representations to investigate holistic versus analytic processing in Mandarin compound word production. The facilitation effect observed for synonym distractors emphasized the role of constituents, supporting the decompositional model. Finally, Chapter 5 examined the concreteness effect, showing that concrete constituents were processed more quickly than abstract constituents during language production. This concreteness effect again demonstrated the influence of individual constituents, reinforcing support for the decompositional model.

Another theoretical implication is about the lemma representation of Mandarin compounds in production. The decompositional model proposed by Levelt et al. (1999) suggests that decomposition during the processing of compounds can occur at any level of lexical selection, allowing for a hybrid lemma representation. There are two main models regarding lemma representation: the two-stage model (Levelt et al., 1999), which posits that only a single lemma is represented, and the hybrid lemma account proposed by Sprenger et al. (2006), which holds the view that the lemma of constituents is represented alongside with the single lemma of compounds. Levelt and colleagues (1999) proposed that compound words are stored holistically at the lemma level, accessing multiple morphemes at the word-form level. After selecting a single lemma, the constituent morphemes are retrieved for phonological encoding before final articulation (Indefrey, 2011; Indefrey & Levelt, 2004). According to this model, compounds are initially stored as single lemmas and decomposed into constituent morphemes at the lexeme level. However, the hybrid lemma model (Sprenger et al., 2006) introduces the possibility of decompositional compound representations, positing that individual constituents coexist with holistic lemmas.

This hybrid lemma account was supported by the results from Chapters 4 and 5, where the synonym distractor effect and concreteness effect demonstrated that semantic features, such as synonymy and concreteness, significantly influenced lexical access and retrieval during language production. These findings indicated that the individual constituents of compound words were stored and retrieved separately at the lemma level in Mandarin compound production.

Specifically, Chapter 4 explored the impact of synonyms on language production, revealing facilitation effects from synonym distractors in compound word naming. This finding suggested that the semantic meaning of the first compound constituent was activated, highlighting the potential role of synonymy of constituents in lexical processing, especially in Mandarin. Chapter 5 revealed a concreteness effect in the constituents of Mandarin compounds. This concreteness effect supported hybrid lemma representation models (Marelli et al., 2012; Sprenger et al., 2006). According to Levelt et al. (1999), compounds are represented as single lemmas that later decompose for phonological encoding. However, the observed concreteness effect suggested separate lemma entries for individual constituents, aligning with the hybrid lemma model.

Besides, traditional lexical competition theories (Levelt et al., 1999) posit related words slow target selection due to competition. However, the results from Chapter 4 found no interference from synonymous distractors, challenging the selection-by-competition model. Moreover, the question of concrete and abstract processes has also insights into lexical selection theories. Abstract words might have weaker semantic-lexical association than concrete words (Hanley et al., 2004; 2013), facing greater competition and ambiguity (Hoffman et al., 2011), complicating retrieval. These findings highlighted concreteness as a critical factor in lexical selection during language production and suggested avenues for further research into its broader effects.

In addition, the findings in the present thesis also provided implications for the cross-linguistic differences between Mandarin and Indo-European languages. Research has suggested that Mandarin's limited morphological priming effect could stem from its relatively simple morphological system, which may reduce reliance on morphological representations during word production compared to morphologically complex languages like Dutch (Janssen et al., 2008; Chen & Chen, 2006). Additionally, using characters and radicals, the Chinese writing system may heighten lexical awareness among speakers through visual cues. As a logographic language, Chinese relies on characters of multiple strokes in square shapes (Tan et al., 2001), which serve as cues to word meanings. Besides, unlike alphabetic languages, where grapheme-phoneme connections are more direct, Chinese characters are based on syllables or morphemes, requiring memorization to link visual forms with pronunciation. This distinct structure, in which characters represent meaningful morphemes rather than individual phonemes, may influence how compound words are mentally represented and retrieved (Plaut, 1996; Tan & Perfetti, 1998).

In summary, the findings of this thesis provide valuable theoretical insights into language production theories, especially regarding decompositional versus full-listing models and the single lemma versus hybrid lemma accounts. Additionally, the results illuminate the roles of synonymy and concreteness effects in lexical selection during the production of Mandarin compounds. Potential cross-linguistic factors are also discussed in this section to highlight differences between Mandarin and Indo-European languages, offering a broader perspective on language processing across linguistic systems.

6.3 Methodological implications

This thesis has significant implications for the long-lag priming paradigm. The long-lag priming approach is well-established and robust, particularly in exploring the morphological representation in the mental lexicon, yielding consistent and replicable results across various experimental methods (e.g., behavioral, electrophysiological, and hemodynamic) and languages (Schiller & Verdonschot, 2018). Previous studies have shown that morphological priming, where the prime is morphologically related to the target, remains effective even with many intervening trials. In contrast, effects related to semantic and phonological priming tend to diminish at longer lags. These findings suggest that priming occurs at a distinct morphological level rather than at the phonological or semantic levels.

In this thesis, we used a long-lag paradigm in Chapter 2 to examine the morphological priming effect. Our results indicate that the long-lag paradigm is not sensitive enough at the behavioral level to capture morphological priming in a language with less morphological richness, such as Mandarin, compared to Dutch or German. While the morphological priming effect is robust in the long-lag paradigm for Dutch and German, it does not hold for Mandarin, highlighting the limitations of this paradigm in languages with simpler morphological structures. Combining the long-lag paradigm with EEG could be a useful supplementary approach for investigating languages with less morphological richness. In Chapter 2, the behavioral results did not show a significant priming effect, but the ERP data successfully captured these subtle effects with its higher temporal resolution, as indicated. The morphological N400 effect observed can be attributed to morphological relatedness between the primes and targets, as semantic/phonological effects are typically short-lived and do not persist across the long-lag paradigm.

In addition, we utilized a single-trial approach for the EEG data analysis in Chapters 2 and 3 and employed linear mixed model (LMM) analysis to examine voltage amplitudes within predefined regions of interest (ROIs) and time windows. This method is particularly advantageous for handling unbalanced datasets, such as those with varying numbers of observations across participants or experimental conditions. LMM also accounts for participant-specific and item-specific effects, eliminating the need to compute grand averages for each condition (Fromer et al., 2008). This approach offers significant advantages over traditional ANOVA-type analyses, which are still commonly used in recent EEG studies. Furthermore, the challenge of pictorially representing abstract and concrete concepts has limited the examination of concreteness effects in language production. Previous studies have attempted to address this limitation using word definitions within sentence contexts (Hanley et al., 2013). In Chapter 5, we tackled this challenge by using a picture-naming task with two sets of concrete images to explore the concreteness effect of constituents, partially circumventing the difficulty of representing abstract words/morphemes. Consequently, the research design in Chapter 5 offers insights for future investigations into abstract imagery in language production.

