THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

PHONOLOGICAL SEGMENTATION NEIGHBORHOODS

KARL DAVID NEERGAARD

Ph.D

The Hong Kong Polytechnic University

2018

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Phonological Segmentation Neighborhoods

Karl David Neergaard

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

March 2017

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Karl David Neergaard

Abstract

Phonological networks constructed from a metric of sound similarity between lexical items, commonly referred to as phonological neighborhood density (PND), have revealed network characteristics of the mental lexicon through the combined use of graph theory and psycholinguistic tasks. PND has long been a metric to account for co-activation of lexical items in the mental lexicon. The purpose of this dissertation was to investigate the role that co-activation plays in speech processing in a tonal language under the lens of a complex system built upon phonological similarity. The study begins by constructing a Mandarin syllable inventory from participant-based phonological associations collected in a neighbor generation task. This inventory was then used to build fourteen phonological networks (8 with tone and 8 without tone) based on existing proposals of Mandarin syllable segmentation. A model selection procedure was used with mixed effect models to identify the optimal fit between the participants' reaction times and the lexical statistics representing each network schema in the remaining 5 tasks. The goal of each analysis was to identify the segmentation schema and neighborhood statistic/s that best accounted for each task. The findings reveal that the spreading of activation through similarity neighborhoods during auditory lexical processing is adaptive to the demands of the task at hand, both in the segmentation schema and the directional effect. In explicit mental search (neighbor generation) lexical items were activated through a network that was nontonal and unsegmented (CGVX), where greater density aided speech production. Through the use of the model selection procedure, a false positive was identified in an auditory shadowing task. In a second task, with stimuli of greater regularity between segment and syllable length, greater density sped reaction times according to the tonal complex rime segmented schema (C_G_VX_T). Finally, an auditory lexical decision task was implemented that featured two classes of Mandarin nonwords: tone gap nonwords, which consist of existing syllables in the syllable inventory combined with one of the four lexical tones that together do not point to an existing lexical item; and syllable gap nonwords, which consist of existing lexical tones assigned to syllables that do not exist in the inventory but whose segments do. Activation spread through a tonal unsegmented schema for monosyllabic words (CGVX_T), while similar to words in the neighbor generation and auditory word repetition tasks, greater density facilitated reaction times. The tone gap nonwords identified a tonal onset/rime schema (C_GVX_T), while the syllable gap nonwords revealed another false positive. Contrary to real words, greater density slowed reaction time for the tone gap nonwords. An account is proposed that places the PND effect at a post-phonological stage, wherein segmentation is indicative of the integration of orthographic and phonological representations (CGVX and CGVX_T), or the lack of an orthographic influence (C_G_VX_T).

Acknowledgements

I would first like to thank Chu-Ren Huang and the members of his research group: Anna, Enrico, Hongchao, Hongzhi, Ihsuan, Jiajuan, Jing Ding, Menghan, Qingqing, Renhui, Ren-feng, Sichang, and Stella. Our many discussions at research meetings and dinner outings stretched through the computational, theoretical, and experimental approaches to addressing questions to both general linguistics and more specifically their implementation to the Chinese language family. I'd like to extend my thanks to the Scan-Unit at the University of Vienna, who adopted me at the end of my doctoral studies. Claus Lamm and his team: Anna, Federica, Isabella, Jasminka, Jessie, Lukas, Markus, Yili; Christoff Eisenegger and his team: Annabel, Hana, Lisa, Nace, Shawn, and many people in between like Sebastian, Giorgia, Igor, and Sigrid, have together been wonderful people to pass the cold Austrian winter with through stirring discussions on empathy and testosterone. Finally, I'm enormously grateful for my time at the Laboratoire Parole et Langage, Aix/Marseille University where James German and his research team: Tim, Oriana, Seung Kyung, LeAnn, Remi, and the AixProsody discussion group: Mariapaola, Pauline, Piera, Axel, and Julia, have made finishing revisions and starting new ventures truly enjoyable. Thank you also to Cigdem for being a friend and collaborator throughout my doctoral studies and hopefully far into the future.

A special heart filled thanks to my wife Utari, who has done an excellent job of dealing

with the effects of doctoral studies on a fragile mind.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

For speech to take place, a multivariate and multidimensional handling of information must be orchestrated at a millisecond scale. Perhaps most fascinating about this process is the fact that both speakers and listeners must identify a single item from a mental lexicon that is, for an average 20-year-old, made up of roughly 42,000 lemmas (uninflected words) plus tens of thousands more inflected forms and proper nouns (Brysbaert, Stevens, Mandera, & Keuleers, 2016). Traditionally, the mental lexicon has been approached according to the divisions common to the study of linguistics, wherein phonology, and morphology refer to the form of lexical items, while syntax, and semantics their meaning (Levelt, 1989). The field has come to recognize, however, that other aspects influence the access of lexical items from the mental lexicon, such as a language's orthography, the number of languages spoken by an individual, and a speaker's working memory capacity, to name a few. As the understanding of the mental lexicon grows so does the need to incorporate it within an explanatory structure that is capable, both empirically and theoretically, to encompass its complexity.

Both language (e.g., Beckner et al., 2009; Gell-Mann, 1994; Kirby, 1999) and the brain (e.g., Morowitz & Singer, 1995) have been called complex adaptive systems (CAS) in that they are both multi-agent systems of adaptive change with competitive forces that take on emergent properties over time. The mental lexicon, taken as a subsystem of an individual's cognition and subsequently the product of an agent within a language community, should reveal aspects of this adaptive complexity. In this chapter, I begin with what is known about lexical access from the psycholinguistic literature, specifically in relation to phonological similarity and the use of phonological networks. I then discuss CAS so as to the stage for an interpretation of the experimental findings. Finally, I review the unique linguistic characteristics of Mandarin Chinese, the test case for this dissertation.

1.1 LEXICAL ACCESS

Lexical information has long been theorized (Caramazza, 1997; Dell, 1986; Levelt, 1989) to traverse three levels of processing. In speech production, the order would be first the retrieval of a concept, followed by its best fitting lexical item, and then retrieval of that lexical item's phonological code. The reverse process is said to occur for the perception of an auditory item. Whether this is theorized to occur in a feedforward manner (e.g., Levelt, Roelofs, & Meyer, 1999) or from a neural network that allows for interactive activation (e.g., Seidenberg & McClelland, 1989), these three levels persist in qualitative and quantitative accounts of lexical access.

The case for interactivity during speech was early on supported through speech errors. While the numerous types of speech errors might occur through the addition, deletion or substitution of larger units such as words, syllables and/or morphemes, they are most commonly found with phonemes (Fromkin, 1973). An early explanation to account for this phenomenon became known as spreading activation (Collins & Loftus, 1975; Quillian, 1962), which initially used semantic priming, i.e., the speeded or slowed access to a lexical item due to a previously exposed semantic neighbor. The concept of spreading activation was further developed between and within levels of speech production (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), and has since become one of the core principles of lexical access for both perception and production. It was later discovered that errors are also more common within a given level rather than between them (Dell, Reed, Adams, & Meyer, 2000).

In the literature devoted to the process of lexical retrieval where phonologically similar words are said to vie for selection, the unit by and large has been the phoneme. Early work in this area began with the Cohort Model (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978) which stressed the linear experience of word recognition in which syllabic units such as onsets and rimes activate candidates for lexical selection, lending increased strength to the onset. For example, if an input word is ‘car’, then possible cohort competitors could include ‘cat’, ‘kiwi’, and ‘Corvette’, while excluding similar sounding words that do not share an onset, such as ‘bar’, ‘far’, and ‘jar’. The evidence to support the Cohort Model came from cross-modal priming studies that found that words with the same rimes but different onsets, for example ‘honing’ (honey), and ‘woning’ (house) did not significantly speed performance (Marslen-Wilson & Zwitserlood, 1989), while activation amongst the same onset cohorts were found to prime semantic associates but not for those words that had semantic associates of the same rime (Connine, Blasko, & Titone, 1993). Evidence to dampen the relevance of the onset in spoken word recognition came from different directions. In a similar cross-modal priming study as the studies done by Marslen-Wilson and colleagues, Allopenna, Magnuson, and Tanenhaus (1998) found evidence for significantly more rime competition than fillers or onset cohorts. Meanwhile, greater focus turned to the phonologically derived Levenshtein edit distance, which is a whole word similarity metric that identifies 'neighboring' words through the addition, deletion or substitution of a single phoneme (Greenberg & Jenkins, 1967; Landauer & Streeter, 1973). Within this metric, currently referred to as phonological neighborhood density (PND), 'car' is a neighbor not only of 'bar', 'far', and 'jar' (substitutions), but also 'are' (deletion), and 'scar' (addition). In auditory lexical decision and phoneme identification tasks, PND effects were found for words of the one-phoneme change rule, but not for those of the onset cohorts (Newman, Sawusch, & Luce, 2005), leading to the conclusion that while identification begins linearly at the onset, the onset alone does not outweigh the strength of whole word activation.

1.1.1 Phonological Neighborhood Density

PND took a central role in the canonic Neighborhood Activation Model (NAM: Luce & Pisoni, 1998). NAM sought to account for the perception of auditory lexical items through the co-activation of phonologically similar words. A principle contribution of NAM was the neighborhood activation equation, in which the probability of mistaking one English CVC monosyllable for another is combined with the target word's PND value, lexical frequency, and lastly its neighborhood frequency (the averaged lexical frequencies of a target word's neighbors). The equation predicts that the more neighbors a word has and the higher in frequency those words are, the harder recognition of that given word is due to increased difficulty in resolving lexical selection. This effect, commonly referred to as lexical competition, is seen in the behavioral response of slower reaction times. Conversely, a word with few neighbors that are low in frequency would produce shorter reaction times. Perhaps due to the difficulty of attaining probability counts for each syllable and their combinations, which would accordingly need to be repeated within differing speech communities and over time due to changes in accent, the single-edit PND metric has been far more productive than the NAM equation.

The PND metric has been used to study a broad list of phenomena, including word learning (Charles-Luce & Luce, 1990, 1995; Storkel, Armbruster, & Hogan, 2006) visual word recognition (Yates, Locker, & Simpson, 2004). short-term (Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002) and long-term memory (Sommers & Lewis, 1999), and conversational speech (Gahl, Yao, & Johnson, 2012), to name a few. As with many phenomena related to mental processes, phonological neighborhood effects differ along a perception/production split. In the literature on spoken word recognition, through auditory tasks such as lexical decision (Luce & Pisoni, 1998; Vitevitch & Luce, 1999) and word repetition (Vitevitch & Luce, 1998, 1999; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997), words with many neighbors (dense words) are recognized more slowly and less accurately than words with few neighbors (sparse words). This inhibitory effect due to greater density of phonologically similar words reflects competition between a given target item and its neighbors in long-term memory. Multiple connectionist models of spoken word recognition account for this effect, including TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000). The opposite is the case in speech production, wherein greater density speeds word production (Vitevitch, 2002; Vitevitch & Sommers, 2003), and proportion of speech errors (Stemberger, 2004; Vitevitch, 2002; Vitevitch et al., 1997; Vitevitch & Sommers, 2003). Multiple studies have modeled the facilitative effect of greater phonological similarity on word production using connectionist models that include co-activation between the phonological and lemma levels of processing (Dell, 1986; Dell & Gordon, 2003; Dell et al., 1997).

No account of lexical access would be complete without connecting lexical selection with working memory. The role of working memory in relation to PND has been investigated in serial recall (Oberauer, 2009; Roodenrys & Miller, 2008; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002; Sommers & Lewis, 1999). The serial recall task brought with it the discussion of redintegration (Schweickert & Boruff, 1986), which is the concept that partial or degraded memory traces of lexical items in short-term memory are reconstructed through access with items in long-term memory. Roodenrys et al., (2002) found that compared to sparse words, words from dense neighborhoods showed better recall, were less likely to be omitted, and more likely to be recalled in the wrong position. Following the narrative thus far on PND, greater density implies greater activation of representations in long-term memory. With a competition account of their results we would expect dense words to show the exact opposite trend. Indeed, Roodenrys and colleagues stated that their results were contrary to prediction, likely because they had depended on the directional effect reported in Luce & Pisoni (1998), where dense words are recognized more slowly.

1.1.1.1 The Question of Polarity

Why would spoken word recognition and production differ in the effect of the co-activation of form similarity between whole words? Dell and Gordon (2003) put forward an account based on the cognitive demands specific to either speaking or listening. While speakers intend to convey meaning and then accordingly select a fitting form, listeners encounter a given form from which they must associate to a certain meaning. For speakers, it is expected to find semantically derived errors in picture naming, rather than phonological ones (Dell et al., 1997). For listeners, errors of a phonological rather than semantic nature are dominant in auditory tasks (e.g., Vitevitch, Chan, & Goldstein, 2014). Their account lies in the respective strength of activation due to the nature of the task. Phonological neighbors are the strongest neighbors of a target word in an auditory task. Having more of them increases feedback between the lexical and phonological levels, converting neighbors into competitors.

Yet, semantic neighbors, which are the strongest neighbors in a production task, facilitate reaction times due to weaker activation of phonological neighbors. Chen and Mirman (2012) replicated this account of activation strength between tasks with a similar interactive activation model that was also switched in direction to simulate recognition (i.e., from phonological, lexical to conceptual) and production (i.e., from conceptual, lexical to phonological). They concluded that lexical items are subject to both inhibitory and facilitative effects, yet inhibition occurs when net activation is strong (i.e., increased interactivity between representations), and facilitation occurs when net activation is weak (i.e., decreased interactivity between representations).

The narrative accounts of the interactive activation models, first put out by Dell and Gordon (2003) and later supported by Chen & Mirman (2012), suffer from two weaknesses. The first of which lies within their explanation of the PND effect in speech production. Neither account explains how a greater number of phonological neighbors is to somehow speed reaction times in picture naming. By their own accounts, facilitation in production occurs despite the existence of phonological neighbors, not because of them. If in production there is a weaker net activation, because phonological neighbors lead to less feedback between levels during a meaning-driven task, we should expect to find a null PND effect in picture naming, not a deterministic speeding of reaction times. The literature does indeed provide some null findings where we would expect them (Gordon & Kurczek, 2014;

Jescheniak & Levelt, 1994; Vitevitch, Armbruster, & Chu, 2004), and some non-significant results that might have been due to unrelated issues, such as mixing photographic and hand-drawn stimuli or due to naming pictures that represent conceptual processes such as verbs (Newman & Bernstein Ratner, 2007; Tabak, Schreuder, & Baayen, 2010). The second problem with the polarity switching narrative rests on the false assumption that there is a consensus in the literature for modality and task specific neighborhood effects. As noted above, given the predictions made possible through the PND literature, greater PND should inhibit not facilitate performance in serial recall (Roodenrys et al., 2002). This contradictory finding has not been restricted to the interaction between PND and working memory. A facilitative effect of PND in recognition was found in auditory lexical decision (Vitevitch & Rodríguez, 2004), while inhibitory effects of PND in production have been found in picture naming (Arnold, Conture, & Ohde, 2005; Sadat, Martin, Costa, & Alario, 2014; Vitevitch & Stamer, 2006, 2009).

Several hypotheses have been made concerning these differences in polarity from the body of behavioral evidence. The first hypothesis places the nexus at psychotypology, i.e., the linguistic differences between the languages being tested (Vitevitch & Stamer, 2006). While it has been predominantly investigated in English, PND has also been shown to effect aspects of speech processing in other languages, including Dutch (Frauenfelder, Baayen, & Hellwig, 1993), French (Muneaux & Ziegler, 2004; Ziegler, Muneaux, & Grainger, 2003), German (Blumenfeld & Marian, 2006; Marian, Blumenfeld, & Boukrina, 2008), Japanese (Amano & Kondo, 1999, 2000), and Spanish (Baus, Costa, & Carreiras, 2008; Sadat et al., 2014; Vitevitch & Rodríguez, 2004). The language-based hypothesis was posited to account for the findings of opposite PND polarity with Spanish speakers (Vitevitch & Rodríguez, 2004; Vitevitch & Stamer, 2009). It is provocative in that it posits that language specific morphology, such as Spanish having a greater number of phonological neighbors that are also semantic neighbors (e.g., *niñ**o**-niñ**a***, boy/girl), can define aspects of mental architecture and thus behavioral outcomes such as reaction time. The second hypothesis refers to the age of the participants being tested. Gordon & Kurczek (2014) advanced age as a determiner of PND strength to explain why older participants saw an inhibitory effect in a picture naming task, while younger patients exhibited no effect. The proposal that age mediates lexical access fits well with the current understanding of cognitive decline and its relation to processing speed (Salthouse, 2011). That a decline in processing speed would also point towards differential outcomes in neighborhood effects implies that as age increases so does our inability to reduce feedback between processing levels. Finally, the third hypothesis points towards the design and methods employed in PND related studies (Sadat et al., 2014), positing that the contradictory findings could be amended through testing with larger stimuli sets and the use of mixed effects models. The re-analyses of PND studies done by Sadat and colleagues clarified important differences between F tests and mixed effects models in the analysis of a variable that is scalar in nature and thus best fit for regression rather than factorial designs.

1.1.1.2 The Question of Processing Levels

The question of polarity rests unsolved in the literature devoted to PND and thus so does part of our understanding of how mental representations interact with one another in the processing of lexical information. The second outstanding question related to the PND effect is where exactly it takes place from concept, lemma, phonology to articulation. As previously mentioned, co-activation is expected to take place more within a given processing level rather than between levels. Given PND depends upon phonological analyses for word similarity to be calculated, it seems intuitive to assume that co-activation between items would take place at the phonological level. This is an assumption shared by all models to date that have directly considered the role of PND in processing. Paradoxically, it has also been argued to occur lexically (Vitevitch & Luce, 1999) based on the fact that the directional effect was opposite for words and nonwords in an auditory word repetition task; in other words, while words were inhibited by greater PND, nonwords were facilitated (Vitevitch & Luce, 1998, 1999). Their argument was one of weak vs strong activation. The inhibitory effect arose from the strong activation from words that when higher in density also increased competition. Nonwords, in contrast, would lead to weak activation, setting in motion the facilitative sublexical aspects of processing, such as the utilization of probabilistic information of phones and biphones.

Concurrent with the proposal that PND occurs at the lexical level should be evidence of interactivity that is indicative of lexical rather than phonological processes in recognition and production tasks that do not require processing of a purely lexical nature. Existing evidence of such interactions comes from orthographic effects in both auditory recognition and production.

Early on, two differing approaches converged on the apparent ability of the alphabetic system to shape knowledge of phonology. Seidenberg and Tanenhaus (1979) found that judgments of auditory word similarity for phonological neighbors were faster when two target items shared the same spelling (e.g., t**ie** & p**ie**) rather than differing spelling (e.g., r**ye** & p**ie**). That same year, Morais, Cary, Alegria, and Bettelson (1979) showed a greater capacity to add, delete, and substitute the onsets of nonwords for literate when compared to illiterate adults. Later, in auditory tasks that do not require the use of the same meta-phonological skills, such as lexical decision (Ziegler & Ferrand, 1998; Ziegler et al., 2003) and word repetition (Ziegler et al., 2003), orthography persisted in influencing processing of phonological information. While Ziegler and Ferrand (1998) found evidence of lexical competition (slower reaction times and greater proportion of errors) for rimes with greater inconsistency in spelling, Ziegler et al., (2003) contrasted PND and its orthographic version (OND: addition, deletion, or substitution of an orthographic unit) to find a switch in polarity: inhibitory effect to greater PND, and facilitative for greater OND. Ziegler et al. (2003) added to the processing level discussion with an interesting proposal. In their account, evidence for a lexical PND effect is attested in increased lateral inhibition for words with greater phonological neighbors in the bi-modal interactive activation model (Grainger & Ferrand, 1996), and in smaller effects (by this they are referring to the relative size of their F statistics) in word repetition rather than lexical decision. Uniquely, they opined that the OND and PND effects could not coexist at the lexical level, which would force the OND effect to also be inhibitory. Instead, it was a sublexical phenomenon, originating between phonological and orthographic mappings. In an alphabetic language, this would be seen in the correlation between OND and PND, such that a word high in both PND and OND would have shared mappings. The facilitative effect of OND, in their opinion, occurs when this mapping coincides.

1.1.2 Beyond Phonological Neighborhood Density

The use of the edit distance rule as a relational parameter between whole words to study phonological processes has extended beyond just PND. As with NAM that incorporated PND in a larger algorithm, quite a few proposals have been put forward (Albright, 2008; Bailey & Hahn, 2001; Benkí, 2003; Frisch, Large, & Pisoni, 2000; Hahn & Bailey, 2005; Strand & Liben-Nowell, 2016; Suárez, Tan, Yap, & Goh, 2011), although, describing each would be outside of the scope of the current dissertation. I will discuss two ways in which PND has been extended beyond the single-edit rule. The first, phonological neighborhood frequency (PNF) was passed over quickly above, while the second represents a complete set of mathematical measures and a new theoretical view of the mental lexicon.

PNF is calculated through averaging the lexical frequency of a given phonological word's neighbors. It has been studied a limited amount in the production and perception of spoken lexical items, but more so with orthography where the relation between targets and neighbors is defined by the items' written rather than spoken form. An inhibitory effect of high orthographic neighborhood frequency (ONF) has been found in the lexical decision task in French (Grainger, 1990; Grainger, O'Regan, Jacobs, & Segui, 1989, 1992) and Spanish (Carreiras, Perea, & Grainger, 1997; Perea & Pollatsek, 1998) but facilitative in word naming for French (Grainger, 1990) and English (Chris R. Sears, Hino, & Lupker, 1995), while inhibitory for Spanish (Carreiras et al., 1997). Sears, Campbell, and Lupker (2006) later surmised that there were no stable results to be found in English. Andrews (1997) noted both the cross-linguistic aspects between the writing systems of each of the languages tested, but also the decision process inherent in the lexical decision task, previously discussed in Balota and Chumbley (1984). While the narrative as to the effect of ONF appears to be rather confusing in alphabetic languages, in Mandarin, the use of the lexical decision and word naming wherein whole characters constitute a single unit (e.g., 果 *guo3* within 水果

shui3guo3, 芒果 *man2guo3*, 果园 *guo3yuan2*, and 果实 *guo3shi2*), OND has been shown to be facilitative while ONF inhibitory (Li, Gao, Chou, & Wu, 2017; Li, Lin, Chou, Yang, & Wu, 2015; Wu, Yang, & Jin, 2013). As with the PND effect previously discussed, differences in polarity across tasks and possible effects of psychotypology make the simple summary of this aspect of the mental lexicon difficult. The story as to the effect of PNF on spoken word processing is not as complex, however, this is likely due to the fact that neighborhood frequency has been less studied in this domain. Its use with verbal production has been investigated in a corpus analysis of malapropisms (Vitevitch, 1997), and within the picture naming task (Baus et al., 2008; Vitevitch & Stamer, 2006, 2009). Unlike with the story of PND, PNF has the same facilitative directional effect on picture naming in both Spanish (Baus et al., 2008; Vitevitch & Stamer, 2009) and English (Vitevitch & Sommers, 2003). Interestingly, the numerous authors that have studied neighborhood frequency, in both its phonological and orthographic forms, have attributed its effect to the interactive activation model. While the principle of PNF, as a frequency based account of a word's neighbors, implies activity between processing levels, its lack of consensus in polarity is not something the current iterations of the interactive activation model (e.g., Dell & Gordon, 2003; Chen & Mirman, 2010) can adequately explain.

While PNF is beyond the single-edit rule of PND, it does not venture far. It is causally tied to the PND metric because one cannot calculate PNF without first identifying a target word's neighbors. What PNF does that PND does not, is tell us about the characteristics of that word's neighborhood. It was an early look into considering words within an interconnected network wherein constraints on lexical processing, such as frequency of occurrence and sound-similarity, act as simultaneous descriptors. It is a fitting transition into the current use of network science in the study of the mental lexicon.

The second manner in which PND has been used outside of its original scope has been the use of network science, wherein lexical items (nodes) are connected (edges) to one another through the one phoneme difference rule. Graph theory adds to the study of the mental lexicon in that it allows for micro-, meso- and macro-levels of analysis. Using 20,000 English lemmas from which a large interconnected component was identified, Vitevitch (2008), showed that the network exhibited macro-level features. The first of which, referred to as 'small world networks' exhibit both a short average path length, and high clustering coefficient (Watts & Strogatz, 1998). A short average path length, or in other terms, the average number of nodes it takes to travel from one extremity to another, is shorter in a small world network when in comparison to a randomly constructed graph. A network's average clustering coefficient is a global assessment of node level interconnectivity. At the micro level, it is a ratio between 0 and 1 that reflects the number of neighbors of a given word that are themselves neighbors of each other. If a target word has 10 neighbors whose neighbors are all interconnected then that word's clustering coefficient (C) equals 1. Likewise, if a target word has 10 neighbors that are not themselves neighbors then that word's C equals 0. In perceptual identification, words high in C were identified correctly significantly more than low C words, while in auditory lexical decision, high C words were responded to slower than words low in C (Chan & Vitevitch, 2009). Interestingly, the effect of C on speech production has thus far revealed greater errors in a corpus of speech errors for high C words, while slower reaction times in a picture naming task (Chan & Vitevitch, 2010).

The second feature found in the English graph constructed by Vitevitch (2008) pertains to assortative and disassortative mixing by degree, which is a measure that reflects the density of nodes with the density of their neighbors. Assortative mixing by degree is the probability that highly connected nodes are connected to other highly connected nodes, while disassortative mixing by degree is the state in which nodes with many neighbors are connected to nodes with few neighbors (Newman, 2002; Newman & Park, 2003). Values for representing assortative mixing by degree come from correlation coefficients that typically fall between .1 and .3 for real world networks. While Newman (2002) found a value of .36 for a co-authorship network of physicists, the English graph of phonological words revealed a value of .62. Vitevitch (2008) conceded that this feature of the lexicon is somewhat difficult to account for, given that greater dissimilarity between lexical items (i.e., lower assortative mixing by degree and lower average clustering coefficient) would limit confusability between words. Interestingly, Arbesman, Strogatz, and Vitevitch (2010) found amongst other natural languages values near to that of English and even higher: Basque, 0.72; Hawaiian, 0.56; Mandarin, 0.65; and Spanish, 0.76.

One question worth asking is whether the addition of network science can aid in resolving the question of lexical activation, or more specifically, differences in PND polarity. With the English graph previously mentioned, Vitevitch (2008) proposed that greater density as seen in a high rate of assortative mixing by degree could lead to the selection of non-target words through partially accessed information. Freedman and Barlow (2013) took this to mean a selection advantage for greater PND. Based on findings from tip-of-the-tongue studies (MacKay & Burke, 1990; Vitevitch & Sommers, 2003) in which dense rather than sparse words garnered more segmental description by participants, their premise was that words with greater PND would share partial information more than words low in PND. They also suggested that words high in PND are more likely to be perceived due to positive degree of assortative mixing, i.e., they would be the more likely candidates due to greater available phonetic information being shared with their neighbors. In an auditory word repetition task in which words were embedded in noise, Freedman and Barlow (2008) found evidence of facilitation through 1) a higher accuracy rate for dense rather than sparse words and 2) a tendency for misperceived words to be high in PND irrespective of lexical frequency. This result is contrary to the existing findings (Luce & Pisoni, 1998; Vitevitch & Luce, 1999) where errors tended to be from dense neighborhoods. However, the Freedman and Barlow

(2008) study featured the addition of noise to the stimuli and also did not report on the reaction time analyses.

The network approach takes what would otherwise be a dictionary-like word list and adds complexity, an essential aspect of biological systems. A current shortcoming in the literature is how phonological networks can account for adaptation of complexity. Adaptation can be seen when examined at a long timescale, through the gain, change, and loss over time of vocabulary during development. At a short timescale, it is the transmission and transformation of information during online processing. Two measures of adaptation have gained attention in network science. The first, growth, can be defined as the adding of nodes to a network over time. Second, preferential attachment, describes a relation in which newly added nodes will likely connect to already highly connected nodes, i.e., a scenario in which the 'rich get richer' (Barabási & Albert, 1999; Barabási, Albert, & Jeong, 1999). Together, growth and preferential attachment lead to a power law distribution of the network's degree (i.e., number of neighbors a given node is connected to). Power law distributions occur when few items are high in a featured value, while most are not, giving them a heavy tail. Power-law distributions have long been associated with language. Zipf (1934) found this distribution between lexical frequency and its rank order, in that a word is inversely proportional to its rank in the frequency table (i.e., word number *n* has a frequency of *n*). Critical to the topic of adaptation is that power-law degree distributions have been associated with the phenomenon known as self-organization, wherein a system naturally evolves to a given state from initial conditions that were ungoverned (Bak, Tang, & Wiesenfeld, 1988). Thus, it would appear as though the tools are available through network science to account for adaptation in the phonological network.

The role of adaptation has been suggested in the use of PND and its use within network science. Vitevitch (2008) made a case for preferential attachment by appealing to word learning studies. Evidence for dense words aiding growth in children's vocabulary was found in facilitative effects for words that shared phonological similarity (Beckman & Edwards, 2000; Gathercole, Hitch, Service, & Martin, 1997; Storkel, 2001, 2003; Storkel & Morrisette, 2002), and a facilitative effect of high PND for young adults (Storkel et al., 2006). Curiously, the English lexicon did not display a power-law degree distribution. Thus, while adaptation in the mental lexicon through PND is likely as seen in word learning, it is not at all obvious how it is involved. In light of this difficulty, it is not arbitrary that the field has depended on connectionist models to simulate complex adaptation. They allow for words to enter a process wherein intermediary representations (hidden layers) dynamically transform information to appeal to experimental findings. As has been discussed, the existing models do not grasp what is currently unfolding in the PND literature, motivating alternative methods that can apprise complexity while also dynamic adaptation.

1.2 COMPLEX ADAPTIVE SYSTEMS

Complex Adaptive Systems (CAS) is an overarching theory and collection of methods originating from physics and mathematics. Examples of CAS range widely to include biological phenomena such as developing embryos (Kauffman, Shymko, & Trabert, 1978), and the formation of leopard's spots (Murray, 1988), to include socially constructed phenomena such as ant colonies (Detrain & Deneubourg, 2006), and technological inventions (Fleming & Sorenson, 2001). The first descriptors of a CAS are easily relatable to a phonological network in that agents (words) within a system are interconnected without a centralized governing mechanism (i.e., distributed control). The next characteristic is the principle of sensitive dependence on initial conditions, sometimes referred to as the 'butterfly effect'. A system exhibits this sensitivity when small changes to the system result in large behavioral differences, or vice versa, a large insult to the system has a negligible effect. This introduces uncertainty into the system, decreasing its predictability. CAS are said to exhibit self-organization. As mentioned above, this is the occurrence of order from an initially disordered state due to local interactions under no governance. Self-organization, and the unpredictability of sensitivity to initial conditions influence how a system is adaptive. CAS, as implied in its adaptation, changes over time and thus are subsequent to certain dynamics. They are said to be self-sustained (not dependent on outside stimuli and thus active in and of itself), moving between transient states (temporary stability) and constant fluctuation due to competing dynamics of inhibitory and excitatory changes. From these conditions arise patterns that extend beyond the characteristics of each agent within the system. This phenomenon, known as emergence, takes on qualitatively new properties that cannot be predicted from its individual components, or in other words, we see in emergence the old adage, 'the whole is not the sum of its parts'.

In the language sciences, CAS has found a productive home. One school of thought that takes what is known as a usage-based approach, sees language as a network of categorized language use developed through experience that is necessarily social. Experimentally, many of these principles can be seen in the iterative learning paradigm (Beckner, Pierrehumbert, & Hay, 2017; Cornish, 2010; Cornish, Tamariz, & Kirby, 2009; Kirby, Cornish, & Smith, 2008; Reali & Griffiths, 2008; Smith & Wonnacott, 2010), wherein the output of a participant attempting to learn an artificial grammar (with errors and innovations included) is passed on to the next participant whose output is then passed on to the next *n* number of times. What has been claimed in these studies is that participants adapt an initially random and unstructured collection of words to successively more regularized forms (i.e., increasing entropy), leading to an emergence of structure that was not present in the initial vocabulary. The second approach in language science to implement methods from CAS focuses on the system's dynamics, primarily between first and second languages (de Bot, Chan, Lowie, Plat, & Verspoor, 2012). The experimental work from this group of studies reveal the non-linear nature of language processing and learning. Language, as a system, exhibits instability in lexical representations, as seen in individual variability (de Bot & Lowie, 2010), according to an individual's chronotype (i.e., between 'early risers', an 'night owls': de Bot, 2015), and as evident in language attrition (Schmid, 2013). While there is evidence of stability in processing and acquisition, these studies show that there is no end state, but rather constant fluctuation and change.