6.4 Limitations and directions for future research

This thesis has some limitations as well. First, we did not control for semantic transparency across the four experiments. Semantic transparent compounds may be processed differently than opaque compounds, where meanings are less obvious. Prior studies show mixed findings on the role of semantic transparency in morphological processing. For instance, while some research (MacGregor & Shtyrov, 2013; Tsang et al., 2022) suggested transparency effects, other researchers (Koester & Schiller, 2008; Zwitserlood et al., 2000, 2002) reported no significant difference between transparent and opaque compounds in morphological priming. Koester and Schiller (2008) observed reduced N400 effects for both types of morphologically related primes without significant distinctions, indicating a need for further investigation into the processing of opaque compounds in Mandarin.

Second, the material's free and bound morphemes may also have influenced the results. Bound morphemes, which cannot stand alone, might be processed differently from free morphemes, potentially affecting compound word encoding and retrieval. The differences between bound and free morphemes could influence the morphological encoding in Mandarin compound production. While a post-hoc analysis in Chapters 2 and 4 found no significant effects, this remains an area worth exploring.

Additionally, while we selected noun compounds as stimuli, we did not control for the word category (e.g., noun, verb) of their constituents in the current thesis due to stimulus selection challenges. Hanley et al. (2013) highlighted the influence of word category on lexical retrieval in their experiments. Therefore, future research could investigate this aspect more comprehensively to provide more apparent evidence regarding the role of word class in compound representation. Additional variables, such as participants' multilingual backgrounds and the age of acquisition (AoA) of compound words, may also have influenced our findings. Furthermore, Mandarin's unique logographic writing system, in which characters serve as basic units, might produce different results for individuals with varying literacy levels, especially regarding compound word representation. Future research could explore these aspects further. Although these are complex issues to address, even modest progress in these areas would advance our scientific understanding, particularly in Mandarin compound word production.

In conclusion, this thesis investigated the central question of Mandarin compound representation through four experiments, each targeting different aspects and effects. While some results showed variability, the overall findings were consistent, with all four experiments supporting the decompositional hypothesis of compound word representation in Mandarin.

References

Abdel Rahman, R., & Aristei, S. (2010). Now you see it… and now again: Semantic interference reflects lexical competition in speech production with and without articulation. *Psychonomic Bulletin & Review*, *17*(5), 657-661. https://doi:10.3758/PBR.17.5.657

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*(5), 713-734. https://doi.org/10.1080/01690960802597250

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*(5), 970-983. https://doi.org/10.1016/j.neuropsychologia.2011.01.003

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

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning.

Psychological Review, *118*(3), 438-481. https://doi.org/10.1037/a0023851

Badecker, W. (2001). Lexical composition and the production of compounds: Evidence from errors in naming. *Language and Cognitive Processes*, *16*(4), 337-366. https://doi.org/10.1080/01690960042000120

Barber, H., Domınguez, A., & de Vega, M. (2002). Human brain potentials indicate morphological decomposition in visual word recognition. *Neuroscience Letters*, *318*(3), 149-152. https://doi.org/10.1016/S0304-3940(01)02500-9

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Bates, D., & Maechler, M. (2020). Bolker [aut, B., cre, Walker, S., Christensen, RHB, Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & simulate. formula. PNK (shared copyright on, 1.1-26.

Belmore, S. M., Yates, J. M., Bellack, D. R., Jones, S. N., & Rosenquist, S. E. (1982). Drawing inferences from concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior*, *21*(3), 338-351. https://doi.org/10.1016/S0022-5371(82)90659-4

Bi, Y., Han, Z., & Shu, H. (2007). Compound frequency effect in word production: Evidence from anomia. *Brain and Language*, *103*(1-2), 55-56. https://doi.org/10.1016/j.bandl.2007.07.042

Bien, H., Levelt, W. J., & Baayen, R. H. (2005). Frequency effects in compound production. *Proceedings of the National Academy of*

Sciences, *102*(49), 17876-17881. https://doi.org/10.1073/pnas.0508431102

Bozic, M., & Marslen-Wilson, W. (2010). Neurocognitive contexts for morphological complexity: Dissociating inflection and derivation. *Language and Linguistics Compass*, *4*(11), 1063-1073. https://doi.org/10.1111/j.1749-818X.2010.00254.x

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34-44. https://doi.org/10.1162/jocn.1993.5.1.34

Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language*, *114*, 104125. https://doi.org/10.1016/j.jml.2020.104125

Butterworth, B. (1983). Lexical Representation. In B. Butterworth (Ed.), *Language production: Vol. II. Development, writing and other language processes* (pp. 257-294). Academic Press.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177-208. https://doi.org/10.1080/026432997381664

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297-332. https://doi.org/10.1016/0010-0277(88)90017-0

Catricalà, E., Della Rosa, P. A., Plebani, V., Vigliocco, G., & Cappa, S. F. (2014). Abstract and concrete categories? Evidences from neurodegenerative diseases. *Neuropsychologia*, *64*, 271-281. https://doi.org/10.1016/j.neuropsychologia.2014.09.041

Chen, T.-M., & Chen, J.-Y. (2006). Morphological encoding in the production of compound words in Mandarin Chinese. *Journal of Memory and Language*, *54*(4), 491-514. https://doi.org/10.1016/j.jml.2005.01.002

Chen, T. M., & Chen, J. Y. (2015). The phonological planning in M andarin spoken production of mono-and bimorphemic words. *Japanese Psychological Research*, *57*(1), 81-89. https://doi.org/10.1111/jpr.12059

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In *Mechanisms of language acquisition* (pp. 1-33). Psychology Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240-247. https://doi.org/10.1016/S0022-5371(69)80069-1

Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *12*(1), 125-131. https://doi.org/10.3758/BF03196357

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, *10*, 111-117. https://doi.org/ 10.3758/bf03196474

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283-321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1-3), 287-314. dictionhttps://doi.org/ 10.1016/0010-0277(92)90046-k

Del Maschio, N., Fedeli, D., Garofalo, G., & Buccino, G. (2021). Evidence for the concreteness of abstract language: a meta-analysis of neuroimaging studies. *Brain Sciences*, *12*(1), 32. https://doi.org/10.3390/brainsci12010032

Dohmes, P., Zwitserlood, P., & Bölte, J. (2004). The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language*, *90*(1-3), 203-212. https://doi.org/10.1016/S0093-934X(03)00433-4

Domínguez, A., De Vega, M., & Barber, H. (2004). Event-related brain potentials elicited by morphological, homographic, orthographic, and semantic priming. *Journal of Cognitive Neuroscience*, *16*(4), 598-608. https://doi.org/10.1162/089892904323057326

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages.

Quarterly Journal of Experimental Psychology, *71*(4), 808-816. https://doi.org/10.1080/17470218.2017.1310261

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. https://doi.org/10.1016/j.cognition.2017.09.014

Feldman, L. B. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1431. https://doi.org/10.1037/0278-7393.26.6.1431

Frauenfelder, U. H., & Schreuder, R. (1991). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In *Yearbook of Morphology 1991* (pp. 165-183). Springer.

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*(6), 790-796. https://doi.org/10.1016/S0010-9452(08)70419-2

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, *22*(7), 953-1000. https://doi.org/10.1080/01690960701190215

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*(2), 182-216. https://doi.org/10.1006/jmla.2000.2716

Fröber, K., Stürmer, B., Frömer, R., & Dreisbach, G. (2017). The role of affective evaluation in conflict adaptation: An LRP study. *Brain and Cognition*, *116*, 9-16. https://doi.org/10.1016/j.bandc.2017.05.003

Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Frontiers in Neuroscience*, *12*, 48. https://doi.org/10.3389/fnins.2018.00048

Fromkin, V. A. (1973). Slips of the tongue. *Scientific American*, *229*(6), 110-117. https://www.jstor.org/stable/24923270

Garrett, M. (1980). Levels of processing in sentence production. In *Language production* Vol. 1: Speech and talk (pp. 177-220). Academic Press.

Gerhand, S., & Barry, C. (2000). When does a deep dyslexic make a semantic error? The roles of age-of-acquisition, concreteness, and frequency. *Brain and Language*, *74*(1), 26-47. https://doi.org/10.1006/brln.2000.2320

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640-654. https://doi.org/10.1037//0096-1523.10.5.640

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*(1), 13. https://doi.org/10.1037//0096-3445.118.1.13

Hanley, J. R., Dell, G. S., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology*, *21*(2-4), 147-158. https://doi.org/10.1080/02643290342000339

Hanley, J. R., Hunt, R. P., Steed, D. A., & Jackman, S. (2013). Concreteness and word production. *Memory & Cognition*, *41*, 365-377. https://doi.org/10.3758/s13421-012-0266-5

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2005). Semantic competition between hierarchically related words during speech planning. *Memory & Cognition*, *33*(6), 984-1000. https://doi.org/ 10.3758/BF03193207

Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience*, *23*, 2432–2446. https://doi.org/10.1162/jocn.2011.21614

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721. https://doi.org/10.1037/0278-7393.25.3.721

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, *2*, 255. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144. https://doi.org/10.1016/j.cognition.2002.06.001

Jacobs, C. L., & Dell, G. S. (2014). 'hotdog', not 'hot''dog': the phonological planning of compound words. *Language, Cognition and Neuroscience*, *29*(4), 512-523. https://doi.org/10.1080/23273798.2014.892144

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. https://doi.org/10.1080/01690960802250900

Janssen, N., Pajtas, P. E., & Caramazza, A. (2014). Task influences on the production and comprehension of compound words. *Memory & Cognition*, *42*, 780-793. https://doi.org/10.3758/s13421-014-0396-z

Jescheniak, J. D., & Levelt, W. J. (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. https://doi.org/10.1037/0278-7393.20.4.824

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. https://doi.org/10.1037/0278-7393.24.5.1256

Kaczer, L., Timmer, K., Bavassi, L., & Schiller, N. O. (2015). Distinct morphological processing of recently learned compound words: An ERP study. *Brain Research*, *1629*, 309-317. https://doi.org/10.1016/j.brainres.2015.10.029

Koester, D., Gunter, T. C., & Wagner, S. (2007). The morphosyntactic decomposition and semantic composition of German compound words investigated by ERPs. *Brain and Language*, *102*(1), 64-79. https://doi.org/10.1016/j.bandl.2006.09.003

Koester, D., Holle, H., & Gunter, T. C. (2009). Electrophysiological evidence for incremental lexical-semantic integration in auditory compound comprehension. *Neuropsychologia*, *47*(8-9), 1854-1864. https://doi.org/10.1016/j.neuropsychologia.2009.02.027

Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, *16*(9), 1647-1668. https://doi.org/10.1162/0898929042568541

Koester, D., & Schiller, N. O. (2008). Morphological priming in overt language production: Electrophysiological evidence from Dutch. *Neuroimage*, *42*(4), 1622-1630. https://doi.org/10.1016/j.neuroimage.2008.06.043

Koester, D., & Schiller, N. O. (2011). The functional neuroanatomy of morphology in language production. *Neuroimage*, *55*(2), 732-741. https://doi.org/10.1016/j.neuroimage.2010.11.044

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 804. https://doi.org/10.1037//0278-7393.20.4.804

Krott, A., Baayen, R. H., & Hagoort, P. (2006). The nature of anterior negativities caused by misapplications of morphological rules. *Journal of Cognitive Neuroscience*, *18*(10), 1616-1630. https://doi.org/10.1162/jocn.2006.18.10.1616

Kuiper, K., Van Egmond, M. E., Kempen, G., & Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, *2*(3), 313-357. https://doi.org/10.1075/ml.2.3.03kui

Kuipers, J. R., & La Heij, W. (2008). Semantic facilitation in category and action naming: Testing the message-congruency account. *Journal of Memory and Language*, *58*(1), 123-139. https://doi.org/10.1016/j.jml.2007.05.005

Kutas, M., & Federmeier, K. D. (2009). N400. *Scholarpedia*, *4*(10), 7790. https://doi.org/10.4249/scholarpedia.7790

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Event-related potentials in cognitive science. In *Handbook of Cognitive Neuroscience* (pp. 387-409). Boston, MA: Springer US.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83-143). Academic Press. https://doi.org/10.1016/B978-012369374-7/50018-3

La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research*, *61*(3), 209-219. https://doi.org/10.1007/s004260050026

Language Teaching and Research Institute of Beijing Language Institute. (1986). *Dictionary of Modern Chinese Frequency*. Beijing Language Institute Press.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933. https://doi.org/10.1038/nrn2532

Laudanna, A., Badecker, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. *Journal of Memory and Language*, *31*(3), 333-348. https://doi.org/10.1016/0749-596X(92)90017-R

Lee, C.-L., & Federmeier, K. D. (2008). To watch, to see, and to differ: An event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and*

Language, *104*(2), 145-158. https://doi.org/10.1016/j.bandl.2007.06.002

Lensink, S. E., Verdonschot, R. G., & Schiller, N. O. (2014). Morphological priming during language switching: an ERP study. *Frontiers in Human Neuroscience*, *8*, 995. https://doi.org/10.3389/fnhum.2014.00995

Lenth R, Singmann H, Love J, Buerkner P, Herve M. (2019) Emmeans: estimated marginal means, aka least-squares means (Version 1.3. 4). Emmeans Estim Marg Means Aka Least-Sq Means HttpsCRAN R-Proj Orgpackage Emmeans.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. MIT press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38. https://doi.org/10.1017/s0140525x99001776

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*(1), 50-64. https://doi.org/10.1016/S0093-934X(02)00520-5

Lieber, R. (2021). *Introducing Morphology*. Cambridge University Press. https://doi.org/10.1017/9781108957960

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

Longtin, C.-M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and*

Language, *53*(1), 26-41. https://doi.org/10.1016/j.jml.2005.02.008

Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.