1.3 MANDARIN

Mandarin Chinese has become a recent focus in the psycholinguistic literature due to set of linguistic features that test the limits of models of speech processing that were developed through European languages. Perhaps the most unique is the status of the syllable, which is tonal, of equivalent size to the primary orthographic unit, and highly homophonic. Unlike English or Dutch, which both have over 10,000+ syllables, the Mandarin syllable inventory is small, featuring ~1,300 syllables plus tone and ~400 without tone. Excluding a select number of high frequency lexical items that do not regularly carry tone and thus are annotated as tone 0 (ex. 的 de0, and 我们 wo3 men0) each syllable carries one of four tones: tone 1 (high level pitch, 55), tone 2 (low rising pitch, 35), tone 3 (low dipping pitch, 214), or tone 4 (high falling pitch, 51). Aside from the dialectal phenomena known as erhua (e.g., Lee, 2005) in which the character 儿 (er2) is added to another character yet pronounced as a single syllable (玩儿, wan2 er2 = war2), each syllable in the inventory matches one or more Chinese characters. Mandarin has been shown to be largely disyllabic in nature, in fact it has been calculated that two-thirds of all Mandarin words consist of two characters (Li et al., 2015; Wu et al., 2013). Yet, characters that do not exist as monosyllabic words, meaning they only exist in multisyllabic words, are still lexical items that contribute to the count of homophone neighbors. In context, the same roughly 1,300 tonal syllables service all lexical combinations from monosyllabic to multisyllabic words. This leads to a homophone density (i.e., the number of homophone neighbors a given word has) of up to 48 when tone is considered (Wang, Li, Ning, & Zhang, 2012), as seen with the syllable yi4 (e.g., 意, 易, 亿, 衣, 亦). To put this in context, 11.6% of Mandarin words have homonyms, compared to 3.15% in English (Wen, 1980). High homophony has been shown to lead to lexical competition in spoken word recognition, as seen by slower reaction times, and lower accuracy (Chen, Chao, Chang, Hsu, & Lee, 2016; Wang et al., 2012). This is uniquely important to Mandarin given the relation of orthography to the syllable.

When dealing with surface features of a language, one option is to disregard their possible import to cognition in general. The opposite would be to consider the possibility that typologically distinct features of a language entail systemic differences between speakers of the world's languages. A growing literature is exploring differences at the cortical level between Chinese speakers and speakers of alphabetic languages. Evidence through the use of inoperative electrocortical stimulation during picture naming and number counting has shown that while there is large individual variation between speakers, the language probability maps between Mandarin and English speakers do not entirely overlap (Wu et al., 2015). Studies examining the mapping between language areas during reading (Cao, Brennan, & Booth, 2015; Cao & Perfetti, 2016; Li et al., 2017; Siok, Niu, Jin, Perfetti, & Tan, 2008; Siok, Perfetti, Jin, & Tan, 2004; Tan et al., 2000, 2001; Tan, Laird, Li, & Fox, 2005) and listening (Brennan, Cao, Pedroarena-Leal, McNorgan, & Booth, 2013; Ge et al., 2015) are beginning to explain why. Brain regions dedicated to reading differ between English and Mandarin speakers likely due to the repetitive physical writing associated with learning Chinese characters (e.g., Cao & Perfetti 2016). The respective mapping between orthographic forms between English, with its 26 letters, and Mandarin, with its 10k+ Chinese characters decomposed without regularity through components known as radicals, has shown differential effects on access of phonological representations. Brennan et al. (2012) showed that during rhyming judgment tasks, the Chinese mapping of a large number of orthographic units to a small number of syllables showed little orthographic influence in phonological processing. In contrast, an orthographic influence likely due to the small number of orthographic units to phonological units, was evident in increased activation in the superior temporal gyrus for English speakers. Their findings supported previous behavioral (e.g., Pattamadilok, Morais, Ventura, & Kolinsky, 2007; Rastle, McCormick, Bayliss, & Davis, 2011; Ventura, Morais, Pattamadilok, & Kolinsky, 2004; Ziegler, Ferrand, & Montant, 2004) and neural (Booth et al., 2002; Pattamadilok, Knierim, Kawabata Duncan, & Devlin, 2010; Pattamadilok, Perre, Dufau, & Ziegler, 2009) evidence of orthographic interference in English, and the lack of this in Mandarin (de Gelder & Vroomen, 1992; Zhou & Marslen-Wilson, 1999).

1.3.1 Lexicality

Sinologist Jerry Norman (1988), when discussing the role of the syllable in the Chinese language family, stated, "…any one dialect contains a fixed number of possible syllables" (p.138). While this statement is a categorical impossibility, seeing as languages would never change if novel syllables were not added or extant syllables lost, it captures a psychologically real constraint. Unlike in an alphabetic language where a novel syllable can be ascribed to a series of letters that themselves allow for pronunciation and the easy transmission of said syllable to a multitude of speakers, Chinese orthography does not lend itself to the required transparency to achieve this. Thus, while it is not the case that the syllable inventory is fixed, it is near immutable.[1] For the psycholinguist, this provides a valuable opportunity to use the

[1] The story of *duang1*, tells us of one avenue to novel syllable-hood in Mandarin Chinese. Its birth began with actor Jacky Chan giving voice to an onomatopoeic action (Chen & Devichand, 2015). The new word's popularity caught on after being a center attraction within a parody-like video (绯色 toy, 2015). It was finally ascribed an orthographic form by combining the two characters from Chan's Chinese stage name, 成龙 *cheng2 long2*, into a single character. Note the entire lack of phonological correspondence between *cheng2 long2* and *duang1*. A very important aspect of the transmission of this novel syllable is the fact that pinyin (Mandarin Romanization) was used prior to the creation of the Chinese character.

natural barrier between what is and isn't a phonological word in Mandarin to test hypotheses about lexical processing.

Mandarin uniquely has three classes of nonwords: tone gap nonwords, syllable gap nonwords (Wang, 1998), and systematic nonwords (Wiener & Turnbull, 2015). A tone gap nonword entails the use of an existing syllable from the syllable inventory combined with one of the four lexical tones. An example of this can be seen with the non-tonal syllable, *mei*. This syllable exists combined with three of the 4 lexical tones, *mei2* (ex: 没), *mei3* (ex: 美), and *mei4* (ex: 妹), however no character is assigned to the first tone, *mei1*. They have been used in few studies to date (Myers, 2002; Myers & Tsay, 2005; Wang, 1998; Wiener & Turnbull, 2016; Zhang & Lai, 2010). Myers (2002), investigated word-likeness, and like Wang (1998) before him, found that tone gap nonwords were given higher acceptability ratings than syllable gap nonwords. Myers and Tsay (2005) continued work with word-likeness judgments while accounting for PND, whose values were calculated from a tonal, fully segmented Mandarin syllable (C_G_V_X_T: elicited through personal correspondence). Syllable gap nonwords entail the use of an existing lexical tone combined with phonemes from the phoneme inventory in such a way as to create a non-existent syllable. Thus, while the bare syllables *man*, *men,* and *min* point to existing syllables, *mon,* and *mun* do not. While a syllable gap might be considered a phonotactic violation, for example through the combination of the onset *f* (e.g., 发 **f**a4, 非 **f**ei1) with the rime *ia* (e.g., 下 x**ia**4,

家 **jia**1) to create *fia4*, a systematic gap might place a phoneme in a location in which they never occur such as the placement of the onset *f* into the coda position: *mif.* Wiener and Turnbull (2015) exploited both tone gap and systematic gap nonwords in a word generation paradigm that required participants to produce monosyllabic lexical items from auditory nonwords. They found that tone categories were the preferred unit of manipulation when participants were allowed to freely elect their target unit. Of particular interest to the current dissertation was their significant PND finding wherein greater PND of the stimuli facilitated the production of phonological neighbors. Note that unlike the Myers and Tsay (2005) study in which PND was calculated with the use of the tonal fully segmented schema, Wiener and Turnbull (2015) used a segmented syllable in which all vowel information was collapsed (C_V_C_T: elicited through personal correspondence). The differences in segmentation schemas between the few studies to have implemented PND in Mandarin are the focus of my next discussion on the nature of the Mandarin syllable.

1.3.2 Syllable Segmentation

Researchers have used two methods to describe how segmental units comprise a syllable in Mandarin. One method recognizes a maximum of 4 segments, CGVX, such that C represents initial consonants, G medial glides (/i, u, y/), the V monophthongs, and the final X the second part of a diphthong, or a final consonant. Early accounts disregarded tone preferring instead to propose segmentation schemas based on the constituents of the rime or whether the medial glide constituted a unique phonological role within the syllable: C_GVX (Xu, 1980); C_G_VX (Cheng, 1966; You, Qian, & Gao, 1980); CG_VX (Bao, 1990); and CG_V_X (Ao, 1992; Duanmu, 2007). Note that here an underscore denotes a separation between phonological units. The second method of describing the Mandarin syllable collapses all vowel information, leaving a maximum of 3 segments: C_V_C (Lee & Zee, 2003). One of the first attempts to analyze PND in Mandarin in fact used this schema to decidedly unclear results (Tsai, 2007). The methods used to arrive at the various schemas, and whose evidence has informed the creation of syllable inventories, come through production tasks that have participants read sentences so as to measure syllable durations (Shih & Ao, 1997; Wu, 2017; Wu & Kenstowicz, 2015), or produce phonological neighbors in rhyming games (e.g., Bao, 1990; Chao, 1931; Yip, 1982).

What is evident thus far in the phonological discussion of syllable segmentation is that a proposal has been made for almost all permutations possible. To muddy the water further is also to acknowledge the multiple accounts of Mandarin syllable inventories (Cheng, 1966; Duanmu, 2010; Lin, 2007; You, Qian, & Gao, 1980; Zhao & Li, 2009) that feature the segments that comprise these syllable schemas. This diverse set of proposals resurrects a rather interesting lesson from the history of phonology. Early on in the emergence of phonological theory Chao (1934) arrived to the debates of competing transcription systems to point out a simple yet poignent insight, known as the non-uniqueness theory. Given that linguists then and now do not agree on the phonemic treatment for the same language, why should we expect a singular system within or across languages? He argued that because there are more than one way to reduce speech to a system of phonemes, there are thus no correct or incorrect versions, but rather those that are either good or bad for a given purpose.

Duanmu (2017) recently revisited the non-uniqueness theory, illustrating that inventories and schemas are intertwined, for example, greater segmentation reduces the phonemic inventory, and vice versa. This can be seen in an example set of syllables [pa, ha, pha]. A ‘fine’ segmentation leads to three phonemes [p, h, a] due to the separation of the stop consonant [p] and aspiration [h], while a ‘course’ segmentation leads to 4 units through the combination of aspiration with the stop consonant [p, h, p^h, a]. Duanmu (2017) calculated the number of units that would exist for speakers of Chengdu Chinese according to four syllable schemas. The ‘finest segmentation’, similar to the above example in which all aspects of speech categorization reflect a unit (CCCVVC, CCCVVV), contained a total of 19 units (12 consonants and 7 vowels). The C_G_VX schema of You et al. (1980), in which rimes are considered single units, had a total of 35 units (19 initial phonemes and 16 rimes). The C_V_C schema (Lee & Zee, 2003), in which all vowel information is collapsed (i.e., G: glides, V: monopthongs, and VV: dipthongs) had a total of 42 units (19 consonants and 23 vowels). Lastly, Duanmu's (2007) CG_V_X schema, which reaches its high number due to complex onsets, had a total of 46 units (39 consonants; 7 vowels).

The importance of the numbers given by Duanmu (2017) is that they represent hypothetical dimensions of a speaker's mental lexicon. The phonemic units housed within a segmentation schema are purported mental categories. Fluctuations in the number of said units would thus have implications either in the time it takes to produce them, or at the very least how that process is interpreted. Differences in this set of categories might also have implications for the interpretation of how native and non-native units are perceived (Best, 1991; Best & Tyler, 2007). The implications of segmentation on speech processing, specifically for Mandarin speakers, has not gone unnoticed. A growing literature is taking place that analyzes speech segmentation within tasks that capture aspects of perception and production. In departure from the theoretical methods of the past, lexical tone is playing a central role in accounting for syllable structure. Of note is that the question of non-uniqueness is still relevant. The non-uniqueness at hand has thus far depended on task modality, i.e., whether greater production or perception is recruited during the task.

O'Séaghdha and colleagues (O'Séaghdha & Chen, 2009; O'Séaghdha, Chen, & Chen, 2010) hypothesized that the first phonological units available for selection below the level of the word or morpheme, titled proximate units, correspond to non-tonal syllables in Mandarin. Their thesis was that unit sizes would vary across languages, granting phonemes and clusters of phonemes in Indo-European languages such as English, while larger units such as morae in Japanese. Speech error analyses have supported this trend, such that in English the dominant unit size is segmental (Fromkin, 1971; Shattuck-Hufnagel, 1979), and in Mandarin syllabic (Chen, 1993, 1999, 2000). For speakers from alphabetic languages, like English and Dutch, sensitivity to syllable onsets between two lexical items has been documented in numerous studies and across multiple paradigms (Jescheniak & Schriefers, 2001; Meyer, 1990; Meyer & Schriefers, 1991; O'Séaghdha et al., 2010; Schiller, 1998, 1999, 2000). These studies show that prior preparation to segmental units shared between lexical primes and target lexical items speeds production of the target word, implying that temporary storage occurs for segmental information. A corresponding series of priming studies have shown syllabic priming results yet no significant onset priming with Chinese orthography in the implicit priming (Chen et al., 2002; O'Seaghdha et al., 2010), and masked priming tasks (Chen, Lin, & Ferrand, 2003; Verdonschot, Nakayama, Zhang, Tamaoka, & Schiller, 2013; You, Zhang, & Verdonschot, 2012). To counter the syllable bias of Chinese characters, similar studies were conducted with picture (Chen & Chen, 2013; You et al., 2012) and auditory stimuli (Chen & Chen, 2013). Supporting evidence for the proximate unit has also been advanced in priming studies with both speakers of Cantonese (Wong & Chen, 2008, 2009; Wong, Huang, & Chen, 2012), and Japanese (Kureta, Fushimi, Sakuma, & Tatsumi, 2015; Kureta, Fushimi, & Tatsumi, 2006; Tamaoka & Makioka, 2009; Verdonschot et al., 2011). While no syllable schema was explicitly proposed by the authors related to the proximate unit proposal, their statement that the primary unit to be selected is non-tonal suggests that either a non-tonal unsegmented schema is the target (CGVX), or its tonal counterpart (CGVX_T) seeing as the syllable is combined with tone prior to production.

To stand in contrast to speech production studies is a growing body of evidence to support the claim that tonal speakers, during speech recognition, process segmental information incrementally and in parallel with tonal information. Two primary methods have been used. The first, combined with recordings of event related potentials (ERP), sought to identify whether categorical differences existed between segmental and tonal processing through the presentation of improbable final monosyllables controlled according to tone, segments and syllables (Brown-Schmidt & Canseco-Gonzalez, 2004; Schirmer, Tang, Penney, Gunter, & Chen, 2005). These studies suffered from the confounding nature of sentence level processing on their N400 measurement, motivating the use of isolated lexical items in the studies to follow. Differential processing between lexical units were then analyzed within a picture-word matching paradigm with both ERP (Malins et al., 2014; Malins & Joanisse, 2012; Zhao et al., 2011) and eye-tracking (Malins & Joanisse, 2010). Malins and colleagues found that whole-syllable mismatches did not produce effects greater than those found with individual components, mirroring results previously found in English (Desroches, Newman, & Joanisse, 2009). This has motivated their claim that processing was segmental, an assertion not supported in Zhao et al., (2011), which found greater evidence for syllable-level processing. One important difference in the latter study however is the fact that they also used Chinese characters during the presentation of their picture stimuli. Their results are likely an effect due to the activation of syllable-sized orthography. One limit to the claims put forward by Malins and colleagues is the fact that while their claim is that Mandarin exhibits incremental and segmental processing, which implies words reside within a tonal fully segmented schema (C_G_V_X_T), they either did not feature mismatch pairs according to glides, nor the X unit (Malins et al., 2014; Malins & Joanisse, 2010, 2012), or did so without regularity (Zhao et al., 2011). Thus, to date these studies only provide evidence for the tonal complex-onset/rime schema (CG_VX_T).

2. EXPERIMENTS

The two models most often referred to in discussions of lexical access (Dell, 1986; Levelt et al., 1999) are comprised of words and morphemes, then phonemes as processing units. Both models depend on a schematic representation of phonological information. In the Dell model, syllable structure is passed down to the phonological level from the representational frame of the higher morphological level, and then filled linearly with phonemes or phoneme clusters. Syllables in the Levelt et al. model are constructed through a linearization process in what's called the metrical frame. O'Seaghdha and Chen (2009, 2010) posited that the Dell and Levelt models do not generalize beyond the European languages they were created to model. They proposed, through the proximate unit principle, that whereas words are selected phonemically in English they are retrieved as syllables in Mandarin. These studies stand in contrast to the perception studies by Malins and colleagues, wherein segmentation has been found to occur incrementally. Through the lens of network phonology, I proceed under the modeling assumption that a given target word within the metrical frame of the Levelt model, or the phonological representation frame of the Dell model, would share the same segmentation properties as the words they are connected to in long-term memory.

Modifying the schematic representation of the syllable modifies each words' selection of neighbors from the lexicon. Given a phonological word, for example the pinyin syllable *xiang4* (example character: 向), the neighbors for the tonal fully-segmented schema (C_G_V_X_T) differ from its non-tonal equivalent (C_G_V_X) by three items (*xiang1*, *xiang2*, and *xiang3*). Differences in what constitutes a neighbor would accordingly arise across all segmentation schemas described above due to both lexical tone and combinations of segments. PND thus offers a unique method in the study of syllable segmentation. Through the identification of the phonological network that corresponds to levels of co-activation of neighbors within a given task, we identify the segmentation of the syllable.

In Table 1 I illustrate 16 segmentation schemas. All of them, save for one, have been explicitly proposed for Mandarin in the literature or suggested in a narrative explanation of experimental findings. In order to reach all permutations, I add the fully segmented diphthongal schema, in its nontonal, C_G_V_C, and tonal form, C_G_V_C_T. The non-tonal schema was proposed for Taiwanese speakers by Lin (1989), based on participant productions of neighbors during a language game. Evidence for diphthongs in Mandarin is also supported in Wan (2006).

Table 1. Mandarin segmentation schemas according to the example monosyllables, lian2 (e.g., 联) and liao3 (e.g., 了)

Without Tone			With Tone		
C_V_C	/l_iɛ_n/	/l_iaʊ/	C_V_C_T	/l_iɛ_n_35/	/l_iaʊ_214/
C_G_V_C	/l_i_ɛ_n/	/l_i_aʊ/	C_G_V_C_T	/l_i_ɛ_n_35/	/l_i_aʊ_214/
C_G_V_X	/l_i_ɛ_n/	/l_i_a_ʊ/	C_G_V_X_T	/l_i_ɛ_n_35/	/l_i_a_ʊ_214/
C_G_VX	/l_i_ɛn/	/l_i_aʊ/	C_G_VX_T	/l_i_ɛn_35/	/l_i_aʊ_214/
C_GVX	/l_iɛn/	/l_iaʊ/	C_GVX_T	/l_iɛn_35/	/l_iaʊ_214/
CG_V_X	/li_ɛ_n/	/li_aʊ/	CG_V_X_T	/li_ɛ_n_35/	/li_aʊ_214/
CG_VX	/li_ɛn/	/li_aʊ/	CG_VX_T	/li_ɛn_35/	/li_aʊ_214/
CGVX	/liɛn/	/liaʊ/	CGVX_T	/liɛn_35/	/liaʊ_214/

The goal of the current dissertation is to use phonological networks to identify syllable segmentation patterns in commonly used tasks in the PND literature. I ask whether additional segmentation information apprises us of what is happening in the featured tasks, and more generally during Mandarin lexical access. Through five experiments, I address the question of PND polarity and the question of whether the PND effect is phonological, lexical, or other. In my interpretation of the dissertation's findings, I conclude with an appeal to complex adaptive systems in order to account for adaptation in processing and the emergence of linguistic structure due to processing demands.

In Experiments 1 and 2 the issue of phonological similarity in Mandarin is addressed through the use of participant created phonological associations in what has been referred to as the neighbor generation task. The purpose of these tasks is to identify a syllable inventory that corresponds to participant-created minimal pairs, i.e., phonological neighbors. From the optimal syllable inventory, a database is constructed that contain PND values and their corresponding lexical statistics for each of the segmentation schema shown in Table 1. The second neighbor generation task is analyzed according to participants' reaction times and the PND values derived from the novel database of neighborhood statistics. Considering previous research with this task that utilized PND, in both Mandarin (Wiener & Turnbull, 2015) and French (Muneaux & Ziegler, 2004), a facilitative effect is expected to greater density. Due to the productive nature of the task, an unsegmented syllable is expected, however, that might belong to either the tonal (CGVX_T) or nontonal (CGVX) variant.

In Experiment 3, an auditory shadowing task is performed and analyzed in the same manner as the second neighbor generation task. This task was performed in order to test the core findings of NAM wherein auditory word repetition was also implemented. Based on their findings, an inhibitory effect is expected to increased density, as seen with English speakers in prior studies (Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999). Due to the recruitment of perceptual processes associated with auditory word repetition, a tonal and segmented schema is predicted. An HD effect is also likely given prior evidence in Mandarin during auditory tasks (Chen et al., 2016; Wang et al., 2012).

In Experiments 4, I revisit auditory word repetition with monosyllabic and disyllabic stimuli to address unanswered questions from Experiment 3 and the role of syllable length on co-activation of form similarity items. Experiment 4 aids the discussion on the question of polarity, while addressing the possible influence of working memory on polarity. Despite previous theories about the role of working memory in lexical processing, to date, this task provides the first evidence of a PND and working memory interaction.

In Experiment 5 the issue of lexical status in Mandarin will be addressed through the use of an auditory lexical decision task. PND values ascribed to Mandarin monosyllabic words, tone gap nonwords, and syllable gap nonwords identify access to similarity neighborhoods that differ according to the item's status. These experiments provide novel evidence towards how form-similarity affects both judgment and recognition of lexical items for Mandarin speakers in a manner that is not applicable to other languages tested thus far. They also set the stage for a re-interpretation of the question of polarity in PND studies. An inhibitory effect to words high in PND is predicted based on prior auditory lexical decision research (Luce & Pisoni, 1998), accompanied by a switch in directional effect for the nonwords (Vitevitch & Luce, 1998, 1999). Considering this task has long been considered a perceptual task, it is expected that the schema identified be tonal and segmented.

2.2 METHOD

The methods in this dissertation are statistical in nature, and specifically suited for the analysis of behavioral experiments in which the dependent variable is reaction time. A further specification is that all present statistical analyses depend upon a database of lexical statistics. The method I present is a validation technique somewhat novel to psycholinguistics.

The tools used however are not novel. They exist in all major statistical tools/programs currently available. I will be accessing them through mixed effects models in the R packages, lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) and mgcv (Wood & Scheipl, 2016).

The goal of the current methods is to aid in the identification of an optimal segmentation schema through the lexical statistics that are tied to them. From the outset, this implies the identification of an optimal model in comparison to many other models. I first address the common problem of colinearity between linguistic variables and the methods used by psycholinguists to avoid it. I then introduce the use of composite variables causally tied to PND. Finally, I give examples[2] of the use of model selection with multiple single-predictor models, while illustrating how they can be used in hypothesis testing.

It is well known that colinearity is highly common in linguistic data, however, when present amongst independent variables (i.e., fixed effects within a mixed effect model) it negatively effects model output. There are several methods in use to deal with colinearity. One such method is to residualize one variable against another and then use the residuals in place of the original values. A researcher might elect to residualize two variables that correlate to artificially evaluate their independent effects. This is a practice that has made its way to several PND studies (Gordon & Kurczek, 2014; Sadat et al., 2014; Storkel, 2004). Conceptually we can see residualization as a means to mask the fact that colinearity between

---

[2] The values pertaining to each model output in Tables 2-5 were created based on possible model outputs and should not be understood as representatives of actual experiments

two variables is an indication that they share variance because they are in some way representing a similar process, or possibly the same underlying effect that is being described by two or more variables. Mathematically, however, it has recently been shown to 1) have no change on the outcome of the residualized variable, 2) possibly flip the direction of the original effect in more complex models, 3) not actually create a 'independent' version of a variable, and 4) ultimately make analysis difficult in that residualization replaces colinearity with a lesser understood problem (Wurm & Fisicaro, 2014). Another method to avoid colinearity is through data reduction by creating principal components from several variables. The use of principal components with language data was detailed in (Baayen, 2008), and recently given a greater explanatory role in (Baayen, Milin, & Ramscar, 2016). The goal of creating principle components is to convert a given number of correlated variables into an experimenter-elected number of uncorrelated variables through orthogonally redistributing variance. This method was not chosen due to the rather abstract and theory-driven discussion that follows such an analysis.

To illustrate the goals of the current analysis, I refer to the nature of the principle variables under scrutiny. In the literature, PND and PNF have been treated as related yet independent variables despite their causal relationship. It is not possible to calculate PNF without having already calculated PND. To remind the reader, PNF describes the likelihood of a given word to be affected by its neighbors due to their frequency of occurrence, i.e., the average frequency of the neighbors that PND represents. It would be more apt, given the current understanding of phonological networks, to state that PND and PNF respectively provide a one-dimensional account of a multidimensional network. Researchers have limited the effect of one or other of the two given variables so as to isolate their respective effects. While this approach makes sense if a one-dimensional account is the explicit goal, it does not make sense if the goal is to identify an underlying structure of multiple dimensions. In the current methods, I assume PNF and PND to be one-dimensional accounts of an underlying schematic structure defined by syllable segmentation.

The route I chose to deal with colinearity while simultaneously meeting the theoretical goal of identifying a dominant syllable segmentation schema through statistical analyses involves combining variables of like effect to create composite variables, a method previously used with language data (Baayen, Feldman, & Schreuder, 2006; Wurm, 2007), and a key method in accounting for effects of PND (Bailey & Hahn, 2001; Luce & Pisoni, 1998). As with the conceit of PND and PNF respectively being one-dimensional representations of the same phonological network, through the combination of variables I am able to increase the number of dimensional representations. The original variables from which I construct composites are PND, PNF, Freq (log10 lexical frequency), and HD (homophone density).

HD and PND consist of the same type of linguistic object, i.e., whole numbers representing lexical items. It is logical then to combine them in order to represent phono-orthographic density (POD). Through POD we are positing that orthographic representations are activated simultaneously with phonological neighbors. Considering the current state of research into HD with Mandarin speakers, a POD effect is highly likely.

Freq and PND are related, but not equivalent measures. Marian et al. (Marian, Bartolotti, Chabal, & Shook, 2012) illustrated across five European languages a tendency for shorter words to have more neighbors and words with greater density to have higher lexical frequency. Freq has repetitively been shown to facilitate language processing. Thus in combining frequency with PND we follow the neighborhood activation equation (Luce & Pisoni, 1998) where Freq is multiplied with PND (fPND). This multiplication accordingly increases the magnitude of highly frequent high-density words while at the same time increasing the spread between them and low-frequency, low-density words. Along the same lines is the multiplication of PNF with PND values (nfPND), which has as its premise the weighting of phonological neighbors by those words whose neighbors are more frequently encountered, and like with fPND, multiplication increases the spread and number of its original values. The same applies to our POD measure above, wherein either the target word's frequency magnifies the spread of combined HD and PND (fPOD), or the target word's phonological neighbor's frequency magnifies the spread of combined HD and PND (nfPOD).

Composite variables alongside the original variables, each as models with a sole predictor, will then be ranked according to model selection. Model selection uses each fitted models' AIC (Akaike's Information Criterion), a measure of how well the given mixed effects fit the data according to maximum likelihood estimation, or BIC (Bayesian Information Criterion), which similarly uses likelihood estimation but is said to perform better with larger datasets. For an in-depth description of AIC, see Burnham and Anderson (2002), and for BIC, see Schwarz (1978). Once all models have been given and AIC/BIC value, those models can then be ranked: lowest value equals best fit. The optimal use of AIC/BIC involves all models differing by only 1 fixed effect. Thus, the manner in which model selection avoids our current concern of colinearity, is to rank all original and composite variables, per each segmentation schema. For the current experiment this amounts to 9 predictors (original plus composite variables), per each of the 16 segmentation schemas, totaling 144 single predictor models. The alternative would be to place all 144 variables in the same model as fixed effects wherein high colinearity between just two would render the model instable. Another alternative would be to use model selection with 16 models (one per each segmentation schema), each with 9 predictors. This approach would evaluate which schematic representation was a best fit, however, without giving a clear understanding of shared variance between individual predictors while still being subject to the issue of colinearity. The method of ranking fitted models with only 1 predictor simultaneously avoids instability while also displaying how shared underlying effects of like predictors can benefit theory building.

Single-predictor models that share an underlying effect produce similar AIC/BIC values, and when ranked, group together. This grouping together of similar models can serve to test hypotheses. The first model outcome scenario I present is a positive identification of a segmentation schema. If in the model selection output the top ranked models belong to a single schema, comprising multiple individual predictors, that tells us that the structural aspects of the schema outweigh the importance of a single variable. In Table 2 I have highlighted the top 9 ranked models in an example model selection output. It is convention to include model output such as the model's estimate (Estimate), standard error (SE) and degree of freedom (df), although they will not be necessary in this discussion. For each example model selection output, I will use only the AIC values, rather than both AIC and BIC. The third column features the AIC ranking. Because the difference, not the absolute value, is what is important in evaluating AIC, I include that difference metric as dAIC. The final two columns pertain to the t and p values. The model's t value details the strength of the individual predictor's effect and its slope. With the current variables, a negative slope means that there is a facilitative effect (greater values sped reaction times), and a positive slope means that there is an inhibitory effect (greater values slowed reaction times). The model's p value is calculated from the t value and generally required for reporting findings.

Table 2. Example model selection output illustrating a dominant schema

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
PND	CGVX_T	0	-25.37	6.84	177.73	-3.70	0.0001
PNF	CGVX_T	0.52	-24.89	6.85	175.93	-3.60	0.0002
nfPND	CGVX_T	0.86	-24.42	6.80	195.51	-3.58	0.0003
nfPOD	CGVX_T	1.02	-24.43	6.87	175.49	-3.55	0.0004
fPND	CGVX_T	1.43	-24.02	6.87	177.21	-3.50	0.0005
fPOD	CGVX_T	1.74	-23.73	6.88	175.48	-3.45	0.0006
POD	CGVX_T	2.15	-23.33	6.89	174.90	-3.39	0.0007
HD	CGVX_T	3.60	-21.63	6.86	199.60	-3.15	0.0010
Freq	CGVX_T	5.92	-19.00	6.92	188.22	-2.75	0.0020
Freq	CGVX	8.45	-15.55	7.02	175.58	-2.22	0.0030
fPND	C_V_VX	8.68	-15.15	7.00	182.87	-2.16	0.0300
fPOD	C_GVX_T	8.71	-15.14	7.02	176.99	-2.16	0.0300
Freq	C_G_V_X	8.94	-14.76	7.03	175.57	-2.10	0.0300
fPND	C_GVX	8.94	-14.73	7.01	183.78	-2.10	0.0300
fPOD	CG_V_X	9.27	-14.20	7.03	177.64	-2.02	0.0400

In Tables 2 through 4, I have highlighted the dAIC value of 2.15. The rule of thumb when using AIC is to claim sufficient difference between two models when that difference is greater than 2. Values less than 2 between models tell us that the difference between them is negligible. In the current example, the tonal unsegmented schema (CGVX_T) outperformed all other models with a dAIC of 8.45 between the next model not belonging to that schema. Meanwhile within the CGVX_T schemas, HD and Freq are the lesser influential dimensions of the network, while PND and PNF are the dominant aspects accounting for lexical processing.

Note that in the current example model selection outputs, Subject and Item are random effects, as is commonly expected in the psycholinguistic literature. dAIC and the t and p values of a single predictor model will not always correspond, meaning a higher ranked model (with lower dAIC) might have a lower t value than a lower ranked model. This would likely be due to random effects behaving differently with different single predictors, as AIC evaluates model performance including both fixed and random effects, while the t and p values assess only individual fixed effects. Given a discrepancy of this sort, the dAIC value between two such models would likely be negligible (i.e., dAIC < 2)

The next possible model outcome scenario I present is when an individual predictor accounts for lexical processing rather than a segmentation schema. If, in the model selection output, the top ranked models belong to a single predictor, regardless of segmentation schema, that tells us that aspects of the given variable outweigh the importance of similarity due to a specific schema. In Table 3 we see that PND is the dominant explanation for the facilitative effect (i.e., negative slope). I have highlighted three POD variables to illustrate the similarity between POD and PND. Remember that POD is causally tied to PND in that it is the result of adding the number of homophone neighbors to the PND count.

Table 3. Example model selection output illustrating a dominant variable

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
PND	CGVX_T	0	-25.37	6.84	177.73	-3.70	0.0001
PND	CGVX	0.52	-24.89	6.85	175.93	-3.60	0.0002
PND	C_V_C	0.86	-24.42	6.80	195.51	-3.58	0.0003
PND	CGVX_T	1.02	-24.43	6.87	175.49	-3.55	0.0004
PND	CG_VX	1.43	-24.02	6.87	177.21	-3.50	0.0005
PND	C_G_V_X	1.74	-23.73	6.88	175.48	-3.45	0.0006
POD	CGVX_T	2.15	-23.33	6.89	174.90	-3.39	0.0007
POD	CGVX	3.60	-21.63	6.86	199.60	-3.15	0.0010
POD	C_V_C	5.92	-19.00	6.92	188.22	-2.75	0.0020
Freq	non-tonal	8.45	-15.55	7.02	175.58	-2.22	0.0030
fPND	C_V_VX	8.68	-15.15	7.00	182.87	-2.16	0.0300
fPOD	C_GVX_T	8.71	-15.14	7.02	176.99	-2.16	0.0300
fPOD	C_G_V_X	8.94	-14.76	7.03	175.57	-2.10	0.0300
fPND	C_GVX	8.94	-14.73	7.01	183.78	-2.10	0.0300
fPOD	CG_V_X	9.27	-14.20	7.03	177.64	-2.02	0.0400

In the following examples, I provide indications of possible false positives (type 1 error). I present two ways in which neither a single predictor nor a schema are identified as the dominant explanation for lexical processing despite significant findings. I then present a case in which all of the indices are wrong and it is required to look at their combined relationship in order to understand the likelihood of the result being real, or as Regina Nuzzo (2014) termed it, “the odds that a hypothesis is correct”.