Luce. R-D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195070019.001.0001

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485-495. https://doi.org/10.3758/BF03198265

Lüttmann, H., Zwitserlood, P., Böhl, A., & Bölte, J. (2011). Evidence for morphological composition at the form level in speech production. *Journal of Cognitive Psychology*, *23*(7), 818-836. https://doi.org/10.1080/20445911.2011.575774

MacGregor, L. J., & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language*, *126*(2), 217-229. https://doi.org/10.1016/j.bandl.2013.04.002

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503. https://doi.org/10.1037/0278-7393.33.3.503

Marelli, M., Aggujaro, S., Molteni, F., & Luzzatti, C. (2012). The multiple-lemma representation of Italian compound nouns: A single case study of deep dyslexia. *Neuropsychologia*, *50*(5), 852-861. https://doi.org/10.1016/j.neuropsychologia.2012.01.021

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3. https://doi.org/10.1037/0033-295X.101.1.3

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315. https://doi.org/10.1016/j.jml.2017.01.001

McKinnon, R., Allen, M., & Osterhout, L. (2003). Morphological decomposition involving non-productive morphemes: ERP evidence. *Neuroreport*, *14*(6), 883-886. https://doi.org/10.1097/00001756-200305060-00022

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69-89. https://doi.org/10.1016/0749-596X(91)90011-8

Mondini, S., Luzzatti, C., Zonca, G., Pistarini, C., & Semenza, C. (2004). The mental representation of Verb–Noun compounds in Italian: Evidence from a multiple single-case study in aphasia. *Brain and Language*, *90*(1-3), 470-477. https://doi.org/10.1016/S0093-934X(03)00458-9

Morris, J., Frank, T., Grainger, J., & Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology*, *44*(4), 506-521. https://doi.org/10.1111/j.1469-8986.2007.00538.x

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 199-203. https://doi.org/10.1002/wics.199

Newton, P. K., & Barry, C. (1997). Concreteness effects in word production but not word comprehension in deep dyslexia. *Cognitive Neuropsychology*, *14*, 481–509. https://doi.org/10.1080/026432997381457

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357. https://doi.org/10.1037/0033-295X.115.2.357

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255. https://doi.org/10.1037/h0084295

Penke, M., Weyerts, H., Gross, M., Zander, E., Münte, T. F., & Clahsen, H. (1997). How the brain processes complex words: an event-related potential study of German verb inflections. *Cognitive Brain Research*, *6*(1), 37-52. https://doi.org/10.1016/S0926-6410(97)00012-8

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. https://doi.org/10.1037/0278-7393.24.3.539

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456-463. https://doi.org/10.1016/S1364-6613(02)01990-3

Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language*, *52*(1), 25-82. https://doi.org/10.1006/brln.1996.0004

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

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

Rodriguez-Fornells, A., Clahsen, H., Lleó, C., Zaake, W., & Münte, T. F. (2001). Event-related brain responses to morphological violations in Catalan. *Cognitive Brain Research*, *11*(1), 47-58. https://doi.org/10.1016/S0926-6410(00)00063-X

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142. https://doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (1996). Morpheme frequency in speech production: Testing WEAVER. In *Yearbook of Morphology 1996* (pp. 135-154). Kluwer, Dordrecht.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and*

Language, *35*(6), 854-876. https://doi.org/10.1006/jmla.1996.0044

Roelofs, A., & Baayen, H. (2002). Morphology by itself in planning the production of spoken words. *Psychonomic Bulletin & Review*, *9*, 132-138. https://doi.org/10.3758/bf03196269

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 353. https://doi.org/ 10.1037/0096-3445.137.2.303

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 922. https://doi.org/10.1037/0278-7393.24.4.922

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 246–251. https://doi.org/10.1037/0278-7393.22.1.246

Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development*, *48*(2), 643–647. https://doi.org/10.2307/1128667.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, *18*(4), 367-379. https://doi.org/10.3758/BF03197126

Sailor, K., & Brooks, P. J. (2014). Do part–whole relations produce facilitation in the picture–word interference task?. *Quarterly Journal of Experimental Psychology*, *67*(9), 1768-1785. https://doi.org/ 10.1080/17470218.2013.870589

Schiller, N. O. (2020). Neurolinguistic approaches in morphology. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.601

Schiller, N. O., & Verdonschot, R. G. (2019). Morphological Theory and Neurolinguistics. In J. A. F. Masini (Ed.), *The Oxford Handbook of Morphological Theory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199668984.013.31

Schwanenflugel, P. J. (1991). Contextual constraint and lexical processing. In *Advances in Psychology* (Vol. 77, pp. 23-45). Elsevier. https://doi.org/10.1016/S0166-4115(08)61528-9

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499-520. https://doi.org/10.1016/0749-596X(88)90022-8

Smolka, E., Gondan, M., & Rösler, F. (2015). Take a stand on understanding: Electrophysiological evidence for stem access in German complex verbs. *Frontiers in Human Neuroscience*, *9*, 62. https://doi.org/10.3389/fnhum.2015.00062

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. https://doi.org/10.1037/0278-7393.6.2.174

Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S., & Ković, V. (2022). How to do better N400 studies: reproducibility, consistency and adherence to research standards in the existing literature. *Neuropsychology Review*, *32*(3), 577-600. https://doi.org/10.1007/s11065-021-09513-4

Spinnato, J., Roubaud, M., Burle, B., & Torrésani, B. (2015). Detecting single-trial EEG evoked potential using a wavelet domain linear mixed model: application to error potentials classification. *Journal of Neural Engineering*, *12*(3), 036013. https://doi.org/10.1088/1741-2560/12/3/036013

Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161-184. https://doi.org/10.1016/j.jml.2005.11.001