First, it is important to understand what false positives are and how they are identified. A false positive is seen in a significant finding to a statistical model that is known to be false. How would someone know that the model’s output is false? Unfortunately, in current methods there is no exact mechanism other than a researcher’s familiarity with their tools and populations to know if a specific model is a false positive without comparing to other findings. This brings us to the false positive rate, or the proportion of instances a given result has been shown to positively or negatively account for a given hypothesis. The false positive rate is an important aspect of medical science. It tells us the probability that a given test is reliable based on knowing a priori that a given disease does or does not exist in their test subjects. Note that I referred to medical sciences, where a ground truth can be established and tested a multitude of times. Basic scientific practice does not have such ground truths when posing new questions and testing new hypotheses. I will show however that the current model selection output, which relies on knowledge of the specific tools available to psycholinguistics and more restrictively to auditory lexical processing, can aid identifying erroneous results in the lack of a ground truth.

Given the examples above of successfully identifying either a dominant schema or a dominant variable, I start with the scenario of identifying neither. In Table 4 I illustrate how a dAIC > 2 does not aid in identifying either schema or predictor. The reason the significant result to the PND.CGVX_T model should be doubted has to do with the structural aspects that are represented by the following models with dAIC < 2. In the given example, only 1 of the 5 models with dAIC < 2 share similarity with the top model (fPOD.CGVX_T). This lack of coherence is what indicates that the significant finding is not due to the structural aspects of PND.CGVX_T, but rather a shared correlation between variables with little in common.

Table 4. Example model selection output illustrating a likely false positive, in that neither schema nor variable are representative of processing

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
PND	CGVX_T	0	-25.37	6.84	177.73	-3.70	0.0001
Freq	non-tonal	0.52	-24.89	6.85	175.93	-3.60	0.0002
PNF	C_V_C	0.86	-24.42	6.80	195.51	-3.58	0.0003
fPOD	CGVX_T	1.02	-24.43	6.87	175.49	-3.55	0.0004
HD	CG_VX	1.43	-24.02	6.87	177.21	-3.50	0.0005
nfPND	C_G_V_X	1.74	-23.73	6.88	175.48	-3.45	0.0006
Freq	tonal	2.15	-23.33	6.89	174.90	-3.39	0.0007
POD	CGVX	3.60	-21.63	6.86	199.60	-3.15	0.0010
nfPOD	C_V_C	5.92	-19.00	6.92	188.22	-2.75	0.0020
PND	CG_VX_T	8.45	-15.55	7.02	175.58	-2.22	0.0030
HD	C_V_VX	8.68	-15.15	7.00	182.87	-2.16	0.0300
PND	C_GVX_T	8.71	-15.14	7.02	176.99	-2.16	0.0300
PNF	C_G_V_X	8.94	-14.76	7.03	175.57	-2.10	0.0300
fPOD	C_GVX	8.94	-14.73	7.01	183.78	-2.10	0.0300
fPOD	CG_V_X	9.27	-14.20	7.03	177.64	-2.02	0.0400

In Table 4 note that dAIC does not reach a value >2 for all of the 15 models present. Low dAIC between models indicates a lack of substantial difference between them. In Table 5 I illustrate another likely false positive wherein dAIC fails to distinguish either a predictor or schema. In this example, all 15 models are mathematically equal representations of the phenomena despite this being theoretically not possible. This example shows how a presence of similarity between the models' dAIC can lead to the conclusion that the models share an underlying relationship that has nothing to do with the structural aspects of the networks they represent.

Table 5. Example model selection output illustrating a likely false positive, in that neither schema nor variable are distinguished through dAIC

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
PND	CGVX_T	0	-25.37	6.84	177.73	-3.62	0.0002
PNF	CGVX_T	0.10	-24.89	6.85	175.93	-3.60	0.0002
nfPND	CGVX_T	0.12	-24.42	6.86	175.40	-3.59	0.0002
nfPOD	CGVX_T	0.12	-24.42	6.86	175.40	-3.59	0.0002
fPND	CGVX_T	0.19	-24.02	6.87	176.50	-3.58	0.0002
fPOD	CGVX	0.23	-23.73	6.88	175.48	-3.57	0.0002
POD	CGVX_T	0.25	-23.33	6.89	175.47	-3.57	0.0002
HD	C_GVX	0.30	-23.20	6.86	180.60	-3.55	0.0002
Freq	CGVX_T	0.45	-22.05	6.92	175.22	-3.50	0.0003
Freq	CGVX	0.45	-22.01	7.02	175.58	-3.50	0.0003
fPND	CGVX	1.34	-20.91	7.00	182.87	-3.40	0.0004
fPOD	CGVX	1.34	-20.87	7.02	182.99	-3.40	0.0004
Freq	C_GVX	1.45	-19.72	7.03	175.57	-3.37	0.0005
fPND	C_GVX	1.50	-18.95	7.01	177.78	-3.33	0.0005
fPOD	C_GVX	1.50	-18.95	7.03	177.64	-3.33	0.0005

With the aid of the previous two tables, I have discussed cases that provide simple and clear indices of why to suspect a false positive. I the final example, I cast doubt on the top model due to its level of significance. For this we must look at the model's t and p values. The p value is generated from the relationship between the predictor's degree of freedom (df) and its corresponding t value. The p value is what has been used in discussions of significance and is accordingly a hot topic in its relation to the current reproducibility crisis. In response to the inability to reproduce previous findings, researchers from multiple fields have proposed changes to how we analyze data. One such proposal would lower the threshold for significance, which now stands at $p < 0.05$ in the social sciences, to $p < 0.005$ (Benjamin et al., 2018). A particular point made by this group is that for exploratory research with low prior odds, an even lower threshold should be used. Prior odds, when ‘informative’, are generated from past experiments, and when ‘uninformative’, are generated from information about the variables being used within the specific nature of the hypotheses being tested. Within these terms, all research into the nature of the mental lexicon as a network is exploratory in nature. This is because even when a researcher utilizes a well-known task, such as lexical decision or word repetition, the hypotheses being tested are made possible through the use of novel variables or populations differing in language background.

As I have illustrated in Table 4 and 5, when using the currently proposed model selection procedure, a weak p value is not the only indicator of a false positive. Other indices include 1) the model’s cohesion (i.e., grouping together of like items) of either a single variable or a single schema, and 2) dAIC > 2 between groups of models representing different networks. What if both of these options fail to explain whether a model output is to be believed? In Table 6, only the top model is below the threshold of 0.05. Next, we see that the dAIC is somewhat deceptive in that it is > 2. Finally, we see a representative schema in the top 5 models. If we were to proceed from the previous examples, PND.CGVX_T should be accepted as the top model and p = 0.02 a meaningful significant result. The last index to consider is that of the t value and its directional effect. Of the 15 models presented in Table 6, 7 have a negative slope. This would not be alarming if the slope matched the schemas to which they are tied to. For example, there is no reason why PND and nfPND from the same schema would differ in polarity.

Table 6. Example model selection output illustrating a likely false positive, in which low t value is indicative of instability

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
PND	CGVX_T	0	-25.37	6.84	177.73	-2.01	0.02
PNF	CGVX_T	2.60	-24.89	6.85	175.93	1.5	0.07
nfPND	CGVX_T	3.15	-24.42	6.86	175.40	1.4	0.08
nfPOD	CGVX_T	3.29	-24.42	6.86	175.40	-1.37	0.09
fPND	CGVX_T	3.33	-24.02	6.87	176.50	-1.30	0.10
fPOD	CGVX	3.36	-23.73	6.88	175.48	1.2	0.12
POD	CGVX_T	3.53	-23.33	6.89	175.47	1.1	0.14
HD	C_GVX	3.88	-23.20	6.86	180.60	-0.99	0.16
Freq	CGVX_T	3.95	-22.05	6.92	175.22	0.91	0.18
Freq	CGVX	4.47	-22.01	7.02	175.58	-0.89	0.19
fPND	CGVX	4.51	-20.91	7.00	182.87	0.80	0.21
fPOD	CGVX	4.53	-20.87	7.02	182.99	-0.79	0.22
Freq	C_GVX	4.60	-19.72	7.03	175.57	0.72	0.24
fPND	C_GVX	4.81	-18.95	7.01	177.78	0.63	0.26
fPOD	C_GVX	4.99	-18.95	7.03	177.64	-0.57	0.28

Next, I discuss the decision process as to what goes in the random effects structure. I start with the question: What if multiple aspects of the stimuli or experimental setup significantly affect reaction times? The issue of random effects is important in evaluating model stability and the likelihood of a false positive. We know that the lack of including Subject and Item as random effects in past studies means the reported p values were much larger than should be expected. The introduction of a variable that significantly accounts for a portion of the variance into the random effects structure will lessen the t value of a given fixed effect. This follows with more complex random effects, as each random intercept introduces adjustments to the intercept, accounting for a portion of the variance. While the standard practice currently is to include both Subject and Item, there are some researchers that believe that the random effects structure could be more complex (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). In the data to follow, random slopes, such as for Trial, were not used because they proved to not be computable. Multiple random intercepts however were used, with Subject and Item as the minimum. Because the goal of each analysis was to rank multiple fitted models in order to evaluate network cohesion, I chose to first evaluate which of my predictors were of significance in single predictor models, prior to model selection. This was done for two reasons: 1) to identify which aspects of the stimuli (Duration, Tone, etc.) or experiment (Block, etc.) would need to go into the random effects structure and 2) identify which non-stimuli or non-experiment related predictors would later go into the final model as fixed effects (post-model selection). In this way, I avoided the unnecessary complexity of placing all stimuli predictors within the same random effects structure, by choosing only those I knew represented processing as seen in the reaction times. Also, because the random effects structure was decided on prior to model selection, each variable was therefore ranked according to the same adjustments to the intercept.

A last issue to discuss is that of data exclusion. Aside from excluding errors, which I did according to the practice within the literature, there is the question of excluding outliers. The exclusion of outliers is done to increase the normality of the distribution. However, current mixed models can also analyze data based on a number of distributions, including, log-normal, gamma, negative binomial, Poisson, heavy-tailed etc. Multiple methods are used for determining how much data must be excluded prior to analysis. Some take a standard approach and exclude at an arbitrary cut-off, usually 2.5, or 2 standard deviations (SDs) above the mean. Some prefer the older boxplot method, in which outlying trials are identified according to a graph. Others still transform their reaction times (e.g., Sadat et al., 2010). While there are likely reasons to use any number of combinations of the above-mentioned options, I will limit my discussion to those which were viable given the current model selection procedure. In each analysis in the experimental chapter to follow I ran the same model selection procedure three times in order to assess the effect of exclusion criteria on cohesion: outliers removed based on the boxplot, 2.5 SDs or 2 SDs above the mean. I found that model selection outputs were more stable with the boxplot method. Recent trends in experimental approaches have moved away from the simple boxplot, most likely because it leads to eliminating more than the ideal number of trials, especially with heavy-tailed distributions. Heavy-tailed distributions are after all the norm in psycholinguistics, not the exception. I found that 2.5 and 2 SDs would often give an entirely different result or an unclear output of mixed predictors and schemas. This is because the boxplot method cuts off the distribution's trailing trials. The results in comparing the three cut-offs reflected the fact that the tail is not the same as the mean. In an attempt to salvage a greater portion of trials I also fit models to the heavy-tailed distribution using a sampling method (Wood, Pya, & Säfken, 2016) from the 'mgcv' R package, which resamples the tail so as to adjust the mean. This often gave similar outputs as those generated from the boxplot method. However, the non-linear aspect of generalized additive models is not always what one looks for in an explanation of the tendency of the mean, while also being computationally expensive. In the experiments to follow I chose to use the simple boxplot. While a greater proportion of trials were excluded due to the boxplot method, it was a conservative approach to a computationally complex question. This is especially important in regards to the direction of the effect for the network variables, where model complexity and or transformations of the distributions might alter the raw output.

To conclude, I present several caveats. The applicability of the model selection procedure as it has been detailed here, is likely only relevant to lexical processing wherein we suspect network like behavior. However, it is altogether possible that it is applicable in evaluating other phenomena wherein a high degree of correlation exists between multiple dimensions of a system that would benefit from being graphed. The experiments in this dissertation are the only studies that I am aware of that have been evaluated in this manner. This implies several things. First, that there are likely scenarios that I have not foreseen, particularly in regards to the issue of identifying a false positive. A second implication is that the current method might be improved upon so as to decrease the dependence on cohesion (shared features) within the models. A final implication is that I have not yet tested when and why this method would fail, which is to say I do not know its limits. Ideally this method can help to inform on the reporting of results related to lexical statistics where a network like approach is already known to occur.

2.3 EXPERIMENT 1: Neighbor generation

Necessary for the purposes of the experiments in this study is the use of a syllable inventory that identifies phonological units in a manner consistent with the theory of minimal pairs. A minimal pair is a pair of words that differ by a single phoneme, the same distinction in identifying a phonological neighbor. Minimal pairs played an important role in early phonological studies (Pike, 1947; Swadesh, 1934), one reason for which was because of their ability to distinguish allophones from phonemes.

An example of the application of minimal pairs can be made in the comparison of two existing Mandarin syllable inventories. For the pinyin syllable bing1 (example character: 兵), Lin (2007) annotates the syllable as /pjəŋ1/, while Zhao & Li (2009) annotate the syllable as /piŋ1/. Both inventories annotate the syllable bin1 as /pin1/ (example character: 宾). With the Zhao and Li inventory (Z&L) we see a single edit distance between bing1 and bin1, however with the Lin inventory, the edit distance is 3, as can be seen in Figure 1. The larger edit distance reflects the greater phonetic attention given the Lin syllables. While there is no doubt some Mandarin speakers produce such exemplars, it is worth questioning whether greater phonetic attention to the syllable represents allophonic or phonemic distinctions.

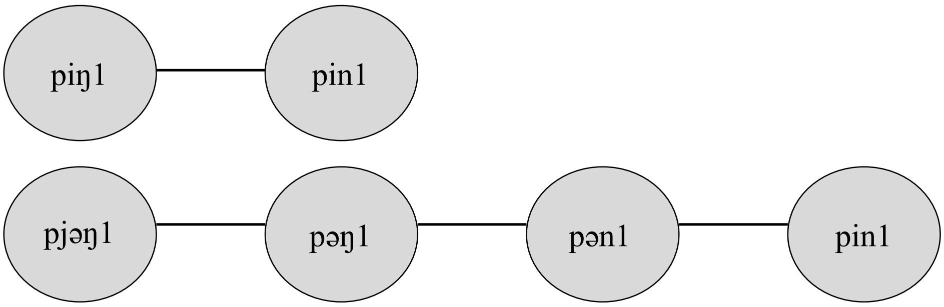

Figure 1. Edit distance for syllables bing1 and bin1 according to Zhao and Li (2009) (top) and Lin (2007) (bottom)

To identify what constitutes a minimal pair in Mandarin, I elicited minimal pairs to monosyllabic Mandarin words, in the neighbor generation task. The neighbor generation task has been used to investigate the weight of vowels and consonants on lexical selection (Cutler, Sebastián-Gallés, Soler-Vilageliu, & van Ooijen, 2000; Marks, Moates, Bond, & Stockmal, 2002; van Ooijen, 1996), the effect of phonological neighborhood density on English nonwords (Luce & Large, 2001), the effect of orthographic representations on phonological processing (Muneaux & Ziegler, 2004), and as a means to model misperceptions (Vitevitch et al., 2014; Vitevitch, Goldstein, & Johnson, 2016). With Mandarin speakers, Wiener and Turnbull (Wiener & Turnbull, 2015) asked participants to change specific segmental units of nonword stimuli in 4 experimental blocks (consonant, vowel, tone, and participant elected units) in order to change them into existing monosyllables. This allowed them to study the varying weights that differing units have on the retrieval of syllabic information. They found that participants' tone changes were the dominant choice in the participant elected block.

The purpose of this study's first experiment was to identify awareness of segmental units within the Mandarin syllable according to edit distance, edit type (whether the manipulation between one word and the next is made through the addition, deletion or substitution of a segmental unit), and edit location (i.e., the structural unit that a manipulated segment corresponds to in a fully segmented syllable: C_G_V_X_T). We opted to allow our participant to elect the unit of their choice throughout the experiment.

2.3.1 Methods

2.3.1.1 Participants

Thirty-four native-Mandarin speaking participants (Female: 21; Age: M, 24.74; SD, 5.29) took part in this experiment. None of the participants reported a history of speech or hearing disorders. Prior to the experiment, participants were asked to complete a short biographical survey. Contents of the survey included, besides age and sex, the name of their home province, self-rated spoken fluency on a scale of 1 (beginner) to 10 (native speaker) in English (M: 6.26; SD: 1.11), and other Chinese languages/dialects and/or other non-Chinese languages. From their home province, I classified the speakers into two groups based on whether the region was traditionally a Mandarin (Guanhua) speaking region (Guanhua: 21; Non-Guanhua: 13). To represent increased competition between similar Chinese languages/dialects, I summed the number of Chinese languages/dialects for all self-rated values from 3-10. This gave us a value that roughly reflects the number of Chinese languages/dialects (M: 2.14; SD: 0.56) that would have words similar to our target Mandarin stimuli. All participants reported native-level proficiency in Mandarin.

The Hong Kong Polytechnic University's Human Subjects Ethics Sub-committee (reference number: HSEARS20140908002) reviewed and approved the details pertinent to all experiments conducted in this study prior to beginning recruitment. The participants gave their informed consent and were compensated with 50HKD for their participation.

2.3.1.2 Stimuli

The material consisted of 157 Mandarin monosyllabic words, which can be seen in Appendix 1. The stimuli belonged to three groups according to the phase in which they were given to the participants: Example minimal pairs, 34; Practice, 10; Test, 113. A female speaker from the Beijing area produced all of the stimuli by speaking target monosyllables at a normal speaking rate 5 times into a high-quality microphone. Clearly produced items that were closest to the group mean were chosen. The pronunciation of each monosyllabic word was verified through transcriptions done by native-Mandarin speaking volunteers. Stimuli that did not have full agreement between transcribers were rerecorded and re-rated until all stimuli were verified by at least 10 volunteers. All stimuli were edited using Audacity 2.1.2 and were 415ms in length.

The example minimal pairs were exposed to the participants prior to the practice phase of the experiment. The idea of providing auditory examples of sound similarity came about during piloting. Upon given instructions to create minimal pairs or similar sounding syllables, our pilot participants were by and large unsure of what constituted similarity. Luce and Large (2001) avoided this possible pitfall by providing their participants the one-phoneme difference rule, while Wiener and Turnbull (2015) made it explicit which segment was to be manipulated in three of their four experimental blocks. I chose to provide example pairs because I did not want to bias our participants towards a tonal fully segmented syllable (C_G_V_X_T), however, it was not possible to provide a perfectly even example per each segmentation schema specifically because syllables can be interpreted in multiple ways depending on the number of units in the syllable, or the interpretation of the segments within the syllable. For example, the syllable pairs *bi3~bian3* can be interpreted as both C_GVX_T, and CG_VX_T with the Z&L inventory (/pi3/, /pian3/), while only representing C_GVX_T with the Lin inventory (Lin: /pi3/, /pjɛn3/). Of the 17 example pairs presented to our participants, 7 consisted of a single segment manipulation according to both inventories (chi3~zi3; ou1~sou1; miu4~you4; nie4~nue4; ka3~kua3; piao4~pao4; shan1~shan4), while 9 consisted of multiple segment manipulations according to either Lin or Z&L (fo2~fei2; ran2~rang2; zhe4~zhen4; huang2~hua2; tian2~tuan2; lie4~luo4; mian2~miao2; bi3~bian3; diu1~di1).

Test stimuli were created with the goal of representing all syllable structures in the Mandarin language. This was done by including all base rime syllables plus two more lexical items by adding a consonant, as can be seen in Table 1.

Table 7. Example test stimuli for Experiment 1

Rime	C + Rime	C + Rime
ai4	gai1	zai4
a1	ba3	ma1
yao4	tiao2	xiao3
ye2	bie2	jie1
wang4	kuang2	zhuang1
yue3	xue2	jue2

2.3.1.3 Procedure

Seated in a quiet room in front of a computer running E-Prime 2.0 (Psychological Software Tools, 2012), and wearing headphones equipped with an adjustable microphone, each participant was exposed to three phases: pre-training, practice, and test. Prior to beginning the experiment participants were instructed to not produce non-items, which included syllables that do not correspond to an existing Chinese character. For the pre-training phase, participants were told to listen and not respond as they were exposed to 17 word pairs as examples of similar sounding syllables. Each pair was presented according to the same procedure: an auditory stimulus was presented during a blank screen that lasted the word's duration followed by a slide that read "听起来像" (sounds like) for 500ms, that was then immediately followed by its minimal pair during a blank screen that lasted the duration of the stimulus. Between each pair, a dark grey slide that featured, "…", in the center of the screen remained for 2000ms.

For the practice phase participants were told to produce a similar sounding syllable for each of the 10 items. Each stimulus was presented on a blank screen with no time limit. Participants were told that their spoken responses would advance the next trial by activating the PST Serial Response Box. A pause of 500ms followed each participant's response, followed by a slide that read "下一个词" (next word) for 500ms before the next trial. The test phase followed the same procedure for all 113 randomized test items. The entire task took an average of 15 minutes to complete. The audio was recorded on a second computer using Audacity 2.0.6 for offline analysis.

2.3.2 Results

2.3.2.1 Edit information

Two native-Mandarin speaking volunteers transcribed the participants' spoken productions, with an agreement rate of 93%. A third transcriber resolved disagreements or classified unresolved items as nonitems. Due to the fact that our participants responded with large numbers of legal syllables that corresponded to existing Chinese characters, but were not monosyllabic words, the online dictionary www.zdic.net was used to identify whether or not a response fell into the non-item list. This was done because while many Chinese characters do not qualify as words, they do qualify as lexical items. Missing (67), identical responses (138), as well as nonitems (260) were removed, accounting for 12% of the total 3842 observations.

Edit distance was used to calculate similarity according to both the Lin and Z&L syllable inventories. Single-segment edits made up over 60% of the responses (Lin: 61%; Z&L: 65%) while two-segment edits comprised over 20% (Lin: 23%; Z&L: 24%), three-segment edits accounted for around 10% (Lin: 12%; Z&L: 9%), and four- and five-segment edits combined were roughly 3% of the responses for both inventories (Lin: 4%; Z&L: 2%).

The single-segment edits can be further described by addressing which segments within the fully segmental schema (C_G_V_X_T) were altered to make a minimal pair (edit location), and the edit type (addition, deletion, or substitution) that was made per manipulation. The predominant segment to be changed within single-edit manipulations was that of lexical tone, which accounted for around 54% of all manipulations (Lin: 56%; Z&L: 52%). The second most often manipulated segment was the initial consonant, accounting for roughly 29% (Lin: 31%; Z&L: 28%). The remaining segments featured in less than 18% of all manipulations: medial glide, 2%; monophthong 7% (Lin: 7%; Z&L: 8%), and the final X at around 7% (Lin: 5%; Z&L: 9%). As for edit type, the vast majority of manipulations were made through substitution (Lin: 90%; Z&L: 85%). Edits made from the addition of a segment accounted for around 10% (Lin: 8%; Z&L: 11%), while deletion type edits accounted for roughly 3% of the total (Lin: 2%; Z&L: 4%).

2.3.2.2 Syllable Inventory Test

A test was created to 1) identify an optimal syllable inventory based on our participants' minimal pair creations, and 2) construct an inventory that improves on both Lin and Z&L in accounting for our participants' phonological associations. A higher edit distance here signified that the inventory did not agree with the responses given by the participants, and conversely a low edit distance, particularly a score of 1, meant that the given inventory and our participants' judgments were aligned.

Results for all 113 syllables helped to distinguish which of the segments from Lin and Z&L fared worse according to a system of phonological similarity. For example, in both Lin and Z&L, pinyin syllables that have an 'ong' rime, such as hong2 (红), or cong2 (从), are annotated as /uŋ/. Participants preferred to produce phonological neighbors that contained /o/ rather than /u/. Once a sufficient number of segments were identified that rated poorly in both the Lin and Z&L inventories, a new inventory was constructed. An example of 10 syllables across the Lin, Z&L and newly constructed Neergaard and Huang (N&H) syllable inventories can be seen in Table 2. See Appendix 2 for the N&H phoneme inventory. An example of the syllable inventory test can be seen in Appendix 3 for the pinyin syllable 'yong4' according to the mean edit distance of the three syllable inventories.

Table 8. Comparisons between syllable inventories

Pinyin	N&H	Lin	Z&L
e	ə	ɤ	ɤ
ai	aɪ	ai	ai
ei	eɪ	ei	ei
o	oʊ	ou	o
ou	oʊ	ou	əu
ao	aʊ	ɑu	ɑu
ang	aŋ	ɑŋ	ɑŋ
yu	y	ɥy	y
yue	yɛ	ɥe	yɛ
yuan	yɛn	ɥɛn	yan

A final step in evaluating the three syllable inventories consisted of an ANOVA between edit distance scores: Lin (M: 1.59; SD: 0.86); N&H (M: 1.48; SD: 0.75); Z&L (M:1.48; SD: 0.74). The main effect was significant ($F=24.29$; $p<0.001$). Pair-wise comparisons showed that both the N&H ($p<0.001$) and Z&L ($p<0.001$) inventories outperformed the Lin inventory. No significant difference was found between the edit distance scores of N&H and Z&L.

2.3.3 Discussion

In this experiment, participant-elicited minimal pairs served in the creation of a novel syllable inventory as well as provide insight into awareness of the units within the Mandarin syllable. As it stands currently, the Lin inventory was outperformed by both Z&L and the newly created N&H inventories with no statistical difference between the latter two. In terms of segmentation, while results show that all units are subject to manipulation, there was a strong prevalence towards two principle units: the unsegmented syllable and lexical tone. The current roughly 55% of single-edit manipulations was only slightly less than what was found for tone changes in Wiener and Turnbull (2015), at 59.5%. These results perhaps do not in themselves lessen the status of each segment but emphasize a tonal route for mental search of minimal pairs. Of another note on this experiment's findings is the fact that our Mandarin-speaking participants produced a lower percentage of single edit responses (Lin: 61%; Z&L: 64%) than did the English speaking participants of Luce and Large (2001) at 71%, Vitevitch et al. (2014) at 74.5%, and Vitevitch et al. (2016) at 84.21%. It is likely safe to assume that the lower values for the Luce and Large study were the result of their participants having given spoken responses, whereas in both studies by Vitevitch and colleagues the recorded responses were written. Another reason for a lower percent of single edit manipulations might be due to the nature of our example pairs. Given our participants did produce examples of manipulations at all units, I decided in a second neighbor generation task to validate the performance of the three syllable inventories using an explicit single-edit example, as was done by Luce and Large. I expected this to increase the number of single-edit manipulations, and in turn aid in discriminating which of the three inventories best aligns with participants' manipulations.

2.4 EXPERIMENT 2: Neighbor generation

The current experiment repeated the neighbor generation task to 1) validate the performance of an optimal syllable inventory, and 2) analyze participants' reaction times according to lexical statistics built upon the optimal syllable inventory. There will thus be two consecutive analyses.

2.4.1 Methods

2.4.1.1 Participants

Thirty-four newly recruited native Mandarin speakers (Female: 23; Age: M, 23; SD, 4) participated in this experiment. None of the participants reported speech, hearing, or visual disorders. Participant characteristics did not differ from those from the first experiment in self-rated spoken English proficiency (M: 6.53; SD: 1.24), traditionally Mandarin speaking region (Guanhua: 23; Non-Guanhua: 11), or number of Chinese languages/dialects spoken (M: 2.38; SD: 0.74). One participant was removed from further consideration because they rated Mandarin as being their non-dominant language. All other participants reported native-level fluency in Mandarin.

As with Experiment 1, all participants gave their informed consent and were compensated with 50HKD for their participation as stipulated by the local ethics committee.

2.4.1.2 Stimuli

In order to increase statistical power additional stimuli were added to the set of 113 previously tested stimuli. This was done by adding two additional columns of consonant + rime syllables as per the examples in Table 3. This brought the total stimuli set to 200 test items and 10 practice items. Two items however were discounted for not existing in the www.zdic.net dictionary, reducing our total to 198 stimuli test items. Test stimuli can be seen in Appendix 4. All additional stimuli were created with the same voice and procedure. Of the original 113 stimuli, 27 were replaced so that lexical frequency could be accounted for using Subtlex-CH (Cai & Brysbaert, 2010). The first analysis will succeed in selecting a syllable inventory from which a database of lexical statistics can be built. The database will allow for the description of the stimuli for a second analysis of the participants' reaction times.

2.4.1.3 Procedure

The procedure differed from Experiment 1 in that no pre-training phase was given. In place of this, participants were given oral instructions as to what consisted of a similar sounding monosyllable through the use of the target syllable ling2 (e.g., 零), which they were told had the neighbors: ling4 (e.g., 另), ning4 (e.g., 宁), lang2 (e.g., 狼), and lin2 (e.g., 磷). All other procedural aspects were the same as in Experiment 1.

2.4.2 RESULTS

2.4.2.1 Edit Information and Syllable Inventory Test

The same transcription procedure as in the first neighbor generation task was followed. Missing (49), identical (209), nonitem (498), and semantically related responses (3) were excluded from the analysis. Edit distance was calculated according to the three syllable inventories, Lin, Z&L, and the newly formed N&H. Single-segment edits accounted for 67-73% (Lin: 67%; N&H: 73%; Z&L: 71%), while two-segment edits comprised 21% (Lin:

21%; N&H: 20%; Z&L: 21%), three-segment edits accounted for 6-10% (Lin: 10%; N&H: 6%; Z&L: 7%), and four- and five-segment edits combined were between 1-2.5%.

Edit location for the single-segment edits again was dominantly at the lexical tone position (Lin: 69%; N&H: 64%; Z&L: 66%). The second most common manipulation again occurred at the initial consonant (Lin: 21%; N&H: 19%; Z&L: 28%). The remaining syllable position saw a combined 5-16% instance of manipulation (Final X: Lin: 4%; N&H: 10%; Z&L: 7%, monophthong: Lin: 4%; N&H: 5%; Z&L: 6%, and medial glide: 1%).

Edit type again was dominantly substitution at 85-90% (Lin: 87%; N&H: 90%; Z&L: 85%). Both edit types, addition and deletion, accounted for a combined 6-10% (Addition: Lin: 4%; N&H: 7%; Z&L: 6%, and deletion: Lin: 1%; N&H: 3%; Z&L: 2%).

Using the syllable inventory test method of averaging edit distances between transcribed responses and their corresponding syllable inventory revealed statistically significant differences between all inventories: Lin (M: 1.48; SD: 0.77); N&H (M: 1.35; SD: 0.65); Z&L (M: 1.39; SD: 0.68). The Lin inventory showed statistically poorer alignment with participant judgments than both the N&H (F: 84.71; $p<0.000$) and Z&L (F: 40.02; $p<0.000$) inventories. Meanwhile the N&H inventory outperformed the Z&L inventory (F: 8.79; $p=0.003$). The fact that the N&H inventory outperformed both the Z&L and Lin inventories is not surprising seeing as neither were constructed specifically for the goal of identifying similarity between items.

2.4.2.2 Database of Neighborhood Statistics

With the N&H inventory validated, it was then used to create a database of neighborhood statistics from each of the 16 segmentation schemas shown in Table 1. The word list from Subtlex-CH (Cai & Brysbaert, 2010) was used for lexical frequency and subsequent neighborhood frequency counts. All PND calculations were taken from the top 30,000 most frequent phonological words. This led to slight differences in PND counts from the existing resource of similar structure and content (Neergaard, Xu, & Huang, 2016) that calculated similarity from the top 17,000 most frequent phonological words. Monosyllables that were featured in the stimuli but that were not present in the Subtlex-CH word list were added for the sake of calculating their PND, but were given a frequency count of 1 and thus were not part of the top 30,000 phonological words from which PND calculations were made. A further note on the database pertains to the non-tonal unsegmented schema (CGVX). Monosyllables in the CGVX schema have two types of neighbors: all other non-tonal monosyllables in the top 30,000 words, and all non-tonal disyllabic words that contain the same target phonological syllable. This meant that each of our 198 stimuli according to the CGVX schema had the same 397 monosyllabic neighbors yet varied in their disyllabic neighbors. Due to this lack of variation among the monosyllables, only the neighborhood statistics derived from the disyllabic content will be used in all further analyses.