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. https://doi.org/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. https://doi.org/10.1037/0278-7393.22.4.896

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*(4), 912-928. https://doi.org/10.1093/cercor/bhp153

Su, I.-F., Yum, Y. N., & Lau, D. K.-Y. (2023). Hong Kong Chinese character psycholinguistic norms: Ratings of 4376 single Chinese characters on semantic radical transparency, age-of-acquisition, familiarity, imageability, and concreteness. *Behavior Research Methods*, *55*(6), 2989-3008. https://doi.org/10.3758/s13428-022-01928-y

Tan, L.-H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. *Reading and Writing*, *10*, 165-200. https://doi.org/10.1023/A:1008086231343

Tan, L. H., Liu, H.-L., Perfetti, C. A., Spinks, J. A., Fox, P. T., & Gao, J.-H. (2001). The neural system underlying Chinese logograph reading. *Neuroimage*, *13*(5), 836-846. https://doi.org/10.1006/nimg.2001.0749

Tsang, Y.-K., Zou, Y., & Tse, C.-Y. (2022). Semantic transparency in Chinese compound word processing: Evidence from mismatch negativity. *Neuroscience*, *490*, 216-223. https://doi.org/10.1016/j.neuroscience.2022.03.007

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition*, *18*, 380-393. https://doi.org/10.3758/BF03197127

Verdonschot, R. G., Middelburg, R., Lensink, S. E., & Schiller, N. O. (2012). Morphological priming survives a language switch. *Cognition*, *124*(3), 343-349. https://doi.org/10.3758/BF03197127

Voeten, C. (2019). permutes: Permutation tests for time series data.

Von Grebmer zu Wolfsthurn, S., Robles, L. P., & Schiller, N. O. (2021a). Cross-linguistic interference in late language learners: An ERP study. *Brain and Language*, *221*, 104993. https://doi.org/10.1016/j.bandl.2021.104993

Von Grebmer zu Wolfsthurn, S., Robles, L. P., & Schiller, N. O. (2021b). Noun-phrase production as a window to language selection: An ERP study. *Neuropsychologia*, *162*, 108055. https://doi.org/10.1016/j.neuropsychologia.2021.108055

Wang, J., Schiller, N. O., & Verdonschot, R. G. (2024). Morphological encoding in language production: Electrophysiological evidence from Mandarin Chinese compound words. *PLoS ONE*, *19*(10), e0310816. https://doi.org/10.1371/journal.pone.0310816

Wang, M., Chen, Y., Jiang, M., & Schiller, N. O. (2021). The time course of speech production revisited: no early orthographic effect, even in Mandarin Chinese. *Language, Cognition and Neuroscience*, *36*(1), 13-24. https://doi.org/10.1080/23273798.2020.1803373

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *Journal of Cognitive*

Neuroscience, *12*(6), 1024-1037. https://doi.org/10.1162/08989290051137558

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge. https://doi.org/10.4324/9781315165547

Winther Balling, L., & Harald Baayen, R. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, *23*(7-8), 1159-1190. https://doi.org/10.1080/01690960802201010

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*, 2031-2046. https://doi.org/10.1016/S1364-6613(00)01545-X

Xu, X., & Li, J. (2020). Concreteness/abstractness ratings for two-character Chinese words in MELD-SCH. *PLoS ONE*, *15*(6), e0232133. https://doi.org/10.1371/journal.pone.0232133

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, *10*(6), 545-600. https://doi.org/10.1080/01690969508407114

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

Zhou, X., & Marslen-Wilson, W. (2014). Words, morphemes and syllables in the Chinese mental lexicon. In *Morphological Structure, Lexical Representation and Lexical Access (RLE Linguistics C: Applied Linguistics) (pp. 393-422)*. Routledge.

Zwitserlood, P., Bölte, J., & Dohmes, P. (2000). Morphological effects on speech production: Evidence from picture naming. *Language and Cognitive Processes*, *15*(4-5), 563-591. https://doi.org/10.1080/01690960050119706

Zwitserlood, P., Bölte, J., & Dohmes, P. (2002). Where and how morphologically complex words interplay with naming pictures. *Brain and Language*, *81*(1-3), 358-367. https://doi.org/10.1006/brln.2001.2530

Zwitserlood, P. (2014). The role of semantic transparency in the processing and representation of Dutch compounds. In *Morphological Structure, Lexical Representation and Lexical Access (RLE Linguistics C: Applied Linguistics)* (pp. 341-368). Routledge. https://doi.org/10.4324/9781315857213

Summary

This dissertation focuses on the central research question of how Mandarin compound words are represented during language production. Specifically, it examines whether compound words are stored in the mental lexicon in a decomposed or holistic manner. If decomposition occurs during production, at what level does it take place? Are the storage mechanisms at the lemma and lexeme levels-two stages in lexical selection-distinct? According to the theoretical model proposed by Levelt et al. (1999), decomposition of compound words can occur at either the lemma or lexeme levels during lexical selection. To explore the representation of Mandarin compound words during lexical selection, this dissertation investigates two primary hypotheses: the morphological decomposition hypothesis and the full-listing hypothesis. The former posits that compound words are represented in the mental lexicon through their constituent morphemes, which play a role in lexical retrieval. In contrast, the latter asserts that compound words are stored as whole units, negating the involvement of morphemes in lexical retrieval. The aim of this research is to uncover how Mandarin Chinese compound words are represented at the two levels of lexical selection during speech production. This dissertation comprises six chapters:

In **Chapter 1**, we introduce the research background and theoretical framework. Grounded in Levelt et al.'s (1999) language production theory. This chapter provides an overview of Levelt's

language production model, summarizes relevant literature, and outlines the key research questions, experimental methods, and main findings presented in subsequent chapters.

Chapter 2 describes the first experiment we conducted, which employed a long-lag priming paradigm to investigate the morphological encoding of Mandarin Chinese disyllabic compound words. Thirty-six native Mandarin speakers participated in a picture-naming task, and both behavioral and electrophysiological data were collected. Unexpectedly, the behavioral data revealed no facilitation effect in response times (RTs) when morphologically related primes preceded target words in the picture-naming task. RTs for morphologically related and unrelated primes showed no significant differences at either the first- or second-character position, indicating no position effects. However, event-related potentials (ERPs) revealed a reduction in N400 amplitude for morphologically related primes compared to unrelated primes, suggesting a priming effect at the morphological encoding level and supporting the decomposition model. Thus, although the behavioral results showed no morphological priming effect, the ERP results revealed an N400 effect, supporting the involvement of morphological priming in the production of Mandarin compound words. The discrepancy between behavioral and electrophysiological data may stem from the use of the long-lag priming paradigm. While this paradigm is designed to isolate morphological effects by eliminating phonological and semantic influences, the interval between prime and target words might reduce behavioral sensitivity to subtle differences across conditions. In contrast, EEG is sufficiently sensitive to detect fine-grained differences in brain responses. Additionally, although position effects were absent, naming times for first-position targets were shorter than for second-position targets, possibly aligning with the serial assumption and modifier-header hypothesis.