I now move on to distributional aspects of the database. Of the top 30,000 phonological words, monosyllables account for 3.80% (n=1,141), disyllables 72.17% (n=21,652), trisyllables 14.84% (n=4453), quadrasyllables 8.73% (n=2618), and less than 1% for the remaining 5-, 6- and 7-syllable phonological words (n=136). Given the current availability of PND values, it is ideal to address the distributional aspects relevant to the non-uniqueness theory. As illustrated by Duanmu (2017), the approach taken in the annotation of phonology determines the size of the phoneme inventory: greater unitization leads to fewer phonemes, and vice versa, clustering of speech qualities increases the phoneme count. What then is the effect of segmentation on the number of phonological neighbors a given word might have? In Figure 2 I illustrate the distribution of mean PND, for monosyllables (left) and disyllables (right), according to the number of units in the schema that was used to build each phonological network. As is evident in both graphs, greater segmentation leads to fewer neighbors on average. The addition of lexical tone to a schema, i.e., the treatment of tone as a phonological unit, is also an indicator of fewer phonological neighbors. The networks built from 5 syllable units, C_G_V_C_T and C_G_V_X_T, are both tonal and have the lowest average PND. Conversely, the segmented syllables that have only two units, CG_VX and C_GVX, are both nontonal and have the highest average PND of the segmented schemas.

Of particular mention is the outlier behavior of the two unsegmented schemas. As can be seen in Figure 2, monosyllabic words have on average 286 neighbors in the tonal unsegmented schema (CGVX_T), yet around 25 for disyllabic words. For monosyllables, this means that every phonological word of a given tone assignment is a neighbor of every other monosyllable with that same tone, leading to 4 distinct subgraphs among monosyllables in the CGVX_T schema organized by tone. Because the edit distance rule only allows for a difference of a single unit between words, these monosyllables are neighbors with only other monosyllables. Figure 4b illustrates what this would look like for an individual monosyllabic word. The complete interconnectedness of neighbors means that these given words have clustering coefficients nearing 1. Disyllabic words of the CGVX_T network, on the other hand, do not divert greatly from the linear trend for mean PND of the segmented schemas. Like monosyllables, disyllables belonging to this network are only neighbors of other disyllables due to the single edit distance metric limiting the link to neighbors of greater edit distance.

Monosyllabic words in the nontonal unsegmented schema (CGVX) have on average 117 neighbors, yet 186 for disyllabic words. For monosyllables, this places nontonal syllables within a network of nontonal disyllabic words. Figure 3a illustrates this network aspect with the nontonal monosyllable *niang*. The number of neighbors almost doubles for nontonal disyllabic words due to the ability of nontonal disyllabic words to link to monosyllables, other disyllables, and trisyllables.

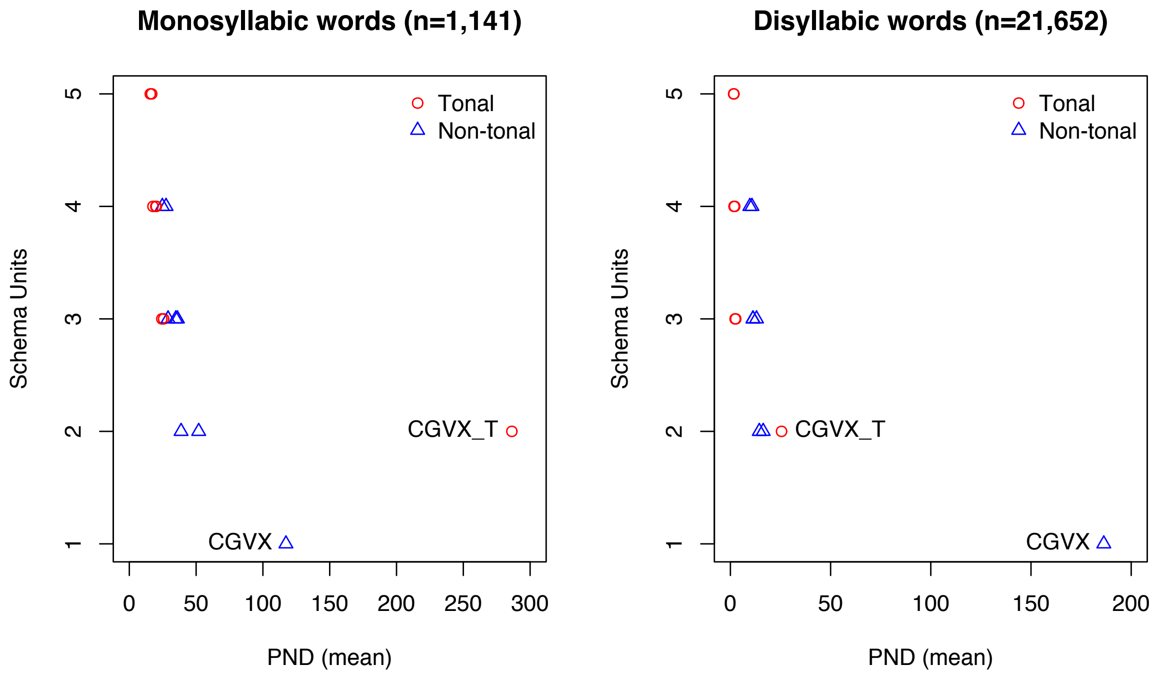

Figure 2. Distributional aspects of PND in comparison to the number of units within a segmentation schema for monosyllables (left) and disyllables (right)

The distributional aspects of PND in relation to segmentation is of particular interest when considering unsegmented syllables. What is obvious from these graphs is that the question of segmentation is gradient for networks built on segmented schemas. That is not the case with the unsegmented schemas. The unique distributional aspects of unsegmented syllables make evident that segmented and unsegmented syllables are not interchangeable. The graphs suggest that while networks of segmented syllables are all likely comparable with each other, networks of unsegmented syllables would require processing of a different nature.

2.4.2.3 Stimuli According to PND Database

The stimuli did not differ from the previous analysis. Lexical statistics for our stimuli were of two types: invariant, and variant. The invariant features did not differ across the 16 schemas, and were treated as nominal variables. They included segment length, which I will refer to as SegLen (SegLen 1: 5, SegLen 2: 48; SegLen 3: 100; SegLen 4: 45), lexical tone (tone 1: 49; tone 2: 48; tone 3: 45; tone 4: 56), and syllable onset, of which there were 28, and seen be seen in Appendix 4. Meanwhile, the variant lexical statistics, which vary per network schema can be seen in Appendix 5 according to their means and standard deviations.

2.4.2.4 Reaction Time Analysis

Reaction times were measured offline using SayWhen (Jansen & Watter, 2008). Three stimuli (*dia3*, *fo2*, *gun3*) were excluded for an error rate greater than a third of the number of participants. From the remaining 6,435 trials, I further excluded reaction times greater than 3000ms and lower than 415ms (17.79%). Thirty-one false starts, 390 nonitems, 187 identical items, 28 missing, and 1 semantically related were excluded (9.90%), giving us a mean of 1540ms (SD: 562ms).

Post exclusion, participants' responses consisted of edit distances between 1-5: edit1, 3582 observations (M: 1506ms; SD: 561ms); edit2, 943 observations (M: 1623ms; SD: 554ms); edit3, 246 observations (M: 1643ms; SD: 552ms); edit4, 49 observations (M: 1766ms; SD; 552ms); edit5, 6 observations (M: 1757ms; SD: 677).

The invariant predictors were evaluated to asses which would be placed into the random effects structure for the model selection procedure. As is customary in reaction times studies, Subject and Item were placed in the random effects to account for by-participant and by-stimulus variation. Trial as a crossed random effect did not reach convergence and was left out. Age, Sex, self-rated English, whether a speaker was from a traditionally Guanhua speaking region, and the number of Chinese languages/dialects spoken by our participants were all non-significant, as were segment length and lexical tone. Our participants' edit distance (Edit), i.e., the number of segments a spoken response differed from an auditory stimulus, significantly accounted for spoken latencies ($t = 4.76$; $p < 0.001$). This revealed that the shorter the edit distance, the faster our participants' reaction times. Because this is an attribute of the response and not the stimuli or our participants' demographics, Edit was not placed into the random effects structure for the model selection procedure. It will however be included in the final model.

Model selection consisted of the ranking of 144 models each with Subject, and Item as random effects, and each featuring a sole predictor [16 schemas * (Freq, HD, PND, PNF, fPND, nfPND, POD, fPOD, nfPOD)].

Table 9. Model selection output for Experiment 2, neighbor generation

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
PND	CGVX	0	-0.025	0.0069	178.74	-3.58	0.0004
nfPND	CGVX	0.90	-0.024	0.0069	175.36	-3.45	0.0007
fPND	CGVX	0.97	-0.024	0.0069	176.45	-3.43	0.0007
POD	CGVX	1.34	-0.023	0.0069	178.19	-3.38	0.0009
nfPOD	CGVX	2.00	-0.023	0.0069	174.78	-3.27	0.0013
fPOD	CGVX	2.03	-0.023	0.0069	175.83	-3.27	0.0013
Freq	Non-tonal*	3.51	-0.021	0.0069	200.48	-3.02	0.0029
Freq	Tonal	5.65	-0.018	0.0069	189.31	-2.63	0.0093
PNF	CGVX	6.90	-0.016	0.0069	188.26	-2.38	0.0186
fPOD	C_V_C	7.92	-0.015	0.0070	176.30	-2.14	0.0338
fPND	C_V_C	8.04	-0.015	0.0070	183.40	-2.11	0.0364
fPOD	C_GVX	8.29	-0.014	0.0070	177.70	-2.05	0.0419
fPOD	C_G_V_X	8.53	-0.014	0.0070	176.26	-1.99	0.0482
fPND	C_GVX	8.55	-0.014	0.0070	184.39	-1.98	0.0489
fPOD	C_G_V_C	8.69	-0.014	0.0070	175.92	-1.95	0.0532

* *Note: Two groups of nontonal schemas belonging to Freq were collapsed in this table. They differed according to dAIC by 0.01 and df by 0.02*

The results reveal a grouping of top ranked single predictor models that belong to the same schema, telling us that the structural aspects of the schema outweighed the importance of a single variable. The dAIC values for the variables belonging to the non-tonal unsegmented schema (CGVX) do not go far outside of a dAIC of 2. Meanwhile, dAIC values steadily increase in the remaining models, revealing that lexical frequency from non-tonal schemas affected the task to a greater proportion than did lexical frequency from tonal schemas, with a dAIC value between the two variables of 2.32. The remaining single predictor models reveal a prevalence for two variables (fPND and fPOD) over that of a specific schema. This suggests that only the CGVX schema presents sufficient evidence to account for processing during the task. While the top ranked CGVX model belonged to PND, it should be noted that all other CGVX predictors were similarly facilitative and within a small range of dAIC. We can infer from this behavior that each variable represents aspects of a single phonological system. We know that PND is at the core of this system because it is causally linked to all other CGVX variables present in the top 7 models.

The final model consists of the top ranked variable from the model selection procedure (PND.CGVX) plus edit distance (Edit), again with Subject and Item as random effects. Table 6 summarizes the final model. Words greater in PND facilitated mental search and production of phonological neighbors. Meanwhile, phonological neighbors were produced faster when stimuli and response differed by less units, i.e., lower in Edit. A generalized additive model was used to visualize a significant interaction between PND.CGVX and Edit (adjusted R-sq. = 0.425; F = 7.75; $p < 0.001$). Figure 1 illustrates, using tensor product smooths within a contour graph, how as edit distance increased, so did facilitation of higher PND stimuli.

Table 10. Model estimates for Experiment 2, neighbor generation

Random effects	Variance	SD			
Item	0.017	0.041			
Subject	0.145	0.380			
Residual	0.181	0.425			
Fixed effects	**Estimate**	**SE**	**df**	**t value**	**p value**
Intercept	1.533	0.068	35	22.56	< 0.0001
PND.CGVX	-0.023	0.007	179	-3.32	0.0011
Edit	0.048	0.010	4753	4.58	< 0.0001

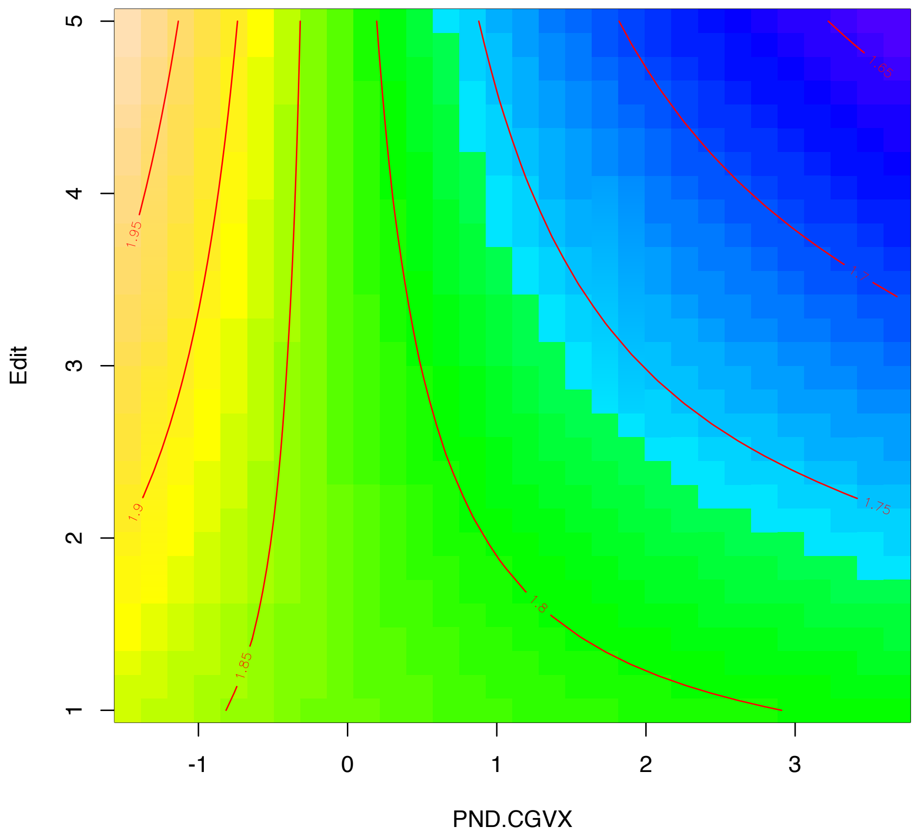

Figure 3. Tensor product smooth for the interaction between phonological neighborhood density according to the non-tonal unsegmented schema (PND.CGVX) and the number of units (segments and/or tone) that differed between auditory stimuli and participant-produced phonological neighbors (Edit)

2.4.3 Discussion

The second neighbor generation task identified an optimal syllable inventory while providing repeated evidence of segmentation biases, specifically towards the manipulation of lexical tone while maintaining a whole syllable. Changes made in comparison to Experiment 1 included changing instructions so as to provide a single edit example, and increasing the number of stimuli, which respectively increased the percentage of single-edit productions (Experiment 1: 61-65%; Experiment 2: 67-73%), and gave greater discriminative power in identifying the newly formed N&H inventory as the optimal syllable inventory. With the N&H inventory, neighborhood statistics were created based on 14 segmentation schemas that were in turn used in a reaction time analysis. Model selection revealed a facilitative effect to greater PND according to the nontonal unsegmented schema (CGVX), which was in line with our initial prediction based on previous findings (Wiener & Turnbull, 2015). I found a significant facilitative effect to lower edit distance values, matched with an interaction with PND. Participant-produced phonological neighbors that shared greater phonological similarity with the stimuli (i.e., lower edit distance) were produced faster when the stimuli had less phonological neighbors. These findings address a question posed by Vitevitch and colleagues (Vitevitch et al., 2016) as to whether the edit distance between the stimuli and participant produced phonological neighbors affects the time it takes to generate a phonological neighbor. In their study, they used neighbor generation as a means to investigate the types of neighbors that would occur to a given target if the target was incorrectly perceived. According to their hypotheses, our current result is suggestive that less time is needed to recover from the misperception of a spoken word when the misperceived item shares greater phonological similarity with its intended target.

Although my initial hypothesis correctly predicted an unsegmented schema whose greater PND values would facilitate production, I was neutral as to whether tone would play a mediating factor in the spreading of activation between neighbors. The question of why the tonal unsegmented schema was not the optimal schema is worth asking, especially given the dominance of single edit tone manipulations made by our participants. To aid in the discussion, in Figure 4 I visualize the differences between the toneless schema, CGVX, and its tonal counterpart, CGVX_T. The presence or absence of lexical tone, i.e., the difference between *niang* /niaŋ/, and *niang2* /niaŋ2/, determines the nature of what is a neighbor. Note that the non-tonal unsegmented network (4a) consists of the featured syllable in a small world network of disyllabic words with high inter-connectivity between neighbors, referred to as the network's clustering coefficient. The addition of tone to the network (4b) limits prospective neighbors to fellow monosyllables that either share the same lexical tone, or the same syllable, which in turn leads to an even higher clustering coefficient because all monosyllabic items that share the same lexical tone in the network are therefore neighbors of each other. For visualization purposes, the non-tonal word-level network was restricted to just 20 visible neighbors due to all CGVX_T PND values being well over 200.

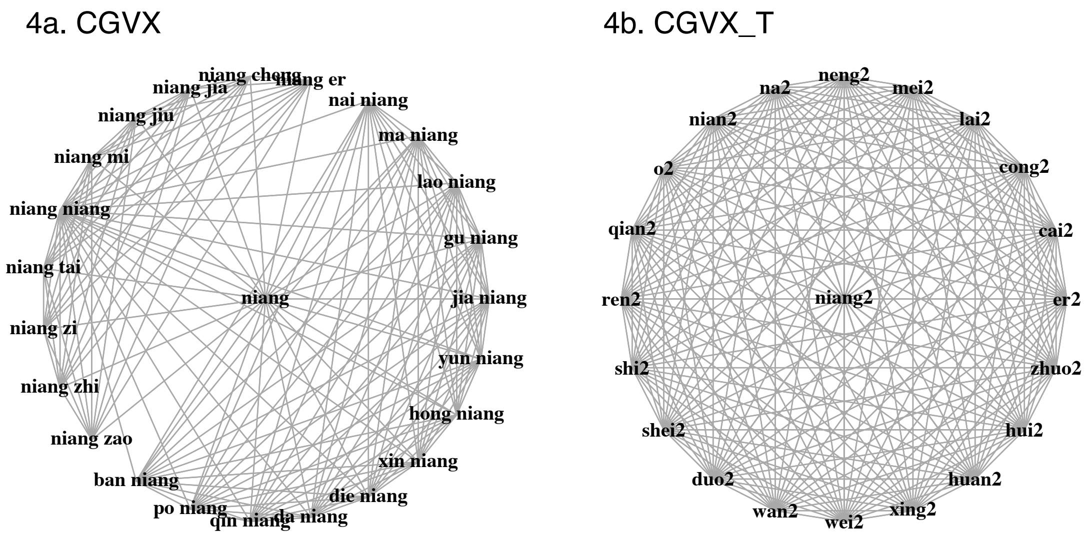

Figure 4. Word-level phonological networks for the monosyllabic word *niang2* /niaŋ2/ "mother" (娘) according to the non-tonal unsegmented schema (4a. CGVX), tonal unsegmented schema (4b. CGVX_T)

Through a comparison of the two schemas, we entertain the possibility that their respective characteristics hint at differences in mental search. First, the tonal schema suggests a search through the entire Mandarin phonological syllable inventory guided by lexical tone. Such a search fits our participants' productions and the goals of the task, however, it doesn't account for all of the mental steps needed in identifying a target monosyllable. While the CGVX_T schema is indicative of a phonological search through the syllable inventory guided by tone, the CGVX schema is indicative of a search through Chinese orthography. The first reason to suspect an orthographic interference lies in the constraints of the task itself: Participants were told to not produce non-items. In this experiment, non-items included syllables that do not correspond to an existing Chinese character. Thus, participants had to verify that a target syllable was in fact an item in their orthographic lexicon prior to producing a response. The second reason to suspect an orthographic influence is due to the fact that lexical tone is not represented with regularity in Chinese orthography. Thus, accessing visual representations of Mandarin syllables would not necessitate its use in limiting the field of prospective neighbors. The third reason for an orthographic influence lies in the nature of the CGVX schema, which consists of a target monosyllabic stimuli within a network of disyllabic nontonal words. It is within this predominantly disyllabic lexicon (Li et al., 2015; Wu et al., 2013) that monosyllabic items present in our participants' productions would likely be found during mental search, i.e., as part of disyllabic words.

2.5 EXPERIMENT 3: Auditory word repetition

In Experiment 2, reaction time analyses of the neighbor generation task provided evidence for the activation of non-tonal syllable-sized units. I now turn to a task that has been used to support claims of activation between phonological neighbors, specifically as a product of word recognition: the auditory word repetition task.

2.5.1 Methods

2.5.1.1 Participants

Thirty-six newly recruited native Mandarin speakers (Female: 22; Age: M, 25; SD, 4) participated in this experiment. None of the participants reported speech, hearing, or visual disorders. The completed survey showed that the current participants were similar to those recruited for Experiment 2, in self-reported English proficiency (M: 6.47; SD: 1.42), proportion from a traditionally Guanhua speaking province (Guanhua: 23; Non-Guanhua: 13); and in the number of other Chinese languages/dialects spoken (M: 2.28; SD: 0.82). In line with Experiment 1 and 2, all participants were compensated with 50HKD for their participation after giving their informed consent as per local ethics committee guidelines.

2.5.1.2 Stimuli

From the stimuli set used in Experiment 1 and 2 we chose 191 items that can be seen in Appendix 6. For the following analysis we excluded Our invariant lexical statistics consisted of the same nominal variables (segment length, lexical tone, and syllable onset). Variant lexical statistics can be seen in Appendix 7.

2.5.1.3 Procedure

Seated in a quiet room in front of a computer running E-Prime 2.0 (Psychological Software Tools, 2012), participants were instructed to repeat the words they heard over headphones into an attached microphone as fast as possible. The beginning of each trial began with the presentation of, “下个词” (next word), at the center of the screen for 1000ms. Next, a black screen was presented with the onset of the target audio. The end of the trial, and a pause of 1000ms, was activated when a participant spoke via a PST Serial Response Box. The entire experiment took less than 15 minutes. The audio was recorded on a second computer using Audacity 2.0.6. Participants were given a practice set of 10 words prior to beginning the experiment.

2.5.2 Results

Reaction times were again measured offline using SayWhen (Jansen & Watter, 2008). One participant was excluded for reaction times 2.5 standard deviations above the participant mean. Three stimuli (qia1, que1, sang1) were excluded due to errors greater than a third the number of participants. From the remaining responses, we excluded incorrect responses (3.22%) and reaction times greater than 975ms and lower than 425ms (5.9%).

None of the demographic variables, nor lexical tone, significantly accounted for a portion of the variance. SegLen significantly accounted for spoken latencies ($t = -4.07$; $p < 0.001$), revealing that the greater number of segments within the syllable, the faster our participants' reaction times. Because SegLen is an attribute of the stimuli we placed it into the random effects structure. Model selection again consisted of the ranking of 144 models each with Subject, and Item as random effects, and each featuring a sole predictor [16 schemas * (Freq, HD, PND, PNF, fPND, nfPND, POD, fPOD, nfPOD)].

Table 11. Model selection output for Experiment 3, auditory word repetition

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
nfPOD	C_V_C	0	0.007	0.002	91.69	4.02	0.0001
nfPOD	C_G_V_C	2.27	0.006	0.002	142.70	3.50	0.0006
nfPOD	C_G_V_X	2.80	0.006	0.002	142.40	3.41	0.0008
nfPOD	CG_V_X	3.46	0.005	0.002	166.30	3.20	0.0017
POD	C_V_C	3.82	0.005	0.002	119.10	3.28	0.0014
nfPND	C_V_C	4.35	0.005	0.002	123.40	3.15	0.0021
fPOD	C_V_C	4.35	0.005	0.002	118.50	3.19	0.0018
nfPOD	C_GVX	4.52	0.005	0.002	139.40	3.07	0.0026
POD	C_G_V_C	5.12	0.005	0.002	156.90	2.92	0.0041
PNF	CG_V_X	5.13	0.005	0.002	100.60	3.01	0.0033
HD	CG_V_X	5.24	0.005	0.002	142.90	2.91	0.0042
HD	CG_VX	5.24	0.005	0.002	142.90	2.91	0.0042
HD	C_G_V_C	5.31	0.005	0.002	144.90	2.89	0.0045
HD	C_G_V_X	5.31	0.005	0.002	144.90	2.89	0.0045
HD	C_G_VX	5.31	0.005	0.002	144.90	2.89	0.0045

The model selection procedure, as seen in Table 11, did not identify a single schema, but rather two influential predictors: PNF and POD. PNF occurred in 7 out of the top 10 models ether as a weighting or the original variable. The PND and HD combined variable, POD, occurred 8 out of 10 top models. Notably, lower down in the output, HD as an original variable was significant across multiple schemas. The effect of the combined variable, nfPOD, endured through the top 4 models without any coherence in segmentation schemas. The next best model to follow had a dAIC of 3.82. which is a good indicator that it and the best ranked model differed significantly despite the fact that they both belonged to the C_V_C schema. Of consistence throughout the model selection results is the complete lack of tonal schemas. Of last note is that all models, including the top ranked nfPOD, revealed an inhibitory effect to reaction times for words greater in similarity. Model estimates for the final model can be seen in Table 12.

Table 12. Model estimates for Experiment 3, auditory word repetition

Random effects	Variance	SD			
Item	0.00032	0.018			
Subject	0.00563	0.075			
SegLen	0.00001	0.003			
Residual	0.00462	0.068			
Fixed effects	Estimate	SE	df	t value	p value
Intercept	0.706	0.013	36.06	54.71	< 0.0001
nfPOD.C_V_C	0.007	0.002	91.69	4.024	0.0001

2.5.3 Discussion

The results from the auditory word repetition task provide a mix of expected outcomes, yet more interestingly a null finding in regards to segmentation schema if we are to consider the results of the model selection output alone.

I will first consider the model estimates from Table 12 independent of the model selection output. As predicted based on evidence from English speakers, greater phonological competition inhibited recognition of auditory stimuli with our Mandarin speaking participants. The second prediction, that HD would play an influential role is also evident in the model in that POD is the combination of both PND and HD. Contrary to the initial predictions is the identification of a model that is non-tonal. Based on data from implicit priming and speech errors, Chen et al. (2002) proposed that the phonemic status of lexical tone did not necessitate unitization. Prior research showing syllable sized speech errors in a tip-of-the-tongue paradigm (Chen, 2000), and tone errors in speech corpora (Chen, 1999; Shen, 1993; Wan, 1996), corroborated the view that lexical tone for Mandarin speakers mechanistically differed from that of segmental information. According to their account, tone, like stress in English and Dutch, is retrieved as prosodic information in parallel with segmental information. The current findings of a toneless segmented syllable would support this hypothesis, that is, if we were to assume the results of Table 12 stood alone.

How does the model selection output from Table 11 help in understanding the output of Table 12? In a traditional analysis, we would only have available to us the Table 11 output and would thus not have further insight into why or why not it is reasonable. Given that there are no a priori findings with the same population and PND values, I cannot with certainty state that the Table 12 findings are erroneous. This is the first experiment to suggest the identification of a variable over a schema with these specific conditions and with the model selection method. A conservative approach would be to state that a significant result has been identified that does not account for the structural aspects of lexical processing through a single segmentation schema. The results for Experiment 3 are likely not false, but instead a correlation with PNF and POD that exists across the multiple networks identified. For example, four networks revealed a nfPOD effect within a dAIC of 2.82, granting us sufficient evidence to state that nfPOD represents a significant aspect of how this specific group of stimuli were processed during auditory word repetition.

As with the prior neighbor generation task, I ask what the likely neighbors of a given word would be based on the demands of the task. In Figure 5 I illustrate features of the neighbors for the non-tonal monosyllable *niang* /niaŋ/, and its tonal counterpart *niang2* /niaŋ2/. As discussed before, the addition of tone, and therefore another unit to the syllable, restricts the number of neighbors available to a target word. While the tonal syllable *niang2* is neighbors with only other monosyllables, the toneless syllable *niang* features disyllabic neighbors such as *yan yang* /iɛn iaŋ/, *you yang* /ioʊ iaŋ/, and *yi yang* /i iaŋ/. While disyllabic neighbors, in Experiment 2, were likely neighbors due to the search aspect of the neighbor generation task, they are not in this auditory word repetition task for two reasons. First, word repetition involves retrieval of an item without explicit search and second there were no disyllabic words present in the stimuli set. Participants thus would have had no reason to expect upcoming disyllabic words given prior monosyllables and because words were not masked or presented in noise neighbors should more closely resemble targets.

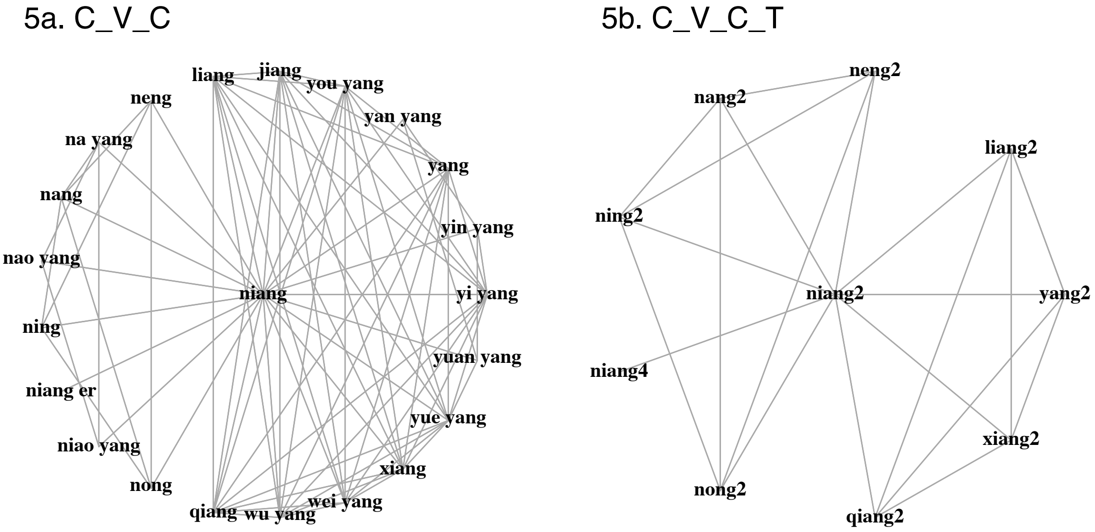

Figure 5. Word-level phonological networks for the monosyllabic word *niang2* /niaŋ2/ "mother" (娘) according to the non-tonal complex vowel segmented schema (5a. C_V_C), and tonal complex-vowel segmented schema (5b. C_V_C_T)

In the experiment to follow I again implement the auditory word repetition task so as to address possible reasons why the results from Table 12 are or are not reflective of a false positive. The first and most obvious question generated from Experiment 3 pertains to the significance of SegLen. Despite the fact that the stimuli were of equal durations, syllables with more segments were produced faster than those with less segments. The sampling of words from several segment lengths that were themselves sampled in unequal proportions might have influenced the likelihood of identifying a single schematic representation. Another concern lies with the inhibitory effect to greater similarity. If the result was a false positive, does that put into question whether or not the direction of the effect was also incorrect?

2.6 EXPERIMENT 4: Auditory word repetition

Experiment 5 intends to address several concerns left unresolved in Experiment 3 while also extending the knowledge of PND in Mandarin. Addressing the curious SegLen effect found in Experiment 3, the current stimuli consist of equal proportions of syllable and segment lengths. The hypothesis being entertained with this decision is that the regularity of the information being processed leads to regularity of activation among lexical items in the mental lexicon, thus allowing for the identification of an underlying segmentation schema. By controlling for SegLen, I am also able to ask the question of whether phonological density measures such as PND, PNF, and their composites can account for differences in syllable length, following those studies that have already tested disyllabic stimuli with PND (Baus et al., 2008; Cluff & Luce, 1990; Sadat et al., 2014; Vitevitch & Rodríguez, 2004; Vitevitch & Stamer, 2006, 2009; Vitevitch, Stamer, & Sereno, 2008). Cluff and Luce (1990), through the use of disyllabic words containing two morphemes (e.g., bedroom), surmised that individual syllables affected perceptual processes much like monosyllables. Vitevitch, Stamer, and Sereno (2008) tested disyllabic words ranging from 3 to 5 phonemes, to find that phoneme length did not differentially effect sparse words being identified more accurately that dense words, while in lexical decision with the same stimuli they reported an inhibitory effect to greater PND and greater accuracy for sparse words. The study of varying syllable and segment lengths is of particular importance in Mandarin. Considering the dominant proportion of words in Mandarin are disyllabic, a monosyllabic PND effect falls short of generalizability to the language processing system. Meanwhile, syllables are known to have a greater influence on speech planning in Mandarin (Mok, 2009), likely due to the fact that each syllable can be identified as belonging to a lexical tone. Thus, a conservative prediction would be that disyllabic Mandarin words would behave as two monosyllabic words, as implied by Cluff and Luce (1990). Vitevitch and colleagues suggested that similarity according to PND is the same across syllable and phoneme length, however, they did not feature phoneme length as a level in their statistical analysis, nor report on whether there was or wasn't an interaction. They also did not report on exclusion standards for testing a normal distribution, which is not positive given that tasks such as perceptual identification and lexical decision are known to produce heavy tailed (i.e., non-normal) distributions. Thus, for current statistical standards their borderline p value of 0.02 should be doubted.

The next issue I address is related to working memory as a possible modulator of the PND directional effect. The studies that have investigated neighborhood effects amongst adults utilize a large range of different sized stimuli sets. In studies that have implemented the auditory word repetition task the story is quite straightforward where large stimuli sets (Luce & Pisoni, 1998: 400 words; Vitevitch & Luce, 1998: 240 words) lead to an inhibitory PND effect. In lexical decision tasks, two experiments using large stimuli sets showed inhibitory PND effects (Luce & Pisoni, 1998: 610 words; Vitevitch et al., 2008: 112 words) while one with a small set of stimuli showed a facilitative PND effect (Vitevitch & Rodríguez, 2004: 80 words). The picture naming literature is where we see inhibitory results with large and small stimuli sets (e.g., Sadat, et al., 2014: 533 pictures; Vitevitch & Stamer, 2006: 48 pictures), facilitative results with small stimuli sets (Baus et al., 2008: 48 pictures; Marian et al., 2008: 57 pictures; Pérez, 2007; Vitevitch, 2002: 48 pictures; Vitevitch & Stamer, 2009: 48 pictures), and non-significant PND effects (Jescheniak & Levelt, 1994: 96 pictures repeated 3 times; Vitevitch et al., 2004: 44 pictures). The possibility that the inhibitory PND effect is a tendency of longer stimuli sets, is suggestive of a compounding of activation between words in working memory and those in long-term memory.

Experimental tasks that investigate working memory have shown that as memory load increases, so does reaction time (Cohen et al., 1997; Jha & McCarthy, 2000). As a participant names a word, that lexical item is temporarily stored in working memory. Behaviorally, it has been shown that when participants sustain attention on a task, memory decay does not happen at the same rate as compared to when they are given a pause or a distractor task that does not interfere with the domain or modality of the main task (Rae & Perfect, 2014). For instance, Baddeley (1986) found that phonological memory decayed within roughly 2 seconds. This does not differ greatly from the recall of orthographic letters after doing simple math problems (Brown, 1958; Peterson & Peterson, 1959). Tolan & Tehan (1999) extended that timeframe past 15 seconds based on results of interference between words read out loud.

Given that during PND related tasks the inter-stimulus pause tends to be between 500-1500ms, i.e., under the rate of decay known to exist for phonological information, it is feasible to predict that activations of multiple word representations overlap and contribute to participant performance.

The possibility that there might be a cumulative working memory effect has not been investigated. In an initial attempt to address this interaction in word recognition, Vitevitch & Luce (1999) referred to the adaptive resonance theory (ART: Grossberg, Boardman, & Cohen, 1997), in which a connectionist model allows for the interactive feedback between a short-term storage of phonetic items and a list categorization network of item sequences. Item selection in such a network depends on transient links between target and context representations in that prior to the encoding of a new list, the context representations are reset. This feature of the ART model can be seen in current computational models of word recognition and speech production, wherein each item is represented as the product of a single event, rather than as the result of cumulative activation due to the sequential exposure to hundreds of items. While computational models of production and perception do not currently account for a cumulative processing effect, the literature does address the interaction of PND and working memory, specifically in the serial recall task (Roodenrys et al., 2002). Roodenrys et al. (2002) found that their results, in which dense words were recalled better, omitted less, and were more likely to be recalled in the wrong order compared to sparse words, stood in contradiction to the lexical competition account tied to NAM. The facilitative effect of neighborhood density was later supported in another lab using similar paradigms (Oberauer, 2009).

The current experiment tests the possibility that the experimental design used in prior PND reaction time studies is in part responsible for the contradictory PND results; namely the length of the task as determined by the number of stimuli a participant is exposed to without sufficient time for decay in memory load. The hypothesis is that by not taxing working memory with sustained activation of lexical items, our analyses will show a facilitative effect to increased density. To test this, I intersperse three blocks of a simple distractor task, known not to interfere with lexical processing (Brown, 1958; Peterson & Peterson, 1959), within an auditory word repetition task consisting of a large stimulus set.

2.6.1 Methods

2.6.1.1 Participants

Forty-seven native-Mandarin speakers participated in this experiment (Female: 29; Ages 19-38, M: 24, SD: 4; Guanhua: 29; Non-Guanhua: 18). None of the participants reported speech, hearing, or visual disorders. The same survey given in the preceding experiments was completed by each participant, providing values (from 1-10) for both self-rated spoken English proficiency (M: 6.79; SD: 0.94), and Num_Chinese (M: 2.06; SD: 0.76).

2.6.1.2 Stimuli

The auditory stimuli for this experiment consisted of 154 Mandarin words (10 practice; 144 test). A female native-speaker of Mandarin from Fujian province produced all of the stimuli by speaking at a normal speaking rate into a high-quality microphone. Stimuli fell into 4 categories according to their syllable or segment length: 36 3-segment monosyllables with a CVN syllable structure (e.g., san1); 36 4-segment monosyllables with a CGVN syllable structure (e.g., bian3); 36 3-segment disyllables with a CV V syllable structure (e.g., da4 yi1); 36 4-segment disyllables with a CV CV syllable structure (e.g., li4 shi3). The 144 test stimuli were made from 20 syllable onsets, whose distributions were non-significant according to both syllable length (SyLen), and SegLen. Eleven stimuli sets were constructed, each where stimuli were pseudo-randomized such that there were no consecutive presentations of items with the same onset or lexical tone (first syllables for disyllabic words). The stimuli list can be seen in Appendix 8.

Because the current stimuli consist of monosyllables and disyllables of both 3 and 4 segments in length, it was not possible to control their durations along all 4 dimensions. Instead, stimuli were chosen in order to minimize durational differences between 3-Segment words (CV V, M: 609.25; SD: 11.59; CVN, M: 609.00; SD: 11.01) and between 4-Segment words (CVCV, M: 784.67; SD: 9.25; CGVN, M: 784.17; SD: 11.02). Stimuli did not differ within their respective segment length groups, but were significantly different across segment lengths (F=9433, p<0.0000). Thus, while a significant difference in reaction times is expected between 3- and 4-Segment words, the same cannot be said between monosyllable and disyllables belonging to their respective segment lengths, which is critical in identifying whether monosyllables and disyllables are processed in an equivalent manner.

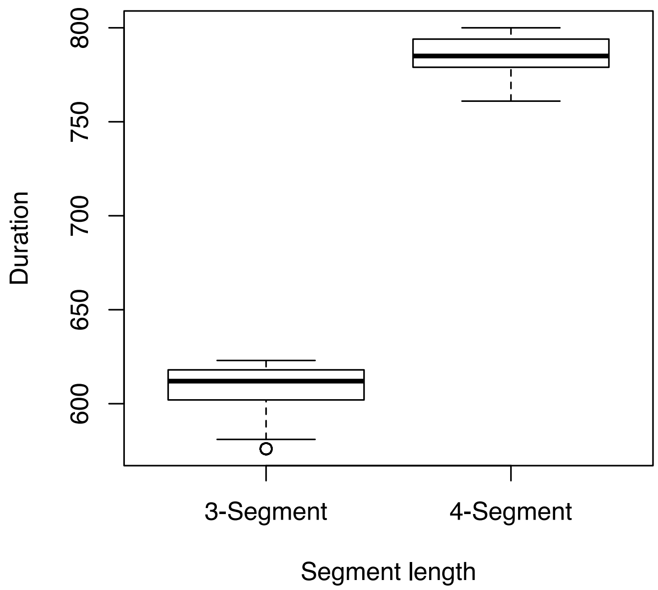

Figure 6. Stimuli durations according to segment length; 3-Segment (CV V, CVN) and 4-Segment (CV CV, CGVN)

Stimuli did not differ in Freq according to the tonal networks for both 3-segment words (CVN, M: 3.08, SD: 0.40; CV V, M: 2.75, SD: 0.46) and 4-segment words (CGVN, M: 2.98, SD: 0.49; CV CV, M: 3.33; SD: 0.21) according to both SegLen (p=0.869) and SyLen (p=0.981). Because Mandarin is a language with high homophony, it was not possible to choose a stimulus set that would both be controlled for equal contributions of lexical frequency, and phonological neighborhood density (PND) without including items with homophone neighbors. According to SegLen (3-segment, M: 1.38, SD: 0.59; 4-segment, M: 1.38, SD: 0.66), the stimuli do not differ (p=1); however, there is a significant difference (p=0.003) in the number of homophones between monosyllabic words (M: 1.53, SD: 0.67) and disyllabic words (M: 1.22, SD: 0.53). This is due to higher homophony among CGVN syllables. Lastly, stimuli were chosen based on the PND values generated from the tonal fully segmented schema C_G_V_X_T. This was done because this schema does not collapse vowel information. It is possible however that another schema will be identified in the model selection process. For this reason, the distributional aspects of PND and all relevant neighborhood statistics are presented in Appendix 9 according to their means and standard deviations.

2.6.1.2 Procedure

Participants sat in a quiet room in front of a computer running E-Prime 2.0 (Psychology Software Tools, 2012). They were instructed to repeat the words they heard over headphones into an attached microphone as fast as possible. Each trial began with a cross '+', in the center of the screen for 1000ms. Next, the onset of the target audio was presented concurrent with the exposure of a blank screen. A PST Serial Response Box was activated by the participants' voice, dependent on their response, which then led to a pause of 1000ms and the end of a trial. The experiment was partitioned into 4 blocks of 36 trials each with 3 interleaved distractor tasks. Stimuli were pseudo-randomized such that no two items were presented sequentially with the same onset or lexical tone. Each distractor task included 4 basic math questions: e.g., “20*2=__”. The distractor task was self-paced. Participants had to press a button to return to the following test block. The entire experiment took less than 15 minutes. The audio was recorded on a second computer using Audacity 2.0.6. Participants were given a practice set of 10 words prior to beginning the experiment.

2.6.2 Results and Discussion

Reaction times were measured offline using SayWhen (Jansen & Watter, 2008). Two participants were excluded from the analysis for reaction times 2.5 standard deviations above the group mean. No participants were excluded due to excessive error rates; however, 3 stimuli were removed for error rates higher than 25% (qing3, san4, sang1). From the new total of 6,345 trials, 142 stimuli were removed due to production errors, accounting for 2.24%, bringing the number of trials to 6,203. A further 122 trials were removed for excessive values below 611ms and above 1418ms, accounting for just 1.97%, leaving our final number of trials to be analyzed at 6,081 (M: 1008ms; SD: 146ms).

With Subject and Item in the random effects structure, preliminary examination of the demographic variables (Age, Sex, self-rated English, Guanhua, and Num_Chinese) showed that none significantly accounted for the participants’ reaction times. Given the possible effect due to the stimuli being displayed within a blocked format, a nominal variable representing the four blocks was tested in the random effects structure. It did not account for a portion of the variance and was thus left out of the final model. Significant effects were found for Duration ($t = 17.27$; $p < 0.001$), and SegLen ($t = 17.49$; $p < 0.001$). In contrast to the strange facilitation to longer segment length found in Experiment 3, 4-Segment words were responded to slower than 3-Segment words. Both Duration and SegLen were expected outcomes given how duration differentiates 3-Segment and 4-Segment stimuli, as illustrated in Figure 4. For this reason, Duration was added to the random effects structure for the model selection procedure, and each lexical statistic was represented in an interaction with SegLen to reflect its distinct levels. In Table 13 SegLen is annotated as SegLen(num) to reflect its numeric coding, which allowed for an easily comparable model selection output (i.e., a single line per model). For the final output, a nominal version of SegLen is used, as seen in Table 14, which accordingly displays both segment lengths.

Table 13. Model selection output for Experiment 4, auditory word repetition

Model interaction	dAIC	Estimate	SE	df	t value	p value
fPND.C_G_VX_T:SegLen(num)	0	-0.008	0.001	106.92	-6.67	< 0.0001
fPOD.C_G_VX_T:SegLen(num)	0.45	-0.008	0.001	107.29	-6.63	< 0.0001
PND.C_G_VX_T:SegLen(num)	4.95	-0.007	0.001	105.16	-6.19	< 0.0001
POD.C_G_VX_T:SegLen(num)	6.07	-0.007	0.001	105.46	-6.08	< 0.0001
fPND.C_G_V_C_T:SegLen(num)	7.34	-0.007	0.001	103.82	-5.94	< 0.0001
fPOD.C_G_V_C_T:SegLen(num)	7.45	-0.007	0.001	104.90	-5.93	< 0.0001
fPND.C_G_V_X_T:SegLen(num)	7.79	-0.007	0.001	103.92	-5.89	< 0.0001
fPOD.C_G_V_X_T:SegLen(num)	7.84	-0.007	0.001	104.90	-5.89	< 0.0001
nfPND.C_G_VX_T:SegLen(num)	7.87	-0.007	0.001	105.38	-5.89	< 0.0001
nfPOD.C_G_VX_T:SegLen(num)	9.46	-0.007	0.001	106.01	-5.73	< 0.0001

Model selection revealed that the tonal complex-rime segmented schema (C_G_VX_T) optimally represented processing in the task over all other schemas and single variables. The model representing the second-best schema had a large dAIC of 7.34. For the final model I included SyLen, in order to test whether monosyllabic and disyllabic words were processed differently. Table 14 illustrates two important points. The significant density variable displays a negative trend for both 3- and 4-Segment words, telling us that contrary to findings with English speakers, greater density speeds reaction times for Mandarin speakers. Meanwhile, critical to the role of the syllable as the planning unit, syllable length did not significantly affect reaction times.

Table 14. Model estimates for Experiment 4, auditory word repetition

Random effects	Variance	SD			
Item	0.001	0.035			
Duration	0.002	0.048			
Subject	0.009	0.094			
Residual	0.008	0.090			
Fixed effects	Estimate	SE	df	t value	p value
Intercept	1.017	0.017	80.91	61.15	< 0.0001
SyLen(disyllables)	-0.014	0.012	109.26	-1.12	0.2659
fPND.C_G_VX_T:3-Segment	-0.028	0.006	110.78	-4.55	< 0.0001
fPND.C_G_VX_T:4-Segment	-0.047	0.013	118.30	-3.75	0.0003

Due to the surprising nature of current facilitative findings, I sought to account for what in the current experiment led to such an outcome through a re-analysis. One concern is the apparent influence of Freq, a variable that is well known to facilitate language processing. While fPND was the highest-ranking model according to dAIC, it enjoyed only a small difference value of 0.45 with fPOD, but a somewhat large dAIC of 4.95 with PND of the same schema. This is evidence of there being both an inconsequential difference between the two frequency weighted variables, and a substantial effect of frequency on PND. To test their independent effects, Freq and PND of the C_G_VX_T network were run as original variables in another model. Interestingly, PND remained influential ($t = -4.70$; $p < .001$), while Freq was not ($t = -1.21$; $p = 0.23$). There were also no correlations between either the variables (0.05) or their residuals (-0.04). Taken together this reveals that the facilitative effect of fPND was not due to Freq, but instead the redistribution of values due to the weighting of PND irrespective of the item's lexical frequency.

Under the auspice that memory load accounts for neighborhood findings, a position entertained due to facilitative findings with small stimuli sets and inhibitory findings with large stimuli sets, we operationalized memory decay as the time spent on the distractor tasks (Decay). While each participant received the same basic math questions, they were given as much time as they saw fit to complete each task before returning to the repetition task. This means that as with participant reaction times, Decay varies per person and can thus be treated as a variable in the same regression manner as our featured neighborhood statistics. In order to maintain consistency, however, it was necessary to subset the trials belonging to the experiment's first block, such that each block under examination entails auditory lexical processing after having received a limited time for memory decay from a previous session of auditory lexical processing. In short, the following re-analysis consists of the addition of a variable we call Decay to the final 3 experimental blocks.

The values for Decay ranged from as short as 3 seconds (3100ms) to as long as 41 seconds (41373ms). Visual inspection of Decay's token values revealed that it was not linearly distributed. We rescaled the variable using a Box Cox transformation (Tukey, 1977) to evenly distribute duration length of non-lexical processing during the distractor task. Figure 5 displays the Decay variable pre- and post-transformation.

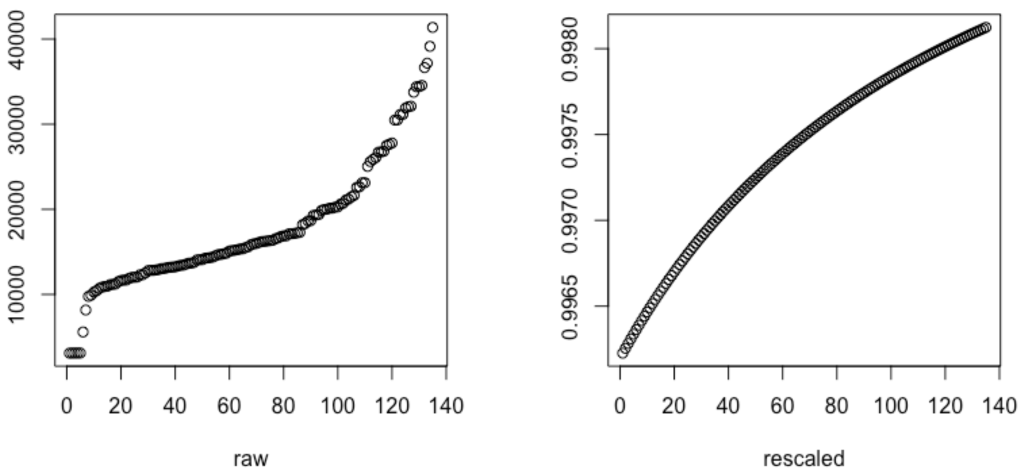

Figure 7. Raw token values for Decay (left) were rescaled using a Box Cox transformation (right)

The analysis of decay was run using a generalized additive model so as to visualize nonlinear relationships through tensor product smooths (Wood, Scheipl, & Faraway, 2013). Block was again tested within the random effects structure. Because it significantly accounted for a portion of the variance, it remained in the final model. As can be seen in Table 14, the effect of fPND on both 3-Segment and 4-Segment were significantly facilitative to reaction times. Novel to this model are the significant interactions between both fPND and Decay, Decay and SegLen, and Decay and SyLen.

Table 15. Model estimates for Experiment 4, auditory word repetition according to time spent on the distractor task (Decay)

Parametric coefficients	Estimate	SE	t value	p value
Intercept	0.999	0.017	59.79	< 0.0001
Smooth terms	**Edf**	**Ref.df**	**F value**	**p value**
ti(fPND.C_G_VX_T):Decay	1.00	1.00	10.45	0.0012
ti(Decay):3-Segment	0.75	0.75	30.90	< 0.0001
ti(Decay):4-Segment	0.75	0.75	30.22	< 0.0001
ti(Decay):Monosyllable	2.13	2.64	7.09	0.0002
ti(Decay):Disyllable	3.17	3.55	12.96	< 0.0001
Random effects	**Edf**	**Ref.df**	**F value**	**p value**
Subject	43.58	44.00	123.82	< 0.0001
Item	128.65	140.00	7.20	< 0.0001
Duration	0.95	1.00	22601.42	< 0.0001
Block	1.51	2.00	21.59	0.0028

I will first discuss the effects of Decay on SegLen and SyLen and then conclude with a discussion on the effect of Decay on phonological similarity. As can be seen in Figure 8 (left) there was a nonlinear effect of Decay on reaction times for both 3- and 4-segment word. The fact that this effect was so similar for each SegLen (3-Sement, $F = 30.90$; 4-Segment, $F = 30.22$) is likely due to the placement of Duration within the random effects structure, which mitigated the difference in reaction times between each SegLen group. A meaningful takeaway from the SegLen interaction with Decay is that greater Decay equated longer subsequent reaction times in a non-linear manner. More interesting is the effect of SyLen as seen in Figure 8 (right). This graph provides an important clue as to whether monosyllables and disyllables are being processed in the same way. The graph reveals that as with SegLen, greater Decay meant longer reaction times in the ensuing repetition block. Unlike SegLen, the effect was not the same between them. The cost for disyllabic words was greater as well as more non-linear than it was for monosyllables.

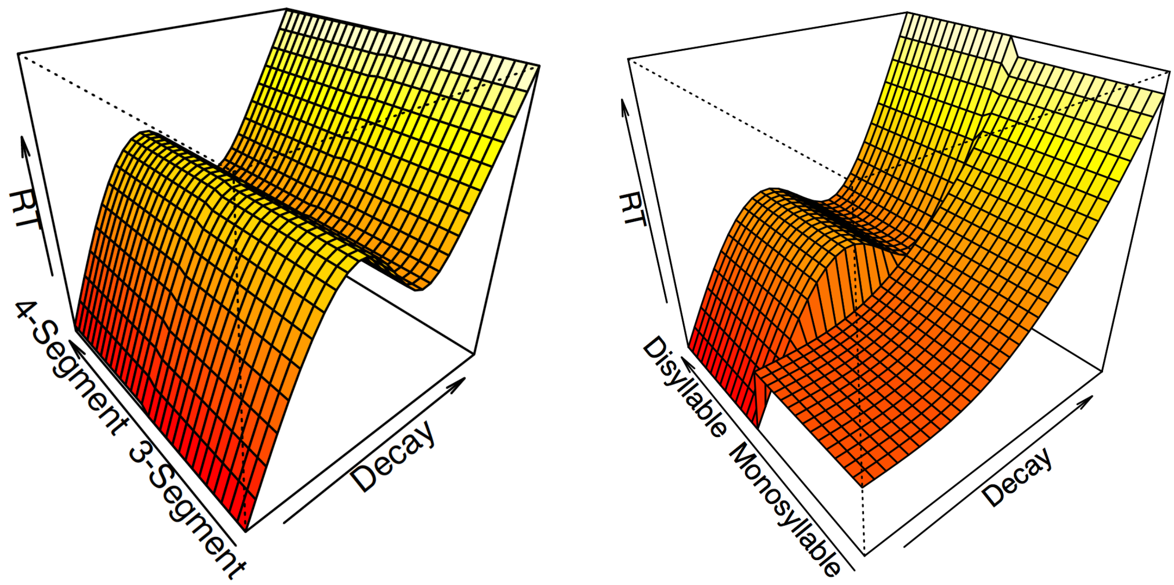

Figure 8. Interaction effects of time spent on the distractor task (Decay) with both 3- and 4-Segment words (SegLen: left), and Monosyllable and Disyllable words (SyLen: right)

The purpose of utilizing an interspersed distractor task within auditory word repetition was to test the hypothesis that the directional effect of PND was a result of cumulative activation between words in working memory and long-term memory. The premise was that the use of large stimulus sets leads to greater cumulative activation which in turn negatively impacts selection in what's known as lexical competition. As seen in Figure 9, when Decay was shortest, the effect of fPND was greatest. This is clear evidence that working memory is indeed a determining factor of neighborhood effects. However, does it necessitate a switch in directional effect due to greater competition? First I consider what the graph does show, and then follow with what it doesn't. The graph in Figure 9 reveals that by providing for Decay we get the expected diminishing of the principle effect tied to the attentional demand of the task (Rae & Perfect, 2014). Or in simpler terms, the PND effect fades because participants change their attention to something else that doesn't require PND-related processing. What the graph does not show is whether such a switch in PND polarity is even possible. Thus, while we know that facilitation can occur if given delay we don't know if inhibition would occur if we did not provide for delay.

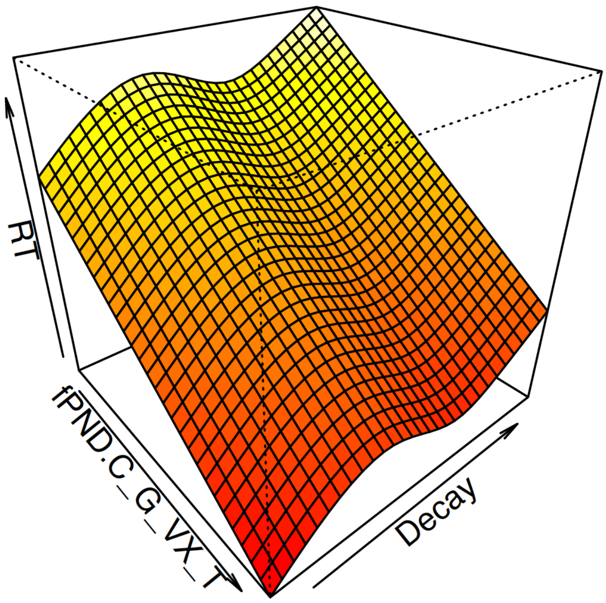

Figure 9. Interaction effect of time spent on the distractor task (Decay) and frequency weighted phonological neighborhood density (fPND) from the tonal complex/rime segmented network (C_G_VX_T)

2.7 EXPERIMENT 5: Auditory lexical decision

In Experiment 4 the addition of decay into the auditory word repetition gave us an inclination towards differential processing for monosyllable and disyllabic words in auditory processing. In Experiment 5, the role of lexicality takes center stage. Through the implementation of an auditory lexical decision task wherein PND values are ascribed to

Mandarin monosyllabic words, tone gap nonwords, and syllable gap nonwords, I identify access to similarity neighborhoods that differ according to the item's status.

2.7.1 Methods

2.7.1.1 Participants

Sixty-one participants took part in this experiment. None of the participants reported a history of speech or hearing disorders. Two participants were excluded due to reporting Mandarin as their non-dominant language. The remaining 59 participants were categorized according to their age (M: 23.95; SD: 3.90), sex (Female: 26; Male: 14), self-rated spoken English proficiency (M: 6.47; SD: 1.29), whether they grew up in a traditionally Mandarin (Guanhua) speaking region (Guanhua: 33; non-Guanhua: 26), and Num_Chinese (M: 2.08; SD: 0.81).

2.6.1.2 Stimuli

The stimuli used in the current experiment consisted of 75 monosyllabic Mandarin words that were presented either with 75 monosyllabic Mandarin tone gap nonwords (TG), or 75 monosyllabic Mandarin syllable gap nonwords (SG). All words and nonwords were 3 segments in length but did vary in syllable structure according to whether the rime ended in a nasal, /n, ŋ/, (words: 43; TG: 41; SG: 38), or vowel, /o, a, ɪ, ʊ, ɛ /, (words: 32; TG: 34; SG: 37). A one-way ANOVA showed that the distributions of onsets (/tsʰ, tʂʰ, f, tɕ, k, kʰ, l, m, n, p, pʰ, tɕʰ, r, s, ʂ, t, tʰ, u, x, ɕ, ts, tʂ, i/) between words and nonwords were not significantly different (F=0.13, p=0.72). A one-way ANOVA showed that the distributions of rimes (/ən, əŋ, aɪ, an, aŋ, aʊ, eɪ, ia, iɛ, in, iŋ, oŋ, oʊ, ua, un, uo, yɛ, yn/) between words and nonwords were also not significant (F= 0; p= 1). The word items can be seen in Appendix 10, while the nonwords can be seen in Appendix 11 for TG and Appendix 12 for SG.

Items varied in duration from 415 to 450ms (M: 430; SD: 9.82) and did not differ between words and nonwords (F=0.241, p=0.79). Due to the distributional aspects of lexical tone in the Mandarin lexicon, tone 2 was more prevalent amongst TG (tone 1: 19; tone 2: 29; tone 3: 16; tone 4: 11), but not so for SG (tone 1: 22; tone 2: 17; tone 3: 19; tone 4: 17), and words (tone 1: 18; tone 2: 11; tone 3: 26; tone 4: 20). Mean and standard deviation for PND, PNF and the remaining composite variables can be seen in Appendix 13 for our monosyllabic words, Appendix 14 tone gap nonwords and Appendix 15 syllable gap nonwords.

2.7.1.3 Procedure

Seated in a quiet room in front of a computer running E-Prime 2.0 (Psychology Software Tools, 2012), participants were instructed to judge whether auditorily presented stimuli were either real words (真词) or nonwords (非词) by pressing the left button or right button, respectively, on a PST response box. Seven lists of pseudo-randomized stimuli were created that did not have consecutive items with the same onset nor lexical tone. Each trial began with a cross '+' in the center of the screen for 500ms, followed by the onset of the stimuli.

No time limit was set to each trial's response time. Participants began the experiment with 10 practice trials. The experiment took approximately 10 minutes.

2.7.2 Results

Data cleaning was first done according to stimuli error rates and mean RTs. Two participants were excluded for having incorrect responses to over a third of the trials. A further two participants was excluded due for mean RTs 2.5 standard deviations above the group mean. This leaves a total of 55 participants, 26 given TG nonwords, and 29 given SG nonwords. Three stimuli were excluded for incorrect responses exceeding a third of the trials; one from TG nonwords, /soŋ2/, and two from SG nonwords, /pʰoŋ2, ʂɛn2/. Words from both SG and TG conditions will be analyze together based on distributional aspects of errors (TG words: 131 [6.71%]; SG words: 124 [5.70%]) and mean reaction times (TG words: 882ms; SG words: 938ms). TG and SG nonwords will be tested separately due to distributional differences in both error rate (TG nonwords: 383 [19.91%]; SG nonwords: 215 [10.16%]) and mean reaction times (TG nonwords: 1176ms; SG nonwords: 1086ms).

2.7.2.1 Analysis of word stimuli

From a total of 3,870 word stimulus trials, reaction times below 430ms and above 1268ms (accounting for 11.32% of trials) were excluded, leaving 3,432 trials (M: 810ms; SD: 173ms).

With Subject and Item in the random effects structure, only Num_Chinese (t = -2.26; p = 0.03), from the demographic variables, significantly accounted for a portion of the variance. Because Num_Chinese describes an attribute of the population rather than the stimuli or experimental setup, I will leave it for the final model. Word onsets (Onset) and rimes (Rime) however did account for a portion of the variance and were added to the random effects structure. The model selection output can be seen in Table 16.

Table 16. Model selection output for Experiment 5, word stimuli from auditory lexical decision

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
POD	CGVX_T	0	-0.023	0.006	55.76	-3.97	0.0002
PND	CGVX_T	0.29	-0.023	0.006	55.85	-3.92	0.0002
nfPOD	CGVX_T	1.91	-0.022	0.006	56.87	-3.66	0.0006
nfPND	CGVX_T	2.12	-0.021	0.006	56.99	-3.62	0.0006
fPOD	CGVX_T	4.90	-0.019	0.006	55.54	-3.17	0.0025
fPND	CGVX_T	4.92	-0.019	0.006	55.92	-3.16	0.0026
nfPND	C_V_C	9.27	0.016	0.007	35.78	2.24	0.0316
PND	C_GVX	10.22	0.014	0.007	58.61	1.99	0.0515
PND	C_V_C	10.60	0.014	0.007	49.67	1.95	0.0573
nfPND	C_GVX	10.89	0.013	0.007	53.80	1.73	0.0893

The top 6 models belonged to the tonal unsegmented syllable (CVVX_T) followed by the non-tonal complex-vowel segmented schema C_V_C. The difference between our top model and the next best schema is a large dAIC value of 9.27. This pattern is indicative of having succeeded in identifying an underlying segmentation schema above and beyond another schema or single variable. Meanwhile, the top two models had a dAIC of just 0.29, revealing no meaningful difference between them. The final model, including both facilitative effects to greater Num_Chinese and POD, can be seen in Table 17. No interaction effect between the two variables was found.

Table 17. Model estimates for Experiment 5, word stimuli from auditory lexical decision

Random effects	Variance	SD			
Item	0.001	0.035			
Duration	0.005	0.071			
Subject	0.001	0.033			
Residual	0.001	0.029			
Fixed effects	**Estimate**	**SE**	**df**	**t value**	**p value**
Intercept	0.875	0.029	66.71	29.91	< 0.0001
POD.CGVX_T	-0.023	0.006	55.76	-3.97	0.0002
Num_Chinese	-0.027	0.012	52.95	-2.26	0.0280

2.7.2.2 Analysis of nonword stimuli

From a total of 1,541 correct nonword trials, extreme values below 430ms and above 2000ms (accounting for 8.59%) were excluded, leaving 1,262 remaining trials (M: 940ms; SD: 227ms). None of the demographic variables significantly accounted for a portion of the variance, but Onset and Rime did and were placed in the random effects structure with Subject, and Item. Model selection for both nonword sets consist of the construction of 44 models, each featuring a sole predictor [16 schemas * (PNF, PND, nfPND)]. The top 10 significant individual predictor models are displayed in Table 18 ranked according to dAIC.

Table 18. Model selection output for Experiment 5, tone gap nonword stimuli from auditory lexical decision

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
nfPND	C_GVX_T	0	0.030	0.009	51.64	3.21	0.0023
PND	C_GVX_T	0.24	0.028	0.009	67.11	3.20	0.0021
nfPND	C_G_V_X_T	2.38	0.026	0.009	29.73	2.77	0.0096
nfPND	C_G_V_C_T	2.51	0.026	0.010	38.62	2.74	0.0092
nfPND	C_G_VX_T	2.53	0.026	0.010	41.72	2.74	0.0089
PNF	CGVX_T	2.62	0.024	0.009	77.59	2.82	0.0061
nfPND	CGVX_T	3.18	0.023	0.009	78.44	2.67	0.0093
PND	C_V_C_T	3.86	0.024	0.010	48.30	2.49	0.0161
PND	C_G_VX_T	3.99	0.024	0.010	38.24	2.44	0.0195
nfPND	C_V_C_T	4.09	0.024	0.010	39.11	2.40	0.0213

The top model belonged to the tonal onset/rime segmented syllable, C_GVX_T, followed by the tonal fully segmented schema, C_G_V_X_T. The difference between our top model and the next best schema is a modest dAIC value of 2.38. The top two models have a miniscule dAIC of 0.24, meaning they are not meaningfully different. The final model is further detailed in Table 19. As is evident from the negative t value, greater nfPND equated longer decision times.