Chapter 3 continues the investigation of compound word representation in the mental lexicon, focusing on morpheme frequency effects and compound word frequency effects. This experiment aimed to further test whether compound words are stored in a decomposed form (as proposed by Levelt et al., 1999) or as holistic units (as proposed by Janssen et al., 2008). Using a newly constructed stimuli list to replicate Janssen et al.'s design, while incorporating EEG data, the study examined how morpheme and compound word frequencies affect the production of Mandarin compound words. The results showed that behavioral data from the picture-naming task did not replicate Janssen et al.'s findings. Reaction times for low-frequency compounds with high-frequency morphemes (L(h)) were comparable to those for high-frequency compounds with high-frequency morphemes (H(h)), indicating a lack of compound frequency effects. However, L(h) responses were faster than those for low-frequency compounds with low-frequency morphemes (L(l)), supporting morpheme frequency effects and the decomposition model, which posits that morphemes play a critical role in word production. Differences between this study and Janssen et al.'s results may stem from the different corpora used: Janssen utilized the Modern Chinese

Frequency Dictionary (1986), which is skewed toward written language, whereas this study used the SUBTLEX-CH corpus (Cai & Brysbaert, 2010), which reflects spoken language. These corpus differences may have influenced frequency distributions, word list selection, and experimental outcomes. Interestingly, EEG data did not align with the behavioral results. Although reaction times supported morpheme frequency effects, no frequency effects were observed in ERPs. This inconsistency underscores the complexity of Mandarin morphological processing, necessitating further research to resolve these disparities. Further investigation could consider frequency effects, language factors (e.g., semantic transparency), and participant-related variables to deepen our understanding of morphological representation in Mandarin and cross-linguistic differences in morphological processing.

Chapter 4 explores the lemma-level representation of Mandarin compound words in the mental lexicon. Using a picture-word interference (PWI) paradigm, native Mandarin speakers were asked to name pictures while ignoring distractor words. Behavioral data were analyzed across three conditions: synonym distractors, semantically related distractors, and control distractors. The results revealed significantly faster naming times for synonym distractors compared to controls, consistent with previous findings of synonym facilitation effects (Dylman & Barry, 2018). In contrast, semantically related distractors delayed naming times compared to controls, aligning with previously documented semantic interference effects (Levelt et al., 1999; Lupker, 1979; Roelofs, 1992; Starreveld & La Heij, 1995, 1996).

These findings suggest that semantically related distractors hinder lexical retrieval, while synonym distractors facilitate it. The results support the hybrid lemma account and challenge lexical competition theories, which typically explain PWI tasks by positing competition among semantically related word representations during lexical selection (Levelt et al., 1999). According to these theories, distractors activated by semantically related words compete with the target, delaying its selection. However, the observed facilitation effect from synonym distractors challenges this competition-based explanation. Overall, the findings suggest a hybrid lemma storage model and provide behavioral evidence for the decomposition model, while questioning the lexical competition hypothesis.

Chapter 5 examines the representation of Mandarin compound words in the mental lexicon, with a focus on the role of morpheme concreteness in compound word production. The experiment included two conditions: concrete condition (cc), in which both morphemes were concrete, and abstract condition (aa), in which both morphemes were abstract. Behavioral results showed significantly faster naming times for the concrete condition, providing evidence that morpheme concreteness affects lexical retrieval in compound word production. This finding has implications for both the full-listing and decompositional models. The observed concreteness effect supports the decompositional model, indicating that each morpheme contributes to lexical retrieval and influences whole-word processing. The findings also align with studies suggesting that abstract words have weaker semantic-lexical connections (Hanley et al., 2004; 2013). Additionally, abstract words tend to have more meanings and higher ambiguity (Hoffman et al., 2011), which may explain their slower retrieval. Besides, these results further support the hybrid lemma account, suggesting that compound words are stored as single units at the lemma level and decomposed into morphemes at the lexeme level for phonological encoding.

In **Chapter 6**, we synthesize the findings from the above four experiments, summarizing their conclusions and theoretical implications. It discusses the limitations encountered and suggests directions for future research. Overall, the results support the decompositional model and the hybrid lemma account of Mandarin compound word representation, indicating that Mandarin compound words are stored in a decomposed form during production, with decomposition occurring at both the lemma and lexeme levels. The dissertation also examines potential cross-linguistic differences in compound word representation, highlighting how Mandarin's unique writing system and phonological structure might contribute to these differences. This research advances our understanding of the processing and representation of Mandarin compound words in the mental lexicon.

Samenvatting

Deze dissertatie richt zich op de centrale onderzoeksvraag hoe samengestelde woorden in het Mandarijn worden gerepresenteerd tijdens taalproductie. Meer specifiek onderzoekt het of samengestelde woorden in het mentale lexicon op een gedecomposeerde of holistische manier worden opgeslagen. Indien decompositie plaatsvindt tijdens de productie, op welk niveau vindt dit dan plaats? Zijn de opslagmechanismen op het lemma - en lexemeniveau - twee fasen in lexicale selectie-verschillend? Volgens het theoretische model van Levelt et al. (1999) kan decompositie van samengestelde woorden plaatsvinden op het lemma- of lexemeniveau tijdens lexicale selectie.

Om de representatie van Mandarijnse samengestelde woorden tijdens lexicale selectie te onderzoeken, richt deze dissertatie zich op twee primaire hypothesen: de hypothese van morfologische decompositie en de hypothese van volledige opslag. De eerste stelt dat samengestelde woorden in het mentale lexicon worden gerepresenteerd door hun samenstellende morfemen, die een rol spelen bij lexicale toegang. De tweede daarentegen beweert dat samengestelde woorden als geheel worden opgeslagen, waarbij morfemen geen rol spelen in lexicale toegang.

Het doel van dit onderzoek is te achterhalen hoe Mandarijnse samengestelde woorden worden gerepresenteerd op de twee niveaus van lexicale selectie tijdens spraakproductie. Deze dissertatie bestaat uit zes hoofdstukken:

Hoofdstuk 1: Dit hoofdstuk introduceert de onderzoeksachtergrond en het theoretische kader, gebaseerd op de taalproductietheorie van Levelt et al. (1999). Het geeft een overzicht van het taalproductiemodel van Levelt, vat relevante literatuur samen, en schetst de kernvragen, experimentele methoden en belangrijkste bevindingen die in de volgende hoofdstukken worden gepresenteerd.