Table 19. Model estimates for Experiment 5, tone gap nonword stimuli from auditory lexical decision

Random effects	Variance	SD			
Item	0.001	0.033			
Subject	0.012	0.110			
Onset	0.001	0.032			
Rime	0.001	0.034			
Residual	0.038	0.195			
Fixed effects	**Estimate**	**SE**	**df**	**t value**	**p value**
Intercept	0.967	0.025	34.68	38.40	0.0000
POD.CGVX_T	0.030	0.009	51.64	3.21	0.0023

From a total of 1,902 correct nonword trials, extreme values below 430ms and above 1540ms (accounting for 12.72%) were excluded, leaving 1,660 remaining trials (M: 913ms; SD: 229ms). Onset and Rime again were placed in the random effects structure with Subject, and Item, while none of the demographic variables significantly accounted for a portion of the variance. Table 20 features the top 10 models from the model selection output.

Table 20. Model selection output for Experiment 5, syllable gap nonword stimuli from auditory lexical decision

Variable	Schema	dAIC	Estimate	SE	df	t value	p value
PNF	C_V_C	0	-0.021	0.008	64.84	-2.49	0.02
PNF	C_G_V_X_T	2.85	0.015	0.008	59.77	1.82	0.07
PNF	C_G_V_C_T	3.27	0.014	0.008	60.73	1.69	0.10
PNF	C_GVX	3.30	-0.015	0.010	41.19	-1.57	0.12
nfPND	C_GVX	3.31	-0.015	0.009	16.23	-1.60	0.13
PND	C_G_VX_T	3.34	0.014	0.009	28.82	1.60	0.12
nfPND	C_V_C	3.53	-0.014	0.009	15.50	-1.53	0.15
nfPND	C_G_VX_T	3.88	0.013	0.009	31.82	1.41	0.17
PND	C_GVX	3.95	-0.013	0.010	14.61	-1.36	0.19
PND	C_V_C	4.47	-0.011	0.009	14.72	-1.17	0.26

With a dAIC of 2.85 between the next ranked model, PNF from the complex-vowel segmented schema (C_V_C) was the only significant model to account for a portion of the variance. Before declaring that model selection successfully identified an underlying segmentation schema for the processing of syllable gap nonwords, the output from Table 20 deserves further attention. There are three aspects of the model output that should raise alarm. The first of which is that despite the fact that a dAIC of 2.85 distinguished the top model (C_V_C) from the next in line (C_G_V_X_T), the top 5 predictors from the output reveal a

PNF tendency, either in the original variable (top 4 models) or as a weighting of PND (C_GVX). This is suggestive of an underlying variable rather than schema despite the dAIC. The second aspect of the output to consider is the lack of cohesion in in directionality. All three variables, PND, PNF, and nfPND, show positive and negative slopes. The final aspect worth concern is the borderline significance at $p = 0.02$, which, as I discussed in the Method section, is above the advised threshold of $p < 0.005$ (Benjamin et al., 2018). Taken together, these three aspects are highly suggestive of a false positive. Note that this discussion would be a labored one without the additional information provided by the model selection output. As is evident in Table 21, nothing is particularly suspicious when considering a model's output in isolation.

Table 21. Model estimates for Experiment 5, syllable gap nonword stimuli from auditory lexical decision

Random effects	Variance	SD			
Item	0.001	0.027			
Subject	0.014	0.120			
Onset	0.001	0.037			
Rime	0.001	0.032			
Residual	0.037	0.193			
Fixed effects	**Estimate**	**SE**	**df**	**t value**	**p value**
Intercept	0.938	0.026	39.30	36.56	< 0.0001
PNF.C_V_C	-0.021	0.008	64.84	-2.49	0.0152

2.7.3 Discussion

Experiment 5 investigated the effect of phonological similarity on lexical access with Mandarin words and nonwords in an auditory lexical decision task, in which similarity neighborhoods were defined by syllable segmentation. Monosyllabic words identified the tonal unsegmented schema (CGVX_T), such that greater density equated faster reaction times, with the top model showing an influence of homophones on PND (i.e., POD). Tone gap nonwords were inhibited by greater phonological similarity according to the tonal onset/rime schema (C_GVX_T). Syllable gap nonwords showed a weak inhibitory effect of PNF according to the C_V_C schema, followed by the prevalence of PNF in multiple models regardless of schema, which I interpret as indicative of a false positive. At first glance the results seem to extend upon the proximate unit account of speech production, yet in a traditional speech recognition paradigm. Prior to assuming the proximate unit proposal fits the current results, however, I question what exactly an auditory lexical decision task with Mandarin speakers tells us about lexical processing and neighborhood effects.

Auditory lexical decision has long been considered one of the go-to recognition tasks that unlike word-likeness judgments requires of a participant to make a yes/no decision rather than a scalar one. The task requires of a participant to identify from their mental lexicon a lexical item that matches to a phonological, orthographic and/or conceptual representation of lexical items, while attempting to isolate phonological processing. However, is it possible that due to language differences, those expectations are not met? Given the contradictory effect between Spanish (Sadat et al., 2014; Vitevitch & Rodríguez, 2004; Vitevitch & Stamer, 2006, 2009) and English speakers (Luce & Pisoni, 1998; Vitevitch, 2002; Vitevitch & Luce,

1998) it was quite logical to approach the differences in PND directionality to typological differences between the speakers' languages. Vitevitch and Stamer (2006) hypothesized that semantically related phonological neighbors, due to inflectional properties of Spanish nouns, influenced their participants' reaction times both in the picture naming task where production was inhibited by high PND (Vitevitch & Stamer, 2006), and in the auditory lexical decision task where recognition was facilitated by high PND (Vitevitch & Rodríguez, 2004). The hypothesis does not bode well with the current Mandarin data seeing as Mandarin has little to no inflection. Either there is an as yet unidentified structural attribute shared between Spanish and Mandarin, or the cause of differing directional effects lies outside of the characteristics of the languages being tested.

I will discuss two possibilities. Under the assumption that auditory lexical decision is a recognition task activating primarily phonological representations, the current facilitative word effect and inhibitory nonword effect can be interpreted as a product of working memory or executive control. Yet, under the premise that this is a mental search task activating primarily orthographic representations, then differences between the directionality of words and nonwords might reflect the push and pull of either matching or mismatching orthographic units within a phonological syllabary.

One possible cue, to the working memory/executive control hypothesis was provided with the significant Num_Chinese result to word stimuli. The purpose of this variable was to better categorize our participants' knowledge of the Chinese language family. Greater values reflect greater knowledge in usage of lexical items similar to auditory stimuli exposed to the participants. Thus, under the competition account, greater values should equate slower reaction times because a greater number of similar items would be in competition for selection. Under this account competition would also increase due to the fact that half of the items each participant was exposed to (tone gape nonwords) were highly similar items that might also be real lexical items in another Chinese language/dialect. Yet, as we have seen, greater knowledge of the Chinese language family led to faster reaction times as did items with greater phonological similarity. To explain this, it might be useful to turn to the bilingual literature on working memory.

While the literature on working memory and bilingualism has shown conflicting results over the years (Calvo, Ibáñez, & García, 2016) it is quite convincing when considering the effect of proficiency of language use, specifically when comparing performance on language and non-language tasks between interpreters and non-interpreter bilinguals. Interpreters have shown greater WM in tasks of categorization and lexical access (Bajo, Padilla, & Padilla, 2000), and reading, word and digit span tasks (Bajo et al., 2000; Christoffels, de Groot, & Kroll, 2006; Yudes, Macizo, & Bajo, 2011). While the simple Num_Chinese measure cannot serve as a stand-in test of proficiency of language use across multiple languages, it is suggestive that greater linguistic knowledge and use might also lead to more proficient processing and therefore shorter reaction times. But how does working memory play a role in our participant's performance? With the current study, we only see a Num_Chinese effect with word stimuli when participants had to decide between words and items highly similar to words (tone gap condition). The effect did not occur when the contrasting items were syllable gap nonwords (syllable gap condition). The ease of the second task compared to the first can be seen in their mean reaction times (tone gap nonwords: 1086ms; syllable gap nonwords: 938ms), wherein syllable gap nonwords were responded to with more ease due to their distance and lack of relative similarity with our word stimuli. Thus, it would appear as though greater knowledge of the Chinese language family was not needed in distinguishing syllable gap nonwords. That it was significant in the tone gap condition points to an executive function advantage in accessing metalinguistic knowledge. Thus, opposite to the competition account, having more items in long-term memory to contrast a given word against, aided our participants. Meanwhile, those participants likely to have a working memory advantage (high in Num_Chinese), outperformed their less lingual peers.

The second possibility to entertain for a facilitative word effect and inhibitory nonword effect deals with orthography within an auditory task. Does orthographic information influence an auditory lexical decision task for Mandarin speakers differently than for either English or Spanish speakers? Instead of retrieving an alphabetic item consisting of at most several letters out of a total of 26 (i.e., for English), educated adult Mandarin speakers search through an orthographic storage of knowledge with upwards of 10,000+ syllable-sized characters. Of these 10k+ lexical items, the correspondence between phonological and orthographic units, i.e., their transparency (Ziegler & Goswami, 2005) is exceedingly low. From the 1971 Xinhua Dictionary, only 3% of the existing characters could be used to reliably predict segment and tonal information (Y. Zhou, 2003). Meanwhile, all 10k+ characters correspond to only 1,300+ phonological syllables resulting in high homophony. Of the 1,300+ phonological syllables available, there are known monosyllabic words (arguable less than the 1,141 monosyllables identified in the database used for this dissertation), and there are syllables that only exist in multisyllabic words and/or obscure characters used in proper names, or pulled from ancient text. Real vs. nonword status depends on one or more Chinese characters corresponding to the given auditorily-presented syllable. It is thus under these constraints where a visual confirmation would aid in deciding on the lexicality of a probable Mandarin syllable.

With the likelihood that orthography is being consulted during the task, I now turn to what in the literature points to orthographic rather than phonological processing compared to the present results. In research on OND with European languages, facilitative effects to greater density have been reported in English (e.g., Andrews, 1989, 1992; Forster & Shen, 1996; Sears, Hino, & Lupker, 1995), Spanish (Carreiras et al., 1997), and French (Grainger & Jacobs, 1996). In the literature devoted to Mandarin speakers, two avenues have been taken: decomposition of component parts (radicals) at the character level (e.g., Li, Bi, Wei, & Chen, 2011; Li, Bi, & Zhang, 2010; Wang & Zhang, 2011), and repetition of whole characters in disyllabic words (e.g., Huang et al., 2006; Li, Gao, Chou, & Wu, 2017; Li, Lin, Chou, Yang, & Wu, 2015; Tsai, Lee, Lin, Tzeng, & Hung, 2006). Given the current findings show activation of whole syllables, the facilitative OND (Li et al., 2015) findings are suggestive of orthographic processing within the current auditorily presented stimuli. For an existing explanation on how OND and PND interact, I return to the Ziegler et al., (2003) study.

The facilitative OND effect in French (Grainger & Jacobs, 1996) was later used to investigate changes in directional effects between words and nonwords (Vitevitch & Luce, 1998, 1999). In alphabetic languages, phonological and orthographic grain sizes overlap. This implies that if speakers from an alphabetic language were using orthographic information in an auditory lexical decision task, the results according to a PND would be affected based on their correlation with OND. Ziegler et al. (2003) teased this overlap apart by featuring stimuli orthogonal in OND and PND to find an inhibition to greater PND yet facilitation to greater OND. The implication of their proposal is that the facilitative OND effect increases with greater overlap of OND and PND. They supported the theory that the inhibitory nonword effect was indicative of a PND effect because it matched the inhibitory effect found in their auditory word repetition task and the ones from English that came before (e.g., Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999).

The Ziegler et al. (2003) account matches the current findings to a certain extent. The existing facilitative OND finding of Li et al. (2015), is mirrored in our word stimuli. Meanwhile, the notion that this facilitative effect is linked to orthography is supported both by the knowledge of what the task entails for Mandarin speakers, but even more so, by the non-tonal network schema identified in the model selection procedure for our word stimuli. It does not match the fact that facilitation was also found in the auditory word repetition task, and not to an unsegmented schema, which would suggest facilitation due to Chinese orthography, but to the tonal fully segmented schema that is indicative of segmental and thus phonological information. The Ziegler et al. account also does not provide an explanation for the null finding with the syllable gap nonwords.

To explain the remaining results, I revert to previous research related to orthographic neighborhoods. One of the few articles to differences in directional effects between words and nonwords was in the OND study of Forster and Shen (1996). Forster and Shen relied on a model put forth by Balota and Chumbley (1984) in which, words high in criterion (i.e., items requiring mental search) would have a high error rate and slower reaction times, while conversely words low in criterion (i.e., items requiring little to no mental search) would have a low error rate and faster reaction times. Snodgrass and Mintzer (1993), reported similar directional effects as in the current experiment, i.e., facilitative for dense words, and inhibitory to dense nonwords. This rather simple model appears to measure mental effort operationalized as error rate and reaction time.

If we take the post-exclusion error rates for the three item types we find that words had a lower error rate (6.71%) and faster average reaction times (910ms) than syllable gap nonwords (10.16%; reaction times: 1086ms), which had a lower error rate than tone gap nonwords (19.91%; reaction times: 1176ms). According to the criterion definition, our word stimuli required less mental search than both tone gap and syllable gap nonwords, and syllable gap nonwords required less mental search than tone gap nonwords: Words < Syllable gap nonwords < Tone gap nonwords. This interpretation fits nicely with the relative strength of the predictions for the three word types, in that we see a likely null effect for syllable gap nonwords in between two opposing directional effects: words ($t = -3.97$; $p = 0.0002$), and tone gap nonwords ($t = -2.49$; $p = 0.023$). Under this interpretation, if yet another nonword class of items were used that further distanced itself from the Mandarin mental syllabary; for example, by using illegal combinations such as consonant clusters; we would see either a null effect or even a positive shift.

3. CONCLUSION

In this dissertation, I set out to explore the Mandarin mental lexicon through applying phonological networks to the question of co-activation of form-similar lexical items during speech. I sought to identify segmentation properties of the words activated in long term memory under the modeling assumption that a given target word within the metrical frame of the Levelt et al. model (Levelt et al., 1999), or the phonological representation frame of the Dell model (Dell, 1986), would share the same segmentation properties as the words they are connected to in long-term memory. While the ability to identify differences in network representations of shared activation was not a foregone conclusion, it was highly likely given the prior evidence from production and perception studies with Mandarin speakers suggesting the selection of unsegmented and segmented syllables respectively.

To address this modeling question through behavioral evidence, many steps were required. Products of this dissertation include, a novel Mandarin syllable inventory, a database of lexical statistics, and a novel statistical method for the evaluation of networked lexical statistics. These tools were used in 5 lexical processing experiments. Below, I first summarize the nature of each experiment and their findings in terms of the predictions made available by previous work in the field. Next, I synthesize the findings to address the issue of PND polarity and processing levels. I then interpret the findings through the lens of complex adaptive systems by introducing the concept of phonological segmentation neighborhoods as emergent structures within the mental lexicon. Finally, I discuss the dissertation's limitations and the avenues it has opened to future research.

3.1 SUMMARY

I began with the question of what constituted a phonological neighbor in Mandarin. In Experiments 1 and 2, participants created phonological associations in the neighbor generation task. In both experiments, participants showed a strong preference for creating minimal pairs by changing the syllable while maintaining the monosyllable's lexical tone. Although tonal manipulations were the preference, manipulations were made with all units within the syllable. In Experiment 1 I used the relationship between the auditory stimuli and the spoken response to test the annotation schemas of two existing syllable inventories. In what I titled the 'Syllable Inventory Test', I used average edit distance for phonologically transcribed word, i.e., the average number that two sets of strings (in phonological annotation) differ from one another, to asses which syllable inventory best aligned with the participants' minimal pairs. Lower edit distance values meant greater fidelity between the target and response. I then constructed a new syllable inventory to improve on either inventory's performance. It was significantly better than the competing inventories at accounting for participants' minimal pairs from Experiment 2.

With the newly formed syllable inventory, a database was constructed that contained PND values and their corresponding lexical statistics for 16 segmentation schemas (8 with tone and 8 without tone). Each schematic representation served to denote a unique phonological network, in that words (as nodes in the networks) were connected to each other (as edges in the networks) according to phonological edit distance determined by the information within the syllable. This meant that the same word would have different neighbors depending on the syllable's segmentation. For example, neighbors for *xia4*, /ɕia51/ from the tonal fully segmented schema (C_G_V_X_T), were all monosyllabic, tonal, and differed by only a single segment, which led to 11 neighbors in total that included: *xiang4* /ɕiaŋ51/ and *xiao4* /ɕiaʊ51/ (addition), *xi4* /ɕi51/and *ya4* /ia51/ (deletion), and *xie4* /ɕiɛ51/ and *xia1* /ɕia55/ (substitution). In contrast, neighbors of the same word from the nontonal complex vowel segmented schema (C_V_C), included nontonal words that were both monosyllabic and disyllabic, leading to 37 neighbors in total that included: *yu xia* /y ɕia/ and *xiu ya* /ɕioʊ ia/ (addition), ya /ia/ (deletion), and *tia* /tʰia/ and *xue* /ɕyɛ/ (substitution). As with this short example, I showed that greater segmentation led to fewer neighbors and that this relationship was gradient for all segmented networks. This was not however the case for the networks from the unsegmented networks (CGVX, and CGVX_T). The differences in average PND between segmented and unsegmented networks suggested that while segmented syllables were all likely comparable with each other, unsegmented syllables are suggestive of processing of a different nature.

Lexical statistics from the 16 networks allowed for the analysis of reaction times. Through the use of model selection, I ranked single predictor models, each containing 1 lexical statistic (as a fixed effect) belonging to each of the 16 networks. The grouping together of like predictors allowed me to assess the likelihood of identifying an underlying schema. I began with the reaction times from Experiment 2. The model selection procedure positively identified the nontonal unsegmented schema (CGVX), with PND as the dominant structural dimension of the network. As with previous research with this task and Mandarin speakers (Wiener & Turnbull, 2015), greater density facilitated mental search. I argued that the identification of the nontonal unsegmented schema was indicative of an orthographic influence on mental search based on aspects of the task, Chinese orthography, and the types of neighbors found in the CGVX network.

Experiment 3, which used the same stimuli from Experiments 1 and 2 consisted of an auditory word repetition task. This task was performed in order to test the core findings of the neighborhood activation model wherein auditory word repetition was also implemented. Based on their findings, an inhibitory effect to greater PND was expected. Meanwhile, a tonal and segmented schema was predicted based on the perceptual processes associated with the task, as well as an HD effect. The reaction times from this task were analyzed with the same model selection procedure. The top model revealed an inhibitory effect to greater neighborhood frequency weighted phono-orthographic density (nfPOD). While the top model provided a strong result ($t = 4.02$; $p = 0.0001$), it was not however indicative of an underlying segmentation schema. Multiple models in the model selection output belonged to nfPOD, meaning that the variable and not the schema was the dominant aspect of the task as represented in the lexical statistics. I argued that while the effect of nfPOD was not arbitrary, it was likely the results of an underlying correlation between networks. However, the fact that a variable rather than a schema was identified in the model selection output was possibly due to the variability in segment lengths of the stimuli. I hypothesized that with reduced variability in the stimuli I would increase the likelihood of identifying an underlying schema.

In Experiment 4, I revisited the auditory word repetition task. The first question I intended to address was that of the influence of segment length hypothesized to influence the results in Experiment 3. This was done by selecting equal proportions of 3- and 4-Segment words. The second question dealt with syllable length by featuring equal proportions of monosyllabic and disyllabic stimuli, orthogonal of segment length. Featuring disyllabic stimuli is critical for understanding lexical processing in Mandarin given 1) the importance of syllables in speech planning (Mok, 2009), and 2) the language is predominantly disyllabic in nature. The third question dealt with the possibility that an inhibitory PND effect from Experiment 3 might have been the result of cumulative working memory in that long stimuli sets in PND studies have shown a tendency for inhibition while short stimuli sets a tendency for facilitation. To address this, I included 3 short distractor tasks interspersed between 4 blocks of test stimuli in which participants completed simple mathematic equations, such as “2*2=?”. The distractor task provided time away from lexical processing so as to reduce the possible cumulative effect. Model selection revealed a dominant segmentation schema above individual predictors. The tonal complex rime segmented schema (C_G_VX_T) accounted for the spread of activation with the top model belonging to the composite variable, frequency weighted phonological neighborhood density (fPND). As the results pertained to the question of variability of segment length, it would seem that the stability of the output is indeed sensitive to the variability of the characteristics that participants sample from their mental lexicon. In regards to the differences between disyllabic and monosyllabic words in relation to the PND effect, syllable length was not significant, meaning that both monosyllabic and disyllabic words were processed in the same manner. Finally, in terms of the possible cumulative working memory effect, I operationalized the time spent on the distractor tasks, which varied per person and trial, as a measure of processing decay. Through a reanalysis, I showed that PND was indeed modulated by working memory in that the longer participants took on the distractor tasks, i.e., not processing lexical items, the more the effect of fPND diminished. While this view of decay adds to our understanding of the effect of PND, it unfortunately did not aid in answering the problem of PND directionality. Thus, as far as is evident in this experiment, Mandarin speakers, contrary to English speakers, are facilitated by greater density.

In Experiment 5, I asked if segmentation neighborhoods differ according to the item's lexical status. I used one group of monosyllabic words and two groups of monosyllabic nonwords specific to the Mandarin mental lexicon in two conditions of an auditory lexical decision task: tone gap nonwords, and syllable gap nonwords. Based on prior research, an inhibitory effect to words high in PND was predicted (Luce & Pisoni, 1998), while a possible flip in directional effect for the nonwords (Vitevitch & Luce, 1998, 1999). Due to the purported perceptual aspect of the task a tonal segmented schema was predicted to be the underlying segmentation. For the word stimuli, the model selection procedure identified a schema above all individual predictors. The tonal unsegmented schema (CGVX_T) accounted for the spread of information through the phonological network with phono-orthographic density (POD) as the top predictor. Contrary to the inhibitory English findings, Mandarin speakers again were facilitated by greater density. Alongside the network variable was also a significant and facilitative Num_Chinese finding, which suggests that greater knowledge of the Chinese language family aids in the judgment of lexical status. For the tone gap nonwords, model selection positively identified the tonal onset complex rime segmented schema (C_GVX_T), with neighborhood frequency weighted phonological neighborhood density (nfPND) as the top predictor. Mandarin speakers were again contrary to the English speakers in that greater density inhibited judgment of nonwords. Interestingly, the switch in directionality found with English speakers was also mirrored in the task, albeit in the exact opposite direction.

As for the final syllable gap nonwords, I argued for a false positive based on a borderline effect (p = 0.02) found for a single predictor that was more representative than a single schema.

The interpretation of Experiment 5 required the merging of a number of extant resources to come to a satisfactory account of the observed phenomena. What accounted for the facilitative effect of greater PND for words was the search through orthographic information to verify word status. This is supported by an existing facilitative OND result with Mandarin speaking participants in which neighbors were unsegmented syllables (Li et al., 2015). It was also supported in the unsegmented network schema identified in the model selection procedure. I argued that the descriptive model of Balota and Chumbley (1984) best accounted for the effects for the words and tone gap nonwords. Words high in criterion require mental search and are accompanied by high error rates and slower reaction times, while words low in criterion require little to no mental search and are accompanied by low error rates and shorter reaction times. The pattern seen with words and nonwords was such that words were responded to faster and more correctly and tone gap nonwords were responded to slower and more incorrectly, while syllable gap nonwords rested in between: Words < Syllable gap nonwords < Tone gap nonwords. The facilitative Num_Chinese effect aids in understanding how the concept of criterion can be operationalized as a measure of mental effort according to error and mean reaction times.

Faster reaction times for our participants who had more knowledge and use of the Chinese language family is supported by advantages in executive function and working memory tasks amongst speakers of multiple languages. The reason a multilingual advantage was revealed in this task was because of the difficulty in identifying the difference between words and tone gap nonwords where similarity was high. A native mandarin speaker would have difficulty with tone gap nonwords because the syllables exist and their tone assignments are probable, whereas syllable gap nonwords cross the restrictive boundary of what is and isn't a lexical item in Mandarin, making their identification easier – and thus their criterion lower.

3.2 SYNTHESIS

At the core of this dissertation has been the PND variable, both as the relational parameter that connects words according to sound similarity for the creation of phonological networks, and as the indicator of co-activation of words in long-term memory. In this section I seek to provide a synthesis of my findings in light of past research. I first discuss the contribution made to the study of lexical access by the model selection procedure, particularly in light of segmentation and lexicality. I then review the current findings in light of the production and perception literature with Mandarin speaking participants. I end with a discussion on PND polarity and the nexus of the PND effect.

The first contribution to the literature that the current research provides pertains to syllable segmentation and the role of lexicality. Model selection amongst multiple network statistics determined by their respective segmentation schemas allowed for a secondary source of information to be evaluated at the same time as the PND directional effect: the schematic representation of the network tied to the mental objects being processed. What this has brought to the table is the ability to judge simultaneous influences on auditory processing in a way that was not possible before. The current evidence shows that orthography was not only an influence to auditory processing but a defining structural aspect of the thought that took place. The neighbor generation task entails a participant starting from an auditory word, searching for one of the word's similar sounding neighbors, verification that the chosen item is indeed a word, and ending by producing that item's spoken form. Auditory lexical decision begins with an auditory item, includes a search through the lexicon to verify its lexical status, and ends in a judgment. While neither task has an explicit orthographic component, there is a suggestion of orthography in that items require verification of their lexical status. I have shown that the verification component of the tasks is not negligible, but a defining feature. The tasks revealed unsegmented schemas, CGVX and CGVX_T respectively, as the end result to repetitive sampling from the lexicon. These tasks stand in contrast to auditory word repetition, wherein the process starts with an auditory word that then leads to either a match or mismatch from the lexicon to end in its spoken production. As long as the participants are never given nonwords, no verification of lexicality is needed. This lack of orthographic consultation ended in a tonal complex rime segmented schema (C_G_VX_T).

How does the current evidence stand in relation to the production literature? O'Seaghdha and Chen (2010) proposed that whole phonological syllables are retrieved from the lexicon for Mandarin speakers during spoken production. In light of the current results that would suggest that orthographic consultation of some sort occurred during the tasks with their native Mandarin speakers. One task heavily used by the authors was form preparation. Form preparation, in its earliest version (Meyer, 1990, 1991) involved holding orthographic information in memory to later reproduce it. We should accordingly expect an orthographic effect from this task, as has already been shown (Damian & Bowers, 2003), and thus syllable-sized processing for Mandarin speakers. O'Seaghdha et al., (2010) were obviously aware of such a confounding issue and accordingly used non-alphanumeric symbols, such as #, @, &, etc. While the alpha-numeric symbols do not belong to the traditional Chinese orthographic system, they are known graphic objects used to represent a single word much like Chinese characters, and accordingly are not a sufficient distance away from the processing of Chinese characters. Another task used in this group of studies was masked priming, in which orthographic items prime, with average 50ms gap, other orthographic items. The glaring problem with these studies is that the authors explicitly used Chinese characters (Verdonschot et al., 2013; You et al., 2012), yet generalized to phonological processing. Picture naming has also been used to support the proximate unit principle, specifically because it does not explicitly require orthography. Yet, if orthography is used to create a link between a picture image and its corresponding phonological form, an orthographic effect should be expected as shown with adult participants learning novel object names (Rastle et al., 2011). Prior to the You et al. (2012) picture naming experiments, participants were familiarized with the pictures accompanied with the orthographic form, much like in the Rastle et al., study. Thus, despite the fact that pictures do not rely on orthography, humans do, especially when given the constraint of having to make the correct lexical association. In all of the above tasks the fact that syllable-sized effects were found is not surprising given the persistence of orthographic processing. Chen and Chen (2013) addressed these concerns by removing orthography from the experimental setup in the form preparation task. In their first experiment, they contrasted spoken versus written prompts while in their second experiment pictures were used without prompts. Both spoken prompts and pictures failed to show a significant onset priming effect. Considering no study has replicated these findings, and each condition only featured 18 participants per condition, researchers should treat these findings skeptically. Meanwhile, in their mixed effect models, variables were repeated multiple times as either single predictors or as part of interactions. Each time a predictor is repeated in the model colinearity is introduced, leaving the output questionable.

Because I have shown that neighbor generation has an unintended orthographic influence, and I did not include another speech production task such as picture naming, I cannot state that the syllable retrieval findings from the proximate unit principle studies aren't indeed an aspect of the Mandarin production system for phonological processing without the concurrent influence of orthography. The results from these studies, when taken together, do not make a convincing case for a lack of segmental processing, but a strong case for that of syllable processing. Future work will need to implement the current methods in picture naming and/or similar production paradigms, without orthographic prompts, in order to provide a schematic representation of the spreading of activation in speech production.

How does the current evidence stand in relation to the perception literature? The current findings are in line with the evidence amounted by Malins and colleagues (e.g., Malins et al., 2014; Malins & Joanisse, 2010, 2012), however, not necessarily with their predictions. Malins and Joanisse (2012) predicted that because PND treats units equally regardless of their place within the syllable, the metric cannot capture their findings of weighted segmental and tonal information. In the recent Mandarin implementation on the TRACE model, Shuai and Malins (2016) represent the Mandarin syllable in a 10-unit schema that features a tonal unit per each segment, excluding the onset. The exclusion of the onset follows findings from incremental gating tasks that do not find tonal discrimination until the vowel in syllables that begin as CV, the most recent of which places the beginning of discrimination at roughly

28ms into the vowel (Connell, Tremblay, & Zhang, 2016). In their model the pinyin syllable 'jia1' (e.g., 家) is annotated as, {j T_{silent} i T_{4L} a T_{4L} a T_{4L} a T_{4L}}, wherein the subscript '4L' annotates the high-level Mandarin first tone. Tones are represented for each unit, reflecting their contour: tone 1, 22222; tone 2, 44322; tone 3, 45554; and tone 4, 11234. As can be seen with the use of a tonal contour to represent the syllable, their goal is to account for online lexical selection. In their rationale, if information is accessed according to a linear unfolding of segments and tonal information, then whole syllable retrieval is not possible. The fact that the model results supported the eye-tracking findings in Malins and Joanisse (2010), likely lend credence to the inapplicability of the single-edit distance metric. Given the current findings, I would agree with their assessment of phonological density metrics not capturing segment by segment processing, despite previous attempts to do so (Magnuson, Dixon, Tanenhaus, & Aslin, 2007). I believe the current dissertation illustrates that PND is the degree of an interconnected network of words, not segments. That this activation spreads through a network in which similarity is defined through greater segmentation implies that online processing is indeed segmental according to the auditory word repetition task. In the introduction of phonological segmentation neighborhoods to follow, I argue that the network, according to a specific schema, is representative of a transient state due to the sampling of many words, not the unfolding of a single word segment by tonally-weighted segment.

The two persistent problems in accounting for lexical access with PND are 1) polarity of the directional effect, and 2) its nexus. The lack of reliability in the polarity of PND casts doubt on much of the core principles in the literature on lexical access. Rather than resolve this issue within existing modeling terms, the current dissertation's findings further complicate the accepted narrative dating back to the earliest studies seeking to account for interactivity between words (e.g., McClelland & Rumelhart, 1981). Meanwhile, the polarity problem naturally bleeds into the question of processing levels, seeing as polarity has been used as the primary evidence in arguments that place PND at the lexical level (Vitevitch & Luce, 1999; Ziegler et al., 2003). Using auditory lexical decision and auditory word repetition, Ziegler et al. (2003) proposed that facilitation was due to dense orthographic neighborhoods, while inhibition to dense phonological neighborhoods. They proposed that the PND effect arose from the lexical level, calling on the same proposal put forth by Vitevitch & Luce (1999). An interesting aspect to the Zeigler et al. proposal is that orthography is further facilitated when a correspondence occurs between the orthographic and phonological neighborhoods. This view of orthography matches the current understanding of the interaction between orthography and phonology, in that consistency between the two allows for greater access (e.g., Andrews, 1982; Baron & Strawson, 1976; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Stanovich & Bauer, 1978; Waters, Seidenberg, & Bruck, 1984). The correspondence between orthography and phonology might be what is needed to explain the facilitative effect found in this dissertation's neighbor generation task and auditory lexical decision task, where the phonological networks identified in the model selection outputs corresponded to orthographic-like syllables. It might likely be that the conflicting results in the previous auditory lexical decision tasks were due to either the facilitative nature of correspondence between orthographic and phonological neighborhoods (Vitevitch & Rodriguez, 2004) or the lack of consistency between orthographic and phonological neighborhoods leading to inhibitory effects (Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999). Following this line of thought would bring us to the facilitative effect in the picture naming task of Vitevitch (2002), in that like Rastle et al. (2011) and You et al., (2012), participants were given the written form of the word corresponding to each picture prior to the picture naming task, implying that participants were facilitated by the recall of the written form of the object with its corresponding picture.

The orthography/phonology correspondence explanation put forth by Ziegler et al. matched their findings because in both their auditory lexical decision and auditory word repetition tasks they found an inhibitory effect to greater PND. Opposite to their findings, I found facilitation to greater PND for words in both tasks. Critically, in the auditory word repetition task the identified schema was segmented, suggesting no orthographic influence. This contrasting of an orthographic effect in lexical decision and lack thereof in word repetition has been seen in several studies (Pattamadilok et al., 2007; Rastle et al., 2011;

Ventura et al., 2004; Ziegler et al., 2004). What we see from this pattern is that phonology can be aided by a correspondence with orthography, but that facilitation or inhibition to greater PND is not dependent on that correspondence alone.

Does an orthography/phonology correspondence explain the polarity problem? The answer appears to be yes, however, it is only fully evident when considering results from languages with differing writing systems. Tasks used in identifying orthographic influence have primarily relied on auditory lexical decision and rhyme judgment. In lexical decision, words critically differ on their level of orthographic and phonological consistency. In rhyme judgment tasks, participants judge whether a pair of words are minimally different, and accordingly feature items with sound-to-spelling consistency, such as “fall-wall”, sound-to-spelling inconsistency, such as “jazz-has”, or words that are not phonological neighbors. Among behavioral studies with speakers from alphabetic languages, an orthographic influence on auditory lexical processing is a known phenomenon (e.g., Pattamadilok, Morais, Ventura, & Kolinsky, 2007; Rastle, McCormick, Bayliss, & Davis, 2011; Ventura, Morais, Pattamadilok, & Kolinsky, 2004; Ziegler, Ferrand, & Montant, 2004). Mandarin speakers in contrast have in the past shown little influence of orthographic information in behavioral speaking tasks (de Gelder & Vroomen, 1992; Zhou & Marslen-Wilson, 1999). Evidence from ERP (Pattamadilok, Perre, & Ziegler, 2011; Perre, Midgley, & Ziegler, 2009; Perre, Pattamadilok, Montant, & Ziegler, 2009), fMRI (Booth et al., 2002) and transcranial magnetic stimulation (Pattamadilok et al., 2010), converge on the presence of orthographic processing during auditory tasks among alphabetic speakers. Spurred on by evidence with Mandarin speaking children and adults (Cao et al., 2015), which was suggestive of a differential pattern between alphabetic languages, Brennan, et al., (2013) analyzed rhyme judgments in children and adult speakers from both an alphabetic language (English) and a logo-graphic language (Mandarin). Results from the children revealed roughly the same pattern across the two languages, which is supported in similar research with only children (Krafnick et al., 2016). The divergence between children and adults is believed to be due to the reorganization of the brain due to increased literacy (e.g., Carreiras et al., 2009; Dehaene et al., 2010). Differences between language groups would then be due to the reorganization of the brain according to the training involved in learning to read language specific scripts. Young English speakers learn just 26 letters that have a relatively one-to-one correspondence with the language's phonology. This makes learning a deciphering process to identify which letters or clusters of letters visually correspond to phonemes. Young Mandarin speakers on the other hand, in order to deal with the lack of orthographic correspondence with phonology as well as the high homophony shared amongst monosyllables, are instructed to learn reading through writing (Lit et al., 2017). This accordingly reveals writing associated activation in the brain during reading (Cao & Perfetti, 2016; Li et al., 2017). For phonological processing this reorganization entails less recruitment of reading associated activity for Mandarin speakers during speaking tasks, but simultaneous activation of speaking and reading associated activity for English speakers, as seen in heightened activity in the superior temporal gyrus (Brennan et al., 2013).

In terms of PND, for English speakers, greater similarity between orthography and phonology should lead to facilitation due to the two streams of information not being in competition with one another. Conversely, when consistency is low, competition increases between processing areas, resulting in inhibition. Future research will need to address whether the inhibitory findings in auditory word repetition in both English and French were due to low orthography/phonology correspondence or yet another area of the brain distinct from Mandarin. Meanwhile, For Mandarin speakers, this lack of dual activation means that phonological information does not require the same resolution of both orthographic and phonological information. PND in this light suggests that facilitation is due to the lack of interactivity between brain processing areas, and thus, activation due to greater PND is always facilitative. This view of PND differs from NAM, in that NAM places competition between words rather than between neural processing centers. NAM, as with other descriptive and computational models, attempts to describe single-trial online lexical selection. In contrast to NAM, the current dissertation suggests that inhibition to greater PND is due to the combination of inconsistent information from more than one processing area while facilitation is the result of either the combination of more than one processing area when both streams of information correspond, or the result of phonological processing without the introduction of orthographic representations. Criterion then enters into this view of lexical processing when explicit search is required. Low criterion is the result of less mental search, accompanied by low error rates and faster reaction times. High criterion is then the result of greater mental search, accompanied by high error rates and slower reaction times. How mental search interacts with the PND effect should be a future focus, particularly in regards to auditory tasks that involve possible orthographic strategies, such as 1) neighbor generation and serial recall; and 2) manipulation or judgment of nonwords, wherein search includes both existing words and a negation of the target word's status based on a gap within the network.

The last concern is the nexus of the PND effect. One difficulty in identifying a place of a given effect within the processing route of a given task is that tasks are not interchangeable. The effect in one task may not allow for generalization to the accepted and discussed stages of processing, particularly the multi-stage speech production account of Indefrey (2011), with its 16 stages from picture to articulation. This is the case with the neighbor generation task, in which the PND values pertained to the stimuli, not the produced items. This was due to there being to much variability in the responses to perform meaningful statistics. The schematic representation is thus the end result of the stimuli in transfer to the spoken target, and unlike the contention put forth by Wiener and Turnbull (2015), is not indicative of production unless the produced items are the items under analysis.

The difficulty in generalizing processing stages between tasks is likely the reason the simplified three-tiered models of the past are so persistent. Under such constraints, the PND effect would necessitate post-phonological processing in that it is the post-combination of orthographic and phonological representations (auditory lexical decision), or the lack of orthographic influence (auditory word repetition). With similar reasoning, Pattamadilok et al. (2010) placed the congruency effect they found in an auditory lexical decision task with the use of transcranial magnetic stimulation as post-phonological, or phono-articulatory. This implies that the schematic representations we see for each task entail planning information for articulatory gestures. Passing tonal unsegmented syllables to articulatory planning would seem to support the strong syllable planning effects seen in Mandarin speech (Mok, 2009), and in the form preparation task (Chen & Chen, 2013). However, given that in lexical decision the process ends with a decision, rather than articulation, this seems unlikely. It is possible that information from the auditory lexical decision task is passed to the motor cortex in syllable-sized units only to be of no further use. It is also likely that the decision component entails further processing to decompose the syllable into segmental articulatory gestures. A future TMS task with Mandarin speakers, such as that performed by Pattamadilok et al. would answer this question. The auditory word repetition task is more representative of the three-tiered model in that an item is identified and then parsed for speech. From the schematic representation of network activation, I can infer that words are segmented prior to being passed to articulators, making the post-phonological argument a reasonable one. The subsequent syllable effect in speech would then likely be due to articulatory planning of lexical tone over the syllable rather than to the existence of phonological syllables.

3.3 PHONOLOGICAL SEGMENTATION NEIGHBORHOODS

In this section I introduce the term phonological segmentation neighborhoods (PSNs) to give name to the multiple network representations built from the differing segmentation schemas, and to an understanding of how more than one schema can account for processing for a single language. Through the model selection procedure, I identified 4 segmentation schemas as representative of the mental processes during the task: words from neighbor generation, CGVX; words from auditory lexical decision, CGVX_T; tone gap nonwords from auditory lexical decision, C_GVX_T; and words from auditory word repetition: C_G_VX_T. To place these findings into context I revisit the non-uniqueness theory (Chao, 1934) that held that the fact that multiple annotations of a language’s phonology are possible suggests that no single system is optimal. Duanmu (2017) extended this discussion to syllable segmentation, yet argued that a single optimal schema is attainable. In Duanmu’s argument, it is not clear if his proposed CG_V_X schema is meant to represent all types of processing, or if this schema would change if given constraints. The current findings identify a single schema per task or word type (tone gap nonwords), which tells us that an optimal schema is indeed likely, but given the specific demands of the task, and the specific lexical items used in the experiment. What this strongly suggests is that that the lexicon adapts to the demands placed upon it and more so that there is no 'best' state but transitions from states due to sampling from the lexicon. I argue that PSNs are emergent structures formed during repetitive sampling from the lexicon and that their strength depends upon executive functions specific to the person, and the regularity of sampling from the lexicon over time and in lexical content.

The claim that PSNs are emergent structures depends upon whether the current findings meet the conditions to be labelled a complex adaptive system, i.e., that the system entails interconnectivity, sensitivity to initial conditions, reorganization, and adaptation. The lexicon constructed in this dissertation used PND as the relational parameter defined according to each schema, i.e., what I am referring to as a PSN. In terms of the Mandarin phonological network, each PSN is a level of the same interconnected network. This implies that the full network includes all connections from each PSN. If we consider the full network as a resting state, i.e., prior to activation through the sampling of auditory items, then no PSN is more or less representative than another. Sensitivity to initial conditions can be seen in how the manipulation of the simple phonological edit distance leads to differing topological statistics as seen in the descriptive statistics of each network. Once activation is added to the network through sampling words from the lexicon, conditions are optimal for reorganization. As sampling occurs, activation spreads to neighbors according to the demands of the task. If orthography is not part of the thought process (auditory word repetition with words), then the PSN that emerges reflects only phonological information and thus is segmented and tonal. If orthography is part of the thought process (auditory lexical decision with words) then the end result reflects the end state of the integration of phonology and orthography. Thus, sampling directs the network towards a given schematic representation and either away from another representation or away from the resting state network. The next condition, adaptation, occurs due to a response to demands both internally and externally. Internal adaptation is due to the nature of integrating both orthographic and phonological information. External adaptation is due to the lexicon belonging to a biological system that has constraints, ones which are commonly referred to as executive functions: working memory, attentional control, inhibition control, etc. External adaptation, will depend on the features of the stimuli given to the participants and the conditions of their sampling from the lexicon.

What aspects of external adaptation explain why one PSN emerges from the resting state network rather than another? As has been shown from statistical learning paradigms, in which participants are given unsegmented strings of phonemes and later asked to identify segmented collections of these phonemes (in monosyllabic or disyllabic form), both infants (e.g., Saffran & Kirkham, 2018) and adults (e.g., Onnis, Monaghan, Richmond, & Chater, 2005; Saffran,

2003) implicitly extract regularities from language stimuli. In the formation of a PSN, it is likely that regularity in the characteristics of the stimuli plays an important role. If participants are given a small range of characteristics, the resulting model selection procedure will reflect this in a positive identification of a PSN. This can be inferred in the outcomes of Experiments 3 and 4, in that contrary to Experiment 3, Experiment 4 was regular in its proportions of segment length and syllable length. When sampling from the lexicon is restricted to characteristics shared between the words, then a subgraph will reflect that stability, and the end state of the mental task. This implies in dynamic terms that first the mental lexicon is at resting state with no schematic representation. As words are sampled, a transient yet stable subgraph emerges between working memory and the activated words and neighbors in long term memory. If sampling from the lexicon were to stop, the trace tied to that structure would fade, as seen in the decay effect of Experiment 4. The PSN that arises due to sampling is emergent because the lexical items themselves do not dictate the end schematic representation. This can be seen in comparing Experiments 2 and 3, in which the same stimuli were used with differing constraints on external adaptation (i.e., task demand on memory and search). While the lexical items do not dictate the end schematic structure, their variability likely does, as well as which characteristics they emphasize or de-emphasize. Like ants building an ant hill, each word contributes to the structural representation without an explicit plan. The researcher can likely influence the PSN, much in the same way that a person with a home ant colony kit can affect its construction by adding immovable structures that must be worked around, or choosing the specific soil for the ants to construct the colony within. This sort of flexibility can be conjectured upon with the current findings. In the auditory word repetition task of Experiment 4, the tonal complex rime segmented schema (C_G_VX_T) would likely be positively identified as the tonal fully segmented schema (C_G_V_X_T) if participants were given a greater proportion of 4-segment syllables. This result would suggest that a nudge in the end outcome is possible because the two schemas are highly similar to one another. The limit to that flexibility, and contrary to all predictions would be the identification of an unsegmented schema (CGVX or CGVX_T) in auditory word repetition task with words. Future work will need to identify how much flexibility is expected in a given task and whether there are individual differences in model selection output. Another future trajectory would be to explore the transience of a PSN emergence, specifically in the transition from one PSN to another PSN.

What are the implications of PSNs to natural speech? While statistical regularities are implicitly learned, and thus lead to the emergence of a PSN, they do not reflect sampling from conversation or other natural language processing where variability is great. This variability is likely exploited during processing (Onnis, Monaghan, Christiansen, & Chater, 2004), implying that a given person would not form a PSN or at least not to the extent shown here. Are there then uses for PSNs? PSNs open the route to addressing further questions into the study of learning, specifically in terms of structural aspects of the input (e.g., Goldstein et al., 2010). Being able to identify a person's behavior in sampling from their environment could help in distinguishing successful learners from unsuccessful learners, as recent research in statistical learning suggests (Onnis, Frank, Yun, & Lou-Magnuson, 2016; Onnis, Lou-Magnuson, Yun, & Thiessen, 2015). PSNs also might be applicable to the study of speech disorders where sensitivity to regularities have shown diagnostic applicability in the study of specific language impairment (Lammertink, Boersma, Wijnen, & Rispens, 2017). Dyslexia is an obvious area of interest given the predicted outcome is an orthographic phonological interaction.

Appendices

Appendix 1: Experiment 1 (neighbor generation) stimuli

Example	bi3	bian3	chi3	di1	diu1	fei2	fo2
	hua2	huang2	ka3	kua3	lie4	luo4	mian2
	miao2	miu4	nie4	nue4	ou1	pao4	piao4
	ran2	rang2	shan1	shan4	sou1	tian2	tuan2
	you4	zen3	zhe4	zhen4	zi3	zong3	
Practice	bing1	cai2	chu1	fa3	guai4	mei2	reng2
	shuo1	song4	zui4				
Test	a1	ai4	an4	ang2	ao4	ba3	bie2
	bo1	che1	chui1	cong2	cuo4	de2	deng3
	di4	dia3	e4	ei4	en1	er2	fen4
	fu4	gai1	gen1	gua4	guan3	hei1	hong2
	hou4	hun4	huo2	ji1	jie1	jin4	jing3
	juan4	jue2	kao4	kuai4	kuang2	lao3	li3
	lin2	lun2	lv4	ma1	ming2	mo2	nan2
	ni3	niang2	niu2	nv3	ou4	pei2	pian4
	pin1	qi3	qian2	qing3	qiong2	qiu2	quan2
	ren2	ri4	rou4	ru2	san1	sen1	shang4
	shuai4	si3	suan4	tang3	tian1	tiao2	ting1
	tui3	wa1	wai4	wan2	wang4	wei2	wen4
	weng1	wo3	wu3	xia4	xiang3	xiao3	xin1
	xiong2	xue2	xun2	ya4	yan3	yang3	yao4
	ye2	yi1	yin1	ying2	yong4	you3	yu3
	yuan2	yue4	yun4	zai4	zheng4	zhua1	zhuang1
	zun1						

Appendix 2: N&H phoneme inventory

	IPA	Sampa	Pinyin word	Sampa word	Ortho word		IPA	Sampa	Pinyin word	Sampa word	Ortho word
Vowels	a	a	b**a**3	p**a**3	把	Plosives	p	p	**b**u4	**p**u4	不
	ə	@	sh**e**4	S**@**4	蛇		pʰ	P	**p**ao3	**P**aU3	跑
	e	e	g**ei**3	k**eI**3	给		k	k	**g**e0	**k**@0	个
	ɛ	E	y**e**3	i**E**3	也		kʰ	K	**k**e4	**K@**4	课
	ɨ	1	zh**i**1	Z**1**1	之		t	t	**d**ou1	**t**oU1	都
	i	i	d**i**4	t**i**4	第		tʰ	T	**t**a1	**T**a1	他
	ɪ	I	su**i**4	sue**I**4	岁	Fricatives	s	s	**s**uo3	**s**uo3	所
	o	o	ru**o**4	ru**o**4	若		f	f	**f**ang4	**f**aN4	放
	ʊ	U	cho**u**3	Co**U**3	丑		x	x	**h**ui4	**x**ueI4	会
	u	u	**w**o3	**u**o3	我	Affricates	ʂ	S	**sh**i4	**S**14	是
	y	y	**yu**an2	**y**En2	元		ɕ	X	**x**ia4	**X**ia4	下
Nasals	m	m	m**a**1	m**a**1	妈		tɕ	J	**j**iu4	**J**ioU4	就
	n	n	**n**eng2	**n**@N2	能		tɕʰ	Q	**q**ing3	**Q**iN3	请
	ŋ	N	xia**ng** 3	Xia**N**3	想		tsʰ	c	**c**ong2	**c**oN2	从
Liquids	l	l	**l**ie4	**l**iE4	列		tʂʰ	C	**ch**u1	**C**u1	出
	r	r	**r**ang4	**r**aN4	让		ts	z	**z**i4	**z**14	字
							tʂ	Z	**zh**e	**Z**@4	这

Appendix 3: Syllable inventory test results for the pinyin syllable 'yong4' according to edit distances between stimuli and response for the three syllable inventories. All stimuli and responses are shown in sampa.

N&H inventory			Z&L inventory			Lin inventory		
Stimuli	Response	Edit	Stimuli	Response	Edit	Stimuli	Response	Edit
ioN4	xoN1	2	iuN4	xuN1	2	juN4	xuN1	2
ioN4	JioN3	2	iuN4	JiuN3	2	juN4	JjuN3	2
ioN4	loN4	1	iuN4	luN4	1	juN4	luN4	1
ioN4	noN4	1	iuN4	nuN4	1	juN4	nuN4	1
ioN4	noN4	1	iuN4	nuN4	1	juN4	nuN4	1
ioN4	noN4	1	iuN4	nuN4	1	juN4	nuN4	1
ioN4	QioN2	2	iuN4	QiuN2	2	juN4	QjuN2	2
ioN4	r@N1	3	iuN4	r@N1	3	juN4	r@N1	3
ioN4	roN2	2	iuN4	ruN2	2	juN4	ruN2	2
ioN4	iaN4	1	iuN4	iAN4	1	juN4	jAN4	1
ioN4	iaU3	3	iuN4	iAu3	3	juN4	jAu3	3
ioN4	iN2	2	iuN4	iN2	2	juN4	j@N2	2
ioN4	ioN1	1	iuN4	iuN1	1	juN4	juN1	1
ioN4	ioN1	1	iuN4	iuN1	1	juN4	juN1	1
ioN4	ioN1	1	iuN4	iuN1	1	juN4	juN1	1
ioN4	ioN1	1	iuN4	iuN1	1	juN4	juN1	1
ioN4	ioN1	1	iuN4	iuN1	1	juN4	juN1	1
ioN4	ioN1	1	iuN4	iuN1	1	juN4	juN1	1
ioN4	ioN1	1	iuN4	iuN1	1	juN4	juN1	1
ioN4	ioN2	1	iuN4	iuN2	1	juN4	juN2	1
ioN4	ioN2	1	iuN4	iuN2	1	juN4	juN2	1
ioN4	ioN3	1	iuN4	iuN3	1	juN4	juN3	1
ioN4	ioU3	2	iuN4	i@u3	3	juN4	jou3	3
ioN4	ioU3	2	iuN4	i@u3	3	juN4	jou3	3
ioN4	ioU4	1	iuN4	i@u4	2	juN4	jou4	2
ioN4	ioU4	1	iuN4	i@u4	2	juN4	jou4	2
ioN4	ioU4	1	iuN4	i@u4	2	juN4	jou4	2
ioN4	yEn2	4	iuN4	yan2	4	juN4	HEn2	4
ioN4	yEn3	4	iuN4	yan3	4	juN4	HEn3	4
ioN4	yn1	4	iuN4	yn1	4	juN4	Hyn1	4
ioN4	yn4	3	iuN4	yn4	3	juN4	Hyn4	3
ioN4	yn4	3	iuN4	yn4	3	juN4	Hyn4	3
	Mean:	1.75			1.91			1.91

Appendix 4: Experiment 2 (neighbor generation) stimuli

a1	ai4	an4	ang2	ao1	ba3	ban1	bei3
bian3	bie2	bing1	bo1	cai2	che1	chi1	chu1
chui1	ci4	cong2	cuo4	dai4	de2	di1	dia3
diao4	die1	diu1	du4	duo1	e4	ei4	en1
er3	fa3	fei2	fen4	feng1	fo2	fu4	gai1
gang1	gao3	ge1	gua4	guai4	guan3	guang1	gun3
hei1	hen3	hong2	hou4	hua2	huai4	huan4	huang2
hun4	huo2	ji1	jia1	jiang1	jie1	jin4	jing3
jiong3	jiu3	ju2	juan4	jue2	jun1	ka3	ke3
kua3	kuai4	kuang2	kun4	la1	lao3	li3	lia3
liang2	lie4	liu1	lun2	lv4	ma1	mai3	mao1
mei2	men2	mian2	miao2	min2	ming2	mo2	nan2
ni3	niang2	niu2	nong4	nue4	nv3	ou4	pao4
pei2	pian4	piao4	pin1	qi3	qia1	qian2	qiang2
qin2	qing3	qiong2	qiu2	qu4	quan2	que1	qun2
ran2	rang4	ren2	reng1	ri4	rou4	ru2	ruan3
ruo4	san3	sang1	shan1	shang4	sheng3	shi2	shua1
shuai4	shuang3	shui3	shuo1	si3	song4	suan4	sui4
tang3	tian2	tiao1	ting1	tuan2	tui3	wa1	wai4
wan2	wang4	wei2	wen4	weng4	wo4	wu3	xia4
xiang3	xiao3	xie2	xin1	xiong2	xu1	xuan3	xue2
xun1	ya4	yan3	yang3	yao4	ye2	yi1	yin1
ying2	yong3	you3	yu3	yuan2	yue4	yun4	zai3
zao3	zeng4	zhan4	zhe4	zhen4	zheng4	zhong1	zhu4
zhua1	zhuang1	zi3	zong3	zui4	zun1		

Appendix 5: Experiment 2 (neighbor generation) stimuli lexical statistics

	Freq M (SD)	HD M (SD)	PND M (SD)	PNF M (SD)	fPND M (SD)	nfPND M (SD)	POD M (SD)	fPOD M (SD)	nfPOD M (SD)
C_V_C	4.15	18.80	33.50	4.91	142.86	166.11	52.30	225.43	259.53
	(0.89)	(14.8)	(9.69)	(0.51)	(58.78)	(55.65)	(21.77)	(120.68)	(120.21)
C_G_V_C	4.16	18.90	26.43	4.89	111.82	130.31	45.33	194.86	223.94
	(0.89)	(14.81)	(7.24)	(0.45)	(44.13)	(41.17)	(18.25)	(102.26)	(100.18)
C_G_V_X	4.16	18.90	23.72	4.90	100.01	116.82	42.63	183.05	210.66
	(0.89)	(14.81)	(6.83)	(0.45)	(39.55)	(36.74)	(17.46)	(96.34)	(94.39)
C_G_VX	4.16	18.90	34.68	4.85	146.82	169.48	53.58	229.86	262.17
	(0.89)	(14.81)	(9.36)	(0.41)	(56.86)	(52.2)	(20.54)	(115.75)	(110.87)
C_GVX	4.15	18.80	51.20	4.91	217.64	252.99	70.00	300.21	346.11
	(0.89)	(14.8)	(13.67)	(0.43)	(84.51)	(76.09)	(26.02)	(147.23)	(139.99)
CG_V_X	4.15	18.82	28.14	4.85	117.71	136.98	46.96	200.36	229.41
	(0.89)	(14.8)	(9.59)	(0.4)	(49.33)	(49.17)	(17.56)	(97.24)	(93.75)
CG_VX	4.15	18.82	37.14	4.80	156.90	179.77	55.95	239.55	271.13
	(0.89)	(14.8)	(11.26)	(0.34)	(62.94)	(59.73)	(20.63)	(116.07)	(110.08)
CGVX	4.15	18.8	119.82	2.47	536.48	314.98	138.62	619.05	363.82
	(0.89)	(14.8)	(85.91)	(0.48)	(430.22)	(250.53)	(97.8)	(492.72)	(287.31)
C_V_C_T	3.73	5.53	16.43	4.23	62.80	70.89	21.95	84.39	94.75
	(0.94)	(4.59)	(6.37)	(0.53)	(32.4)	(32.14)	(8.28)	(42.75)	(42.64)
C_G_V_C_T	3.73	5.53	15.23	4.26	58.64	66.05	20.76	80.24	90.08
	(0.94)	(4.59)	(6.08)	(0.54)	(31.13)	(30.62)	(8.03)	(41.36)	(41.2)
C_G_V_X_T	3.73	5.53	14.16	4.28	54.29	61.39	19.68	75.89	85.47
	(0.94)	(4.59)	(5.84)	(0.55)	(28.92)	(28.89)	(7.78)	(38.98)	(39.19)
C_G_VX_T	3.73	5.53	18.13	4.22	69.92	77.64	23.65	91.52	101.36
	(0.94)	(4.59)	(7.5)	(0.5)	(37.33)	(36.53)	(9.25)	(47.06)	(46.09)
C_GVX_T	3.73	5.53	24.95	4.22	94.39	106.43	30.47	115.99	130.10
	(0.94)	(4.59)	(8.2)	(0.46)	(42.17)	(40.72)	(10.02)	(52.33)	(50.43)
CG_V_X_T	3.73	5.53	18.81	4.22	70.96	80.58	24.33	92.56	104.24
	(0.94)	(4.59)	(7.77)	(0.56)	(36.39)	(37.71)	(8.8)	(44.08)	(44.24)
CG_VX_T	3.73	5.53	22.85	4.17	86.73	96.57	28.38	108.33	119.93
	(0.94)	(4.59)	(8.16)	(0.49)	(40.77)	(39.37)	(9.23)	(48.9)	(45.9)
CGVX_T	3.73	5.53	286.07	4.18	1073.06	1200.60	291.60	1094.66	1223.62
	(0.94)	(4.59)	(34.22)	(0.18)	(313.48)	(177.76)	(34.3)	(320.31)	(178.34)

Appendix 6: Experiment 3 (auditory word repetition) stimuli

ai4	an4	ang2	ao1	ba3	ban1	bei3	bian3
bie2	bing1	bo1	cai2	che1	chi1	chu1	chui1
ci4	cong2	cuo4	dai4	de2	di1	diao4	die1
diu1	du4	duo1	e4	en1	er3	fa3	fei2
fen4	feng1	fo2	fu4	gai1	gang1	gao3	ge1
gua4	guai4	guan3	guang1	gun3	hei1	hen3	hong2
hou4	hua2	huai4	huan4	huang2	hun4	huo2	ji1
jia1	jiang1	jie1	jin4	jing3	jiong3	jiu3	juan4
jue2	jun1	jv2	ka3	ke3	kua3	kuai4	kuang2
kun4	la1	lao3	li3	liang2	lie4	liu1	lun2
lv4	ma1	mai3	mao1	mei2	men2	mian2	miao2
min2	ming2	mo2	nan2	ni3	niang2	niu2	nong4
nv3	pao4	pei2	pian4	piao4	pin1	qi3	qia1
qian2	qiang2	qin2	qing3	qiong2	qiu2	quan2	que1
qun2	qv4	ran2	rang4	ren2	reng1	rou4	ru2
ruan3	ruo4	san3	sang1	shan1	shang4	sheng3	shi2
shua1	shuai4	shuang3	shui3	shuo1	si3	song4	suan4
sui4	tang3	tian2	tiao1	ting1	tuan2	tui3	wa1
wai4	wan2	wang4	wei2	wen4	weng4	wo4	wu3
xia4	xiang3	xiao3	xie2	xin1	xiong2	xu1	xuan3
xue2	xun1	ya4	yan3	yang3	yao4	ye2	yi1
yin1	ying2	yong3	you3	yu3	yuan2	yue4	yun4
zai3	zao3	zeng4	zhan4	zhe4	zhen4	zheng4	zhong1
zhu4	zhua1	zhuang1	zi3	zong3	zui4	zun1	

Appendix 7: Experiment 3 (auditory word repetition) stimuli lexical statistics

	Freq M (SD)	HD M (SD)	PND M (SD)	PNF M (SD)	fPND M (SD)	nfPND M (SD)	POD M (SD)	fPOD M (SD)	nfPOD M (SD)
C_V_C	4.19 (0.8)	18.43 (13.04)	33.28 (9.58)	4.88 (0.51)	142.73 (56.24)	164.24 (55.13)	51.71 (19.99)	223.12 (109.92)	255.3 (111.02)
C_G_V_C	4.19 (0.79)	18.57 (13.08)	26.55 (7.21)	4.87 (0.46)	113.02 (42.67)	130.56 (41.21)	45.12 (16.74)	194.02 (93.34)	222.23 (93.31)
C_G_V_X	4.19 (0.79)	18.57 (13.08)	23.93 (6.78)	4.89 (0.46)	101.53 (38.26)	117.55 (36.68)	42.5 (16.02)	182.53 (87.87)	209.38 (87.56)
C_G_VX	4.19 (0.79)	18.57 (13.08)	34.78 (9.26)	4.84 (0.42)	147.99 (54.62)	169.65 (52.07)	53.35 (18.86)	228.99 (105.81)	260.53 (103.9)
C_GVX	4.19 (0.8)	18.43 (13.04)	50.99 (13.45)	4.9 (0.44)	217.84 (80.31)	251.36 (75.54)	69.42 (24.12)	298.23 (134.98)	342.41 (131.44)
CG_V_X	4.19 (0.8)	18.45 (13.04)	28.38 (9.66)	4.83 (0.4)	119.7 (48.3)	137.58 (49.43)	46.83 (16.49)	200.18 (89.61)	227.78 (88.24)
CG_VX	4.19 (0.8)	18.45 (13.04)	37.2 (11.25)	4.79 (0.34)	158.05 (60.82)	179.47 (59.76)	55.65 (19.13)	238.53 (106.36)	268.79 (103.35)
CGVX	4.19 (0.8)	18.43 (13.04)	120.61 (82.58)	2.5 (0.4)	538.14 (413.77)	315.9 (239.24)	139.04 (92.96)	618.53 (468.55)	363.59 (270.94)
C_V_C_T	3.76 (0.88)	5.49 (4.47)	15.95 (5.9)	4.21 (0.53)	60.98 (29.01)	68.25 (29.54)	21.44 (7.66)	82.37 (38.63)	91.81 (39.44)
C_G_V_C_T	3.76 (0.88)	5.49 (4.47)	15.09 (5.88)	4.25 (0.54)	57.97 (29.44)	65.14 (29.67)	20.58 (7.64)	79.36 (38.84)	88.98 (39.67)
C_G_V_X_T	3.76 (0.88)	5.49 (4.47)	14.12 (5.73)	4.26 (0.55)	54.02 (27.7)	60.91 (28.32)	19.61 (7.49)	75.41 (36.97)	84.78 (38.07)
C_G_VX_T	3.76 (0.88)	5.49 (4.47)	17.88 (7.17)	4.2 (0.5)	68.75 (34.93)	76.25 (35.01)	23.37 (8.73)	90.14 (43.75)	99.77 (43.93)
C_GVX_T	3.76 (0.88)	5.49 (4.47)	24.36 (7.54)	4.2 (0.46)	92.17 (37.16)	103.29 (37.35)	29.85 (9.22)	113.56 (46.56)	126.72 (46.47)
CG_V_X_T	3.76 (0.88)	5.49 (4.47)	18.83 (7.83)	4.2 (0.55)	71.54 (35.86)	80.27 (37.79)	24.32 (8.83)	92.94 (43.1)	103.74 (44.36)
CG_VX_T	3.76 (0.88)	5.49 (4.47)	22.67 (8.02)	4.16 (0.49)	86.39 (38.94)	95.41 (38.49)	28.16 (8.98)	107.78 (46.38)	118.6 (44.75)
CGVX_T	3.76 (0.88)	5.49 (4.47)	284.98 (34.61)	4.18 (0.18)	1076.15 (299.85)	1194.33 (179.49)	290.48 (34.71)	1097.55 (306.04)	1217.18 (180.19)