Hoofdstuk 2: Hier wordt het eerste experiment beschreven, waarin een long-lag primingparadigma werd gebruikt om de morfologische codering van Mandarijnse disyllabische samengestelde woorden te onderzoeken. Zesendertig moedertaalsprekers van het Mandarijn namen deel aan een benoemingstaak met afbeeldingen, waarbij zowel gedrags- als elektrofysiologische gegevens werden verzameld. Gedragsdata toonden onverwacht geen facilitatie-effect in reactietijden (RT's) wanneer morfologisch gerelateerde primes voorafgingen aan doelwoorden. De RT's tussen morfologisch gerelateerde en niet-gerelateerde primes verschilden niet significant, zonder positie-effecten. ERPs lieten echter een reductie in N400-amplitude zien bij morfologisch gerelateerde primes, wat een primingeffect op het niveau van morfologische codering suggereert en de decompositiemodel ondersteunt. De discrepantie tussen gedrags- en elektrofysiologische data kan te wijten zijn aan de lange interval tussen prime en doelwoord, wat gedragsgevoeligheid kan verminderen, terwijl EEG subtiele verschillen kan detecteren.

Hoofdstuk 3: Dit hoofdstuk gaat verder met de studie van de representatie van samengestelde woorden in het mentale lexicon, met de nadruk op morfeemfrequentie- en woordfrequentie-effecten. De resultaten lieten zien dat de reactietijden geen effecten van samenstellingsfrequentie toonden, maar dat lage-frequentie samengestelde woorden met hoge-frequentie morfemen (L(h)) sneller werden benoemd dan lage-frequentie samengestelde woorden met lage-frequentie morfemen (L(l)). Deze bevindingen ondersteunen de decompositiemodel, wat suggereert dat morfemen een kritieke rol spelen in woordproductie. ERP-data waren echter niet consistent met de gedragsresultaten, wat de complexiteit van Mandarijnse morfologische verwerking benadrukt.

Hoofdstuk 4: Hier wordt onderzocht hoe Mandarijnse samengestelde woorden op lemma-niveau worden gerepresenteerd in het mentale lexicon, met behulp van een picture-word interference (PWI)-paradigma. Synoniemen veroorzaakten snellere benoemingstijden, terwijl semantisch gerelateerde afleiders vertragingen veroorzaakten. Deze resultaten ondersteunen een hybride lemma-opslagmodel en betwisten hypothesen gebaseerd op lexicale competitie.

Hoofdstuk 5: Dit hoofdstuk onderzoekt de rol van morfeemconcreetheid bij de productie van samengestelde woorden. Concreetheid bleek significant snellere reactietijden te veroorzaken, wat de decompositiemodel ondersteunt. De resultaten suggereren dat Mandarijnse samengestelde woorden als enkele eenheden op lemma-niveau worden opgeslagen en op lexemeniveau worden gedecomposeerd.

Hoofdstuk 6: Hier worden de bevindingen uit de vorige experimenten samengevat, evenals hun theoretische implicaties. Het bespreekt beperkingen en suggereert richtingen voor toekomstig onderzoek. Over het algemeen ondersteunen de resultaten de decompositiemodel en het hybride lemma-account van de representatie van Mandarijnse samengestelde woorden, met decompositie op zowel lemma- als lexemeniveau. Deze dissertatie draagt bij aan ons begrip van de verwerking en representatie van Mandarijnse samengestelde woorden in het mentale lexicon.

摘要

本论文的主要研究问题是普通话产出过程中汉语复合词的表征方式，即汉语复合词在我们心理词典中是以分解的方式储存，还是整合的方式储存？如果普通话在产出过程中包括分解储存方式，那么这种分解发生在什么层级？在词汇选择过程中的两个层次，即词条层和词素层的储存方式是否相同？根据 Levelt 等人的理论模型（1999），复合词的分解储存方式可以发生在词汇选择的任何一个层级，即词条层级或者词素层级。为深入探讨在词汇选择过程中汉语复合词的表征方式，本论文重点关注两种主要假设：形态分解假设和整体列存假设。前者认为复合词在心理词典中是通过其构成语素来表征的，词素在词语提取中发挥着作用；而后者则主张复合词在心理词典中以整体形式存储，并否定词素在该词语提取中可能发挥的作用。本研究旨在探讨在言语产出过程中普通话复合词在词汇选择中的表征方式。

本论文共分为六章。**第一章**介绍了研究背景及理论基础。本论文基于 Levelt 等人的语言产出理论，将此理论从单词产出延展并应用到复合词产出过程。本章首先介绍了该语言产出理论并总结了相关文献，同时概况总结了各章节的主要研究问题、实验方法和简要结论。

第二章介绍了本论文的第一个实验。该实验使用长时滞启动范式检验普通话双音节复合词的形态编码过程，具体考察了普

通话复合词是否会产生形态启动效应这一研究问题。36 名汉语母语者完成了命名图片任务，他们的行为及电生理数据被记录下来。出乎意料的是，行为数据显示，在图片命名任务中，当形态相关启动词出现在目标词之前，反应时并未缩短，即未出现促进效应。此外，形态相关和不相关启动词之间的反应时在复合词的第一个字位置上或者第二个字位置上均无显著差异，因此并未发现位置效应。但是，事件相关电位分析显示，形态相关启动词相比于不相关启动词诱发的 N400 幅度减少，表明形态相关启动词在形态编码层面上发生了启动效应，支持了词汇分解模型。综上所述，此实验的行为结果未出现形态学启动效应，但电生理学结果发现 N400效应，支持了汉语普通话复合词在产出过程中词素成分具有形态学启动效应这一观点。本文认为两种数据类型的差异可能在于长时滞启动范式的使用。虽然根据之前的研究此范式可用来排除语音和语义的启动效应从而捕捉形态促进效应，但由于该范式的目标词和启动词的间隔试次过长，可能导致行为数据不够灵敏因此无法捕捉不同条件之间的细微差别。但是脑电技术足够灵敏，且擅长捕捉大脑在处理不同数据时释放出来的细微差异，因此脑电数据发现了效应。此外，位置效应虽然缺失，但是第一位置的启动词命名时间比第二位置的启动词命名时间要短，这可能与顺序处理假说和中心词假说有关。

第三章继续探讨了普通话复合词在心理词典中的表征方式，重点关注了词素词频效应和复合词频率效应。此实验旨在继续探

究复合词是以分解形式存储还是整词形式存储。本实验设计了新的词表来复制 Janssen 等人（2008）的实验并加入了 EEG 数据，以考察词素词频和复合词词频对普通话复合词产出过程中的影响。实验结果表明，图片命名任务的行为数据结果未能复制Janssen等人的发现，即低词频/高词素频率（L(h)）条件的反应时与高词频/高词素频率（H(h)）条件相当，表明复合词词频效应的缺失；L(h)比低词频/低词素频率（L(l)）命名速度快，这支持了词素频率效应，表明词素在词语生产中发挥了关键作用，进而支持了分解模型。本实验结果与Janssen等人的实验结果存在的差异可能源于不同语料库的使用，即Janssen使用了《现代汉语频率词典》（1986），该语料库更倾向于书写语料；本研究则基于电影字幕语料库SUBTLEX-CH，该语料库更倾向于口语语料。语料库类型的差异可能影响了频率分布，进行影响了词表选择和实验结果。有趣的是，本实验的 ERP 数据与行为结果不一致。尽管反应时显示了词素频率效应，但 ERP 数据未能支持该词素频率效应。在本实验中，行为和 ERP 结果的不一致凸显了汉语形态加工的复杂性，需要进一步的研究以澄清这些差异。综上所述，本实验行为结果表明普通话复合词产出过程受到词素频率的影响，这支持了分解模型；然而，ERP 数据未显示词素频率效应。未来的研究可以更加深入地研究这些频率效应，同时考虑其他语言因素，如语义透明度和被试相关变量，以加深对普通话复合词形态表征以及跨语言形态加工差异的理解。

第四章探讨了普通话复合词在心理词典中词条层面的表征方式。本实验使用图词干扰范式，要求普通话母语者在忽略干扰词的同时对图片进行命名。本实验收集了被试图片命名的反应时作为数据进行分析。实验设计涵盖了三种条件：同义干扰词条件、语义相关干扰词条件和控制条件。本章的研究结果表明，同义干扰词的图片命名显著快于控制组，发现了与先前研究一致的同义词促进效应。相比之下，语义相关干扰词比控制组的命名时间更长，这与之前研究中记录的抑制效应一致。本章结果表明语义相关干扰词阻碍了词语提取过程，而同义干扰词则促进了词语提取过程，缩短了词语加工时长。本研究支持了混合词条储存模型。图词干扰任务通常通过词汇竞争理论来解释，即图片命名过程中被激活的语义相关词语之间的竞争会阻碍目标词语选择。在该模型中，当干扰词激活了非目标词时，它会成为竞争者，延迟目标名称的选择。Finkbeiner 等人（2006）指出，密切相关的词汇表征会使目标词选择变得复杂，他们认为这种机制也适用于同义词。然而本研究发现同义干扰词对目标词的命名产生了促进作用，这对竞争理论提出了挑战。综上所述，本实验结果为混合词条储存模型提高了数据支持，进而为分解模型提供了行为学依据。

在**第五章**中，我们探讨了汉语复合词在心理词典中的表征方式，特别关注了词素具体性对复合词产出的影响。本实验设有两种条件：具体词素条件指包含两个具体词素的复合词，而抽象词素条件指包含两个抽象词素的复合词。行为实验结果显示，具

体词素条件的反应时显著快于抽象词素条件，为复合词产出过程中词素的具体性对词语提取产生影响提供这一论点提供了数据支持。此发现对两种词汇表征模型的验证具有重要意义：整词存储模型和分解模型。整词存储模型假设复合词作为整体单元存储在心理词典中，且不受词素的影响。相反，分解模型认为复合词通过其组成部分被提取，预测会有词素效应。本实验观察到的词素具体性效应支持分解模型，表明在词汇生成过程中复合词的词素会被分别提取并影响整词的处理。此前研究对于具体性效应的解释之一是，抽象词的语义-词汇联结较弱，这一观点可能超越了单词层面，扩展到词素（复合词组成部分）层面，这也符合 Newton 和 Barry（1997）关于抽象词更难提取的主张。他们认为抽象词更容易受到语义邻近词的竞争影响。此外，Hoffman 等人（2011）通过潜在语义分析显示，抽象词具有更多的义项并且更具歧义性。总之，本实验研究结果突显了具体性在普通话语言产出中的重要性，未来研究可以从更广的角度探索具体性对复合词在语言产出过程中的影响。此外，本章的实验结果与词汇表征的混合词条理论一致。本实验观察到的具体性效应表明词素的词条在整个复合词生成中发挥了作用，进一步支持了混合词位观点而非单一词位观点。

本论文**最后一章**对上述的四组实验结果进行了归纳总结，并阐述了每个实验的研究结论和理论启发。本论文还探讨了这些结果对理论框架的启示以及局限性。未来的研究可以在本论文成

果的基础上开展进一步研究。综合以上四个实验的研究结果，本论文结果支持普通话复合词分解模型表征观点，即普通话复合词在产出加工过程中采取分解储存的方式，且该分解过程可能发生在词汇选择的两个层级，即词条层级和词素层级。此外，本文还探讨了汉语与印欧语言在复合词表征上产生差异的潜在原因。在汉语独特的书写系统中，字符能够直接提供语义线索，加之其音韵结构中字符与语素的映射关系，而非与音素的对应关系，可能导致了跨语言差异。总体而言，本论文结果深化了对普通话复合词在心理词典中加工与表征方式的理解，尤其是对形态加工处理机制的理解。

Curriculum vitae

Jiaqi Wang was born on the 14th of March 1997 in Tianjin, China. She was born and raised in Hebei Province, where she completed her primary, secondary, and high school education. From 2012 to 2016, she attended Hebei University of Technology, earning a bachelor's degree in English. In 2016, she was admitted without examination to the research master's program in Foreign Language and Literature at the China University of Petroleum (Beijing). After completing her master's degree, she worked as a project manager at the China National Building Materials Exhibition & Trade Center in Beijing from July 2018 to December 2020. In January 2021, she began her PhD studies at Leiden University Centre for Linguistics (The Netherlands), with funding from the China Scholarship Council for 48 months. From February 2023 to February 2024, she was a visiting PhD student in the Neurobiology of Language Department at the Max Planck Institute for Psycholinguistics in Nijmegen (The Netherlands). She collected data as a visiting PhD student at Beijing Normal University from August 2023 to October 2023. Then, she obtained a visiting PhD student scholarship at the Language and Translation Department in the City University of Hong Kong from February 2025 to June 2025. This thesis presents the key findings of her doctoral research investigating the morphological encoding of Mandarin compounds in language production. Upon graduation, she will join the English Department of China University of Geosciences (Beijing) as a faculty member.