Appendix 8: Experiment 4 (auditory word repetition) stimuli

3-segment monosyllable CVN structure		4-segment monosyllable CGVN structure		3-segment disyllable CV V structure		4-segment disyllable CV CV structure	
ben3	men1	bian3	luan3	bi3 yu4	ji1 e4	ba1 li2	ju4 da4
can3	men2	chuan3	luan4	bu4 yi3	ji4 yi4	bu2 bi4	ju4 li2
cang2	ming4	chuan4	mian2	chu3 yu2	ke3 wu4	bu4 xu3	ka3 che1
ceng2	nin2	chuang1	niang2	da4 yi1	ke3 yi2	bu4 zhi1	ke4 hu4
chong3	nong4	chuang3	nuan3	de2 yi3	ke4 yi4	chu3 li3	ku4 zi0
chun3	pin1	chuang4	pian1	de2 yi4	li3 wu4	da3 du3	li4 shi3
cun2	qing3	cuan4	quan1	de2 yu3	li4 yi4	di4 si4	mi4 ma3
dan3	qun2	duan1	quan3	di2 yi4	qi3 e2	di4 tu2	mu4 shi1
deng3	san1	guang3	quan4	di4 wu3	shi2 wu3	di4 zhi3	qi3 su4
dong3	san4	guang4	shuan1	di4 yi1	shi4 wu4	du2 zi4	sha1 si3
fan3	sang1	huan1	suan1	di4 yu4	shu3 yu2	fa1 shi4	shi1 qu4
fang4	shan3	huan3	tian3	du2 wu4	si4 yu3	fu4 za2	shu4 zi4
gang3	song1	huang1	tuan2	fa3 yi1	te4 yi4	ge1 qu3	te4 shu1
gun4	tang4	huang3	xiang2	fu2 wu4	ti3 yu4	gu1 ji4	za2 zhi4
hen4	zan3	juan1	xiong2	fu2 yi4	xi1 yi4	he2 zi0	zhu4 he4
kong4	zen3	kuan1	xuan3	fu4 yu3	zhi1 yi1	ji1 hu1	zi4 sha1
kun4	zheng3	kuan3	zhuan1	gu3 wu3	zhi2 wu4	ji2 le0	zu3 zhi1
leng3	zong3	kuang2	zuan1	gu4 yi4	zhu3 yi4	ji4 zhu4	zu3 zhi3

Appendix 9: Experiment 4 (auditory word repetition) stimuli lexical statistics

	Freq M (SD)	HD M (SD)	PND M (SD)	PNF M (SD)	fPND M (SD)	nfPND M (SD)	POD M (SD)	fPOD M (SD)	nfPOD M (SD)
C_V_C	4.07 (0.52)	15.04 (8.70)	33.08 (8.01)	4.83 (0.54)	137.03 (45.59)	160.56 (44.41)	48.12 (14.54)	200.15 (76.78)	233.66 (79.03)
C_G_V_C	4.07 (0.52)	15.05 (8.71)	26.96 (5.22)	4.85 (0.47)	110.81 (29.69)	131.37 (29.72)	42.01 (12.15)	173.99 (62.65)	204.73 (64.84)
C_G_V_X	4.07 (0.52)	15.05 (8.71)	25.11 (5.75)	4.85 (0.47)	103.09 (29.95)	122.04 (30.69)	40.16 (12.55)	166.26 (62.69)	195.37 (65.91)
C_G_VX	4.07 (0.52)	15.05 (8.71)	34.53 (7.09)	4.76 (0.38)	141.18 (35.83)	164.78 (37.68)	49.59 (13.35)	204.35 (66.69)	236.82 (70.36)
C_GVX	4.07 (0.52)	15.04 (8.7)	49.12 (10.95)	4.80 (0.48)	203.32 (62.53)	236.73 (60.54)	64.16 (18.02)	266.44 (95.18)	309.33 (96.23)
CG_V_X	4.07 (0.52)	15.05 (8.71)	29.75 (10.19)	4.83 (0.46)	121.52 (45.91)	143.70 (51.24)	44.80 (15.49)	184.70 (72.58)	216.76 (79.09)
CG_VX	4.07 (0.52)	15.05 (8.71)	39.93 (10.89)	4.76 (0.34)	163.15 (50.45)	190.68 (55.52)	54.99 (16.00)	226.32 (76.75)	262.63 (83.07)
CGVX	4.07 (0.52)	15.04 (8.7)	122.64 (76.71)	2.51 (0.42)	525.17 (360.12)	316.12 (216.88)	137.68 (82.2)	588.29 (387.34)	354.48 (233.18)
C_V_C_T	3.58 (0.36)	2.81 (1.81)	16.59 (4.88)	4.14 (0.55)	59.33 (18.77)	68.84 (22.3)	19.40 (5.29)	69.37 (20.56)	80.67 (25.22)
C_G_V_C_T	3.58 (0.36)	2.81 (1.81)	16.65 (4.56)	4.20 (0.51)	59.66 (18)	69.79 (20.05)	19.47 (4.99)	69.71 (19.88)	81.76 (22.9)
C_G_V_X_T	3.58 (0.36)	2.81 (1.81)	15.96 (5.15)	4.22 (0.51)	57.14 (19.78)	66.99 (21.7)	18.77 (5.53)	67.19 (21.48)	78.99 (24.05)
C_G_VX_T	3.58 (0.36)	2.81 (1.81)	21.05 (6.13)	4.08 (0.39)	75.50 (23.86)	85.57 (25.21)	23.87 (6.36)	85.54 (25.15)	97.26 (27.33)
C_GVX_T	3.58 (0.36)	2.81 (1.81)	25.29 (5.53)	4.12 (0.54)	90.59 (22.58)	103.89 (25.89)	28.11 (6.25)	100.63 (25.34)	115.62 (30.26)
CG_V_X_T	3.58 (0.36)	2.81 (1.81)	19.48 (8.85)	4.20 (0.54)	69.85 (32.96)	81.89 (39.3)	22.29 (9.08)	79.89 (34.24)	93.84 (40.75)
CG_VX_T	3.58 (0.36)	2.81 (1.81)	25.25 (8.58)	4.12 (0.4)	90.55 (32.68)	104.37 (37.85)	28.07 (8.78)	100.59 (33.94)	116.14 (39.56)
CGVX_T	3.58 (0.36)	2.81 (1.81)	290.19 (29.36)	4.22 (0.17)	1038.15 (148.4)	1228.28 (152.61)	293.00 (29.22)	1048.19 (148.72)	1240.13 (152.18)

Appendix 10: Experiment 5 (auditory lexical decision) word stimuli

bai2	bei1	ben3	cai1	can3
cao1	chang3	cheng1	chong2	chou1
dai1	dan1	dao1	ding1	dong3
fan1	fang4	gong4	gun3	han3
hua1	huo3	jia4	jun1	kao3
ken3	kong1	kun4	lao2	lei2
leng3	ling4	lou2	man3	mao1
meng4	ming4	mou3	nei4	nin2
nong4	pai2	pan2	pang4	pei4
peng4	qun2	rou4	ruo4	sai4
suo1	shao3	shen2	sheng3	shou1
shua3	tan4	tang4	tong4	tou2
wang4	xia1	xie3	xin4	xing3
yang3	yen3	zao3	zen3	zong3
zuo3	zhang3	zheng3	zhun3	zhuo1

Appendix 11: Experiment 5 (auditory lexical decision) tone gap nonword stimuli

tsʰən3	tsʰəŋ3	tsʰaŋ3	tʂʰua2	tʂʰun4
tʂʰuo3	foʊ1	fuo4	tɕiŋ2	kaŋ2
keɪ1	kua2	kʰəŋ4	kʰan2	kʰaʊ2
kʰua2	kʰuo3	laɪ1	lan1	lia1
lyɛ1	mən3	maɪ1	meɪ1	min4
nəŋ1	nən3	naɪ1	niŋ1	noʊ1
nyɛ3	pan2	paŋ2	pin2	pʰiɛ2
pʰiŋ4	pʰoʊ4	tɕʰyɛ3	rən1	rəŋ4
ran4	raʊ1	run1	səŋ4	san2
soŋ2	soʊ2	sun4	ʂaɪ2	ʂaŋ2
ʂeɪ1	ʂun1	ʂuo3	təŋ2	teɪ4
tia2	toŋ2	tun2	tʰəŋ3	tʰiɛ2
uəŋ2	uaɪ2	xən1	xəŋ3	xaŋ3
xeɪ3	ɕyn3	tsəŋ2	tsaɪ2	tsaŋ2
tseɪ1	tsun2	tʂeɪ3	tʂoŋ2	tʂua2

Appendix 12: Experiment 5 (auditory lexical decision) syllable gap nonword stimuli

tsʰiŋ1	tsʰua2	tsʰyŋ1	tʂʰyŋ1	faɪ2
faʊ2	fiɛ1	iaɪ4	tɕiə2	kɛn2
kia2	kiɛ1	kiŋ3	kuŋ2	kʰia2
kʰiŋ4	kʰyɛ3	kʰyn2	lən4	lɛn1
lua4	lyŋ1	miə1	mon4	mun1
muŋ1	nɛn1	niə3	nia4	nyn3
nyŋ4	pia1	pon3	poŋ4	poʊ4
puŋ2	pʰɛn3	pʰoŋ2	pʰun1	pʰuŋ3
pʰyɛ3	tɕʰyŋ3	raɪ4	reɪ4	riɛ1
riŋ2	ruŋ1	seɪ4	siɛ1	sin1
siŋ4	sua1	ʂon2	ʂoŋ3	ʂɛn2
ʂiŋ3	tɛn2	tiə4	tua3	tyɛ3
tʰeɪ4	tʰiə2	tʰia1	tʰua1	uaʊ3
uoʊ4	xin2	xon1	xyɛ3	xyn3
ɕiə3	tsia3	tsiɛ4	tsiŋ3	tsua1

Appendix 13: Experiment 5 (auditory lexical decision) word stimuli lexical statistics

	Freq M (SD)	HD M (SD)	PND M (SD)	PNF M (SD)	fPND M (SD)	nfPND M (SD)	POD M (SD)	fPOD M (SD)	nfPOD M (SD)
C_V_C	4.07 (0.52)	15.04 (8.70)	33.08 (8.01)	4.83 (0.54)	137.03 (45.59)	160.56 (44.41)	48.12 (14.54)	200.15 (76.78)	233.66 (79.03)
C_G_V_C	4.07 (0.52)	15.05 (8.71)	26.96 (5.22)	4.85 (0.47)	110.81 (29.69)	131.37 (29.72)	42.01 (12.15)	173.99 (62.65)	204.73 (64.84)
C_G_V_X	4.07 (0.52)	15.05 (8.71)	25.11 (5.75)	4.85 (0.47)	103.09 (29.95)	122.04 (30.69)	40.16 (12.55)	166.26 (62.69)	195.37 (65.91)
C_G_VX	4.07 (0.52)	15.05 (8.71)	34.53 (7.09)	4.76 (0.38)	141.18 (35.83)	164.78 (37.68)	49.59 (13.35)	204.35 (66.69)	236.82 (70.36)
C_GVX	4.07 (0.52)	15.04 (8.70)	49.12 (10.95)	4.80 (0.48)	203.32 (62.53)	236.73 (60.54)	64.16 (18.02)	266.44 (95.18)	309.33 (96.23)
CG_V_X	4.07 (0.52)	15.05 (8.71)	29.75 (10.19)	4.83 (0.46)	121.52 (45.91)	143.70 (51.24)	44.80 (15.49)	184.70 (72.58)	216.76 (79.09)
CG_VX	4.07 (0.52)	15.05 (8.71)	39.93 (10.89)	4.76 (0.34)	163.15 (50.45)	190.68 (55.52)	54.99 (16.00)	226.32 (76.75)	262.63 (83.07)
CGVX	4.07 (0.52)	15.04 (8.70)	122.64 (76.71)	2.51 (0.42)	525.17 (360.12)	316.12 (216.88)	137.68 (82.20)	588.29 (387.34)	354.48 (233.18)
C_V_C_T	3.58 (0.36)	2.81 (1.81)	16.59 (4.88)	4.14 (0.55)	59.33 (18.77)	68.84 (22.3)	19.40 (5.29)	69.37 (20.56)	80.67 (25.22)
C_G_V_C_T	3.58 (0.36)	2.81 (1.81)	16.65 (4.56)	4.20 (0.51)	59.66 (18.00)	69.79 (20.05)	19.47 (4.99)	69.71 (19.88)	81.76 (22.9)
C_G_V_X_T	3.58 (0.36)	2.81 (1.81)	15.96 (5.15)	4.22 (0.51)	57.14 (19.78)	66.99 (21.70)	18.77 (5.53)	67.19 (21.48)	78.99 (24.05)
C_G_VX_T	3.58 (0.36)	2.81 (1.81)	21.05 (6.13)	4.08 (0.39)	75.50 (23.86)	85.57 (25.21)	23.87 (6.36)	85.54 (25.15)	97.26 (27.33)
C_GVX_T	3.58 (0.36)	2.81 (1.81)	25.29 (5.53)	4.12 (0.54)	90.59 (22.58)	103.89 (25.89)	28.11 (6.25)	100.63 (25.34)	115.62 (30.26)
CG_V_X_T	3.58 (0.36)	2.81 (1.81)	19.48 (8.85)	4.20 (0.54)	69.85 (32.96)	81.89 (39.3)	22.29 (9.08)	79.89 (34.24)	93.84 (40.75)
CG_VX_T	3.58 (0.36)	2.81 (1.81)	25.25 (8.58)	4.12 (0.4)	90.55 (32.68)	104.37 (37.85)	28.07 (8.78)	100.59 (33.94)	116.14 (39.56)
CGVX_T	3.58 (0.36)	2.81 (1.81)	290.19 (29.36)	4.22 (0.17)	1038.15 (148.4)	1228.28 (152.61)	293.00 (29.22)	1048.19 (148.72)	1240.13 (152.18)

Appendix 14: Experiment 5 (auditory lexical decision) tone gap nonword lexical statistics

	PND M (SD)	PNF M (SD)	nfPND M (SD)		PND M (SD)	PNF M (SD)	nfPND M (SD)
C_V_C	28.03 (6.77)	4.79 (0.54)	134.75 (36.8)	C_V_C_T	14.13 (3.96)	4.09 (0.61)	58.99 (22.76)
C_G_V_C	24.12 (5.04)	4.79 (0.49)	116.21 (29.49)	C_G_V_C_T	14.04 (4.26)	4.13 (0.57)	58.57 (20.9)
C_G_V_X	22.48 (5.59)	4.80 (0.49)	108.06 (29.45)	C_G_V_X_T	13.28 (4.49)	4.15 (0.58)	55.51 (21.16)
C_G_VX	30.52 (7.42)	4.75 (0.44)	145.25 (37.66)	C_G_VX_T	16.33 (5.6)	4.02 (0.62)	66.29 (25.51)
C_GVX	40.29 (10.14)	4.77 (0.46)	192.44 (51.47)	C_GVX_T	19.11 (5.46)	4.05 (0.64)	78.51 (29.49)
CG_V_X	26.85 (8.35)	4.76 (0.45)	128.49 (42.83)	CG_V_X_T	16.73 (5.62)	4.08 (0.49)	68.59 (25.12)
CG_VX	34.93 (10.42)	4.73 (0.36)	165.57 (50.87)	CG_VX_T	20.21 (6.3)	4.00 (0.51)	81.23 (28.33)
CGVX	51.27 (59.78)	2.17 (0.75)	131.77 (169.81)	CGVX_T	273.41 (35.68)	4.13 (0.17)	1133.58 (182.02)

Appendix 15: Experiment 5 (auditory lexical decision) syllable gap nonword lexical statistics

	PND M (SD)	PNF M (SD)	nfPND M (SD)		PND M (SD)	PNF M (SD)	nfPND M (SD)
C_V_C	17.08 (8.07)	4.59 (0.47)	80.17 (41.10)	C_V_C_T	9.17 (6.12)	3.56 (1.13)	36.83 (26.95)
C_G_V_C	14.64 (7.65)	4.63 (0.51)	69.31 (40.39)	C_G_V_C_T	7.64 (5.97)	3.79 (0.73)	30.88 (27.19)
C_G_V_X	13.43 (6.66)	4.63 (0.51)	63.33 (34.45)	C_G_V_X_T	7.09 (5.39)	3.78 (0.72)	28.12 (23.13)
C_G_VX	21.28 (8.36)	4.72 (0.45)	101.10 (43.05)	C_G_VX_T	11.00 (7.5)	3.94 (0.67)	44.82 (32.6)
C_GVX	28.89 (8.69)	4.68 (0.35)	135.95 (44.66)	C_GVX_T	16.13 (7.97)	3.93 (0.52)	64.77 (34.14)
CG_V_X	19.04 (9.75)	4.67 (0.51)	91.73 (53.13)	CG_V_X_T	12.45 (8.24)	3.84 (0.93)	52.50 (40.87)
CG_VX	26.35 (9.96)	4.75 (0.42)	126.87 (53.48)	CG_VX_T	16.63 (8.51)	4.08 (0.56)	69.98 (42.27)
CGVX	0 (0)	0 (0)	0 (0)	CGVX_T	283.87 (33.25)	4.18 (0.18)	1188.98 (171.13)

References

Albright, A. (2008). Gradient phonological acceptability as a grammatical effect. *Phonology*, 1–31. Retrieved from https://link.springer.com/content/pdf/10.3758/BF03208845.pdf

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements : Evidence for Continuous Mapping Models. *Journal of Memory and Language*, (38), 419–439.

Amano, S., & Kondo, T. (1999). Neighborhood effects on spoken word recognition in Japanese. In *Eurospeech* (pp. 1663–1666). Retrieved from https://pdfs.semanticscholar.org/e0ad/362dbc168ab47f981c84eb9a5a3acc36fbc0.pdf

Amano, S., & Kondo, T. (2000). Neighborhood and cohort in lexical processing of Japanese spoken words. In A. Cutler, J. McQueen, & R. Zondervan (Eds.), *SWAP-2000: SCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes* (pp. 91–94). Nijmegen.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, *10*(6), 565–575. https://doi.org/10.3758/BF03202439

Andrews, S. (1989). Frequency and neighbourhod effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(5), 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 234–254. https://doi.org/10.1037/0278-7393.18.2.234

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

Ao, B. X. (1992). The non-uniqueness condition and the segmentation of the Chinese syllable. *Working Papers in Linguistics*, *42*, 1–25.

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The Structure of Phonological Networks Across Multiple Languages. *International Journal of Bifurcation and Chaos*, *20*(3), 679–685.

Arnold, H. S., Conture, E. G., & Ohde, R. N. (2005). Phonological neighborhood density in the picture naming of young children who stutter : Preliminary study. *Journal of Fluency Disorders*, *30*(2), 125–148. https://doi.org/10.1016/j.jfludis.2005.01.001

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R. Cambridge University Press* (Vol. 2). https://doi.org/10.1558/sols.v2i3.471

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., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55*, 290–313.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174–1220. https://doi.org/10.1080/02687038.2016.1147767

Baddeley, A. D. (1986). *Working memory* (Oxford psy). Oxford: Oxford University Press.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness : Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, *44*, 568–591. https://doi.org/10.1006/jmla.2000.2756

Bajo, T., Padilla, F., & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. Gallardo San Salvador, & Y. Gambier (Eds.), *Translation in Context: Selected papers from EST Congress, Granada 1998* (Benjamins, pp. 127–142). Amsterdam: John Benjamins.

Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A*, *38*(1), 364–374. https://doi.org/10.1103/PhysRevA.38.364

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Bao, Z. M. (1990). Fan-Qie Languages and Reduplication. *Linguistic Inquiry*, *21*.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barabási, A. L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and Its Applications*, *272*(1), 173–187. https://doi.org/10.1016/S0378-4371(99)00291-5

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 386–393. https://doi.org/10.1037/0096-1523.2.3.386

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

Baus, C., Costa, A., & Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes*, *23*(6), 866–888. https://doi.org/10.1080/01690960801962372

Beckman, M. E., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*, *71*(1), 240–249. https://doi.org/10.1111/1467-8624.00139

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., … Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(SUPPL. 1), 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, *2*(2), 160–176. https://doi.org/10.1093/jole/lzx001

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E., LotsofOtherPeople, & Tingley, D. (2018). Redefine Statistical Significance. *Nature Human Behavior*, *2*(1), 1–18. https://doi.org/10.17605/OSF.IO/MKY9J

Benkí, J. R. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *The Journal of the Acoustical Society of America*, *113*(3), 1689–1705. https://doi.org/10.1121/1.1534102

Best, C. T. (1991). *The emergence of native-language phonological influences in infants: A perceptual assimilation model*. *Haskins Laboratories Status Report on Speech Research* (Vol. SR-107). Retrieved from https://s3.amazonaws.com/academia.edu.documents/45189083/The_emergence_of_native-language_phonolo20160428-10583-1w6yx66.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1520857569&Signature=CZILuWdHja1xXKnp9fJYMiv5%2FsU%3D&response-content-disposition=inli

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production.* (pp. 13–34). Amsterdam: John Benjamins. Retrieved from https://pdfs.semanticscholar.org/d952/272d14c57d27f6e56f2eb417dc196e4a945e.pdf

Blumenfeld, H. K., & Marian, V. (2006). Phonological neighborhood density in native and non-native word production. *Proceedings of the Twenty-Eighth Annual Meeting of the Cognitive Science Society*, *2006*, 71–76. Retrieved from http://comm.soc.northwestern.edu/bilingualism-psycholinguistics/files/phonological.pdf

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2002). Functional anatomy of intra- and cross-modal lexical tasks. *NeuroImage*, *16*(1), 7–22. https://doi.org/10.1006/nimg.2002.1081

Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, C., & Booth, J. R. (2013). Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Human Brain Mapping*, *34*(12), 3354–3368. https://doi.org/10.1002/hbm.22147

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research*, *33*(2), 103–135. https://doi.org/10.1023/B:JOPR.0000017223.98667.10

Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology*, *10*, 12–21. https://doi.org/10.1080/17470215808416249

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How Many Words Do We Know ? Practical Estimates of Vocabulary Size Dependent on Word Definition , the Degree of Language Input and the Participant ' s Age. *Frontiers in Psychology*, *7*(July), 1–11. https://doi.org/10.3389/fpsyg.2016.01116

Burnham, K. P., & Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information- Theoretic Approach, Second Edition*. Springer.

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

Calvo, N., Ibáñez, A., & García, A. M. (2016). The Impact of Bilingualism on Working Memory: A Null Effect on the Whole May Not Be So on the Parts. *Frontiers in Psychology*, *7*(February), 265. https://doi.org/10.3389/fpsyg.2016.00265

Cao, F., Brennan, C., & Booth, J. R. (2015). The brain adapts to orthography with experience: Evidence from English and Chinese. *Developmental Science*, *18*(5), 785–798. https://doi.org/10.1111/desc.12245

Cao, F., & Perfetti, C. A. (2016). Neural signatures of the reading-writing connection: Greater involvement of writing in Chinese reading than english reading. *PLoS ONE*, *11*(12). https://doi.org/10.1371/journal.pone.0168414

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

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *23*(4), 857–871. https://doi.org/10.1037/0278-7393.23.4.857

Carreiras, M., Seghier, M. L., Baquero, S., Estévez, A., Lozano, A., Devlin, J. T., & Price, C. J. (2009). An anatomical signature for literacy. *Nature*, *461*(7266), 983–986. https://doi.org/10.1038/nature08461

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1934–1949. https://doi.org/10.1037/a0016902.The

Chan, K. Y., & Vitevitch, M. S. (2010). Network Structure Influences Speech Production. *Cognitive Science*, *34*, 685–697. https://doi.org/10.1111/j.1551-6709.2010.01100.x

Chao, Y.-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica*, *4*(4), 363–398. Retrieved from http://www.airitilibrary.com/Publication/alDetailedMesh?docid=10124195-193401-201011160124-201011160124-363-398

Chao, Y. R. (1931). *Fan-qie Yu Ba Zhong (Trans. Eight Varieties of Secret Language Based on the Principle of Fan-qie)*. *Bulletin of the Institute of History and Philology*.

Charles-Luce, J., & Luce, P. A. (1990). Similarity Neighbourhoods of Words In Young Children's Lexicons. *Journal of Child Language*, *17*, 205–215. https://doi.org/10.1017/S0305000900013180

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, *22*(3), 727–735. https://doi.org/10.1017/S0305000900010023

Chen, H., & Devichand, M. (2015). Millions share new Chinese character. Retrieved January 9, 2016, from http://www.bbc.com/news/blogs-trending-31689148

Chen, J.-Y. (1993). A small corpus of speech errors in Mandarin Chinese and their classification. *Word of Chinese Language*, *69*, 26–41.

Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics*, *20*(2), 289–301. https://doi.org/doi:10.1017/S0142716499002064

Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia: An International Journal of Psychology in the Orient*.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-Form Encoding in Mandarin Chinese as Assessed by the Implicit Priming Task. *Journal of Memory and Language*, *46*(January), 751–781. https://doi.org/10.1006/jmla.2001.2825

Chen, J.-Y., Lin, W.-C., & Ferrand, L. (2003). Masked Priming of the Syllable in Mandarin Chinese Speech Production. *Chinese Journal of Psychology*, *45*(1), 107–120.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417–430. https://doi.org/10.1037/a0027175

Chen, T.-M., & Chen, J.-Y. (2013). The syllable as the proximate unit in Mandarin Chinese word production: An intrinsic or accidental property of the production system? *Psychonomic Bulletin & Review*, *20*, 154–162. https://doi.org/10.3758/s13423-012-0326-7

Chen, W.-F., Chao, P.-C., Chang, Y.-N., Hsu, C.-H., & Lee, C.-Y. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Language*, *157–158*, 51–62. https://doi.org/10.1016/j.bandl.2016.04.005

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*(2), 135–158. https://doi.org/10.1017/S0022226700001444

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324–345. https://doi.org/10.1016/j.jml.2005.12.004

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception & Performance*, *16*(3), 551–563.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*. https://doi.org/10.1038/386604a0

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Connell, K., Tremblay, A., & Zhang, J. (2016). The Timing of Acoustic vs . Perceptual Availability of Segmental and Suprasegmental Information. In *Proceedings of the 5th International Symposium on Tonal Aspects of Languages (TAL 2016)* (pp. 99–102). Buffalo, New York. https://doi.org/10.21437/TAL.2016-21

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193–210.

Cornish, H. (2010). Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems. *Interaction Studies*, *11*(1), 112–137. https://doi.org/10.1075/is.11.1.02cor

Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: What experiments can tell us. *Language Learning*, *59*(SUPPL. 1), 187–205. https://doi.org/10.1111/j.1467-9922.2009.00540.x

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory & Cognition*, *28*(5), 746–755. https://doi.org/10.3758/BF03198409

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

de Bot, K. (2015). Circadian rhythms and second language development. *International Journal of Bilingualism*, *19*(2), 142–155. https://doi.org/10.1177/1367006913489201 de Bot, K., Chan, H., Lowie, W., Plat, R., & Verspoor, M. (2012). A dynamic perspective on language processing and development. *Dutch Journal of Applied Linguistics*, *1*(2), 188–218. https://doi.org/10.1075/dujal.1.2.03deb

de Bot, K., & Lowie, W. (2010). Conceptual representations in the multilingual mind. In M. Pütz & L. Sicola (Eds.), *Cognitive Processing in Second Language Acquisition* (pp. 117–133). Amsterdam/Philadelphia: John Benjamins Publishing Company. Retrieved from https://www.researchgate.net/profile/Marjolijn_Verspoor/publication/236234839_Conceptual_representations_in_the_multilingual_mind/links/0deec52cc18bd1b59a000000.pdf#page=126

de Gelder, B., & Vroomen, J. (1992). Auditory and visual speech perception in alphabetic and non-alphabetic Chinese-Dutch bilinguals. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (pp. 413–426). Amsterdam: Elsevier.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., … Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359–1364. https://doi.org/10.1126/science.1194140

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283. Retrieved from http://129.237.66.221/P800/dell1986.pdf

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes? In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9–37).

Dell, G. S., Reed, K. ristopher D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(6), 1355–1367. https://doi.org/10.1037/0278-7393.26.6.1355

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the Time Course of Spoken Word Recognition: Electrophysiological Evidence for the Influences of Phonological Similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893–1906. https://doi.org/10.1162/jocn.2008.21142

Detrain, C., & Deneubourg, J.-L. (2006). Self-organized structures in a superorganism: do ants “behave” like molecules? *Physics of Life Reviews*, *3*, 162–187. https://doi.org/10.1016/j.plrev.2006.07.001

Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed.). Oxford: Oxford University Press.

Duanmu, S. (2010). Chinese Syllable Structure. In *The Blackwell companion to phonology* (pp. 2151–2777).

Duanmu, S. (2017). From non-uniqueness to the best solution in phonemic analysis: evidence from Chengdu Chinese. *Lingua Sinica*, *3*(1), 15. https://doi.org/10.1186/s40655-017-0030-7

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039. https://doi.org/10.1016/S0048-7333(00)00135-9

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 696–713. https://doi.org/10.1037/0278-7393.22.3.696

Frauenfelder, U. H., Baayen, R. H., & Hellwig, F. M. (1993). Neighborhood Density and Frequency Across Languages and Modalities. *Journal of Memory and Language*, *32*(6), 781–804.

Freedman, S. E., & Barlow, J. A. (2013). Effects of Neighborhood Density on Adult Word Repetition. *International Journal of English Linguistics*, *3*(2), 1–19. https://doi.org/10.1007/s10549-015-3663-1.Progestin

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords. *Journal of Memory and Language*, *42*(4), 481–496. https://doi.org/10.1006/jmla.1999.2692.Perception

Fromkin, V. A. (1971). The nonanomalous nature of anomalous utterances. *Language*, *47*, 27–52.

Fromkin, V. A. (1973). *Speech errrors as linguistic evidence*. The Hague: Mouton.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806. https://doi.org/10.1016/j.jml.2011.11.006

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966–979. https://doi.org/10.1037//0012-1649.33.6.966

Ge, J., Peng, G., Lyu, B., Wang, Y., Zhuo, Y., Niu, Z., … Gao, J.-H. (2015). Cross-language differences in the brain network subserving intelligible speech. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 2972–2977. https://doi.org/10.1073/pnas.1416000112

Gell-Mann, M. (1994). *The Quark and the Jaguar*. New York: Freeman & Co.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming Lexical Neighbors of Spoken Words: Effects of Competition and Inhibition. *Journal of Memory and Language*, *28*(5), 501–518. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3901307/

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, *14*(6), 249–258. https://doi.org/10.1016/j.tics.2010.02.004

Gordon, J. K., & Dell, G. S. (2001). Phonological Neighborhood Effects: Evidence from Aphasia and Connectionist Modeling. *Brain and Language*, *79*, 21–31. https://doi.org/10.1006/brln.2001.2574

Gordon, J. K., & Kurczek, J. C. (2014). The Aging Neighborhood: Phonological Density in Naming. *Language and Cognitive Processes*, *29*(3), 326–344. https://doi.org/10.1016/j.sbspro.2011.09.208

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*, 228–244.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, *35*(5), 623–647. https://doi.org/10.1006/jmla.1996.0033

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*, *103*(3), 518–565. https://doi.org/10.1037/0033-295X.103.3.518

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

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception and Psychophysics*, *51*(1), 49–56. https://doi.org/10.3758/BF03205073

Greenberg, J. H., & Jenkins, J. J. (1967). Studies in the psychological correlates of the sound system of American English. In A. Jacobivits & M. S. Miron (Eds.), *Readings in the psychology of language* (pp. 186–200). Englewood Cliffs, New Jersey: Prentice Hall.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 481–503. https://doi.org/10.1037/0096-1523.23.2.481

Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition*, *97*(3), 227–267. https://doi.org/10.1016/j.cognition.2004.09.006

Huang, H.-W., Lee, C.-Y., Tsai, J.-L., Lee, C.-L., Hung, D. L., & Tzeng, O. J.-L. (2006). Orthographic neighborhood effects in reading Chinese two-character words. *Neuroreport*, *17*(10), 1061–1065. https://doi.org/10.1097/01.wnr.0000224761.77206.1d

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, *2*(Article 255), 1–16. https://doi.org/10.3389/fpsyg.2011.00255

Jansen, P. A., & Watter, S. (2008). SayWhen: an automated method for high-accuracy speech onset detection. *Behavior Research Methods*, *40*(3), 744–751. https://doi.org/10.3758/BRM.40.3.744

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture-naming interference. *Quarterly Journal of Experimental Psychology*, *54*, 371–382.

Jha, A. P., & McCarthy, G. (2000). The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. *Journal of Cognitive Neuroscience*, *12*, 90–105. https://doi.org/10.1162/089892900564091

Kauffman, S. A., Shymko, R. M., & Trabert, K. (1978). Control of sequential compartment formation in Drosophila. *Science*, *199*(4326), 259–270. https://doi.org/10.1126/science.413193

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10681–6. https://doi.org/10.1073/pnas.0707835105

Kirby, S. M. (1999). *Function, Selection and Innateness*. Oxford: Oxford University Press.

Krafnick, A. J., Tan, L.-H., Flowers, D. L., Luetje, M. M., Napoliello, E. M., Siok, W.-T., … Eden, G. F. (2016). Chinese Character and English Word processing in children's ventral occipitotemporal cortex: fMRI evidence for script invariance. *Neuroimage*, *133*, 302–312. https://doi.org/10.1016/j.neuroimage.2016.03.021

Kureta, Y., Fushimi, T., Sakuma, N., & Tatsumi, I. F. (2015). Orthographic influences on the word-onset phoneme preparation effect in native Japanese speakers: Evidence from the word-form preparation paradigm. *Japanese Psychological Research*, *57*(1), 50–60. https://doi.org/10.1111/jpr.12067

Kureta, Y., Fushimi, T., & Tatsumi, I. F. (2006). The functional unit in phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1102–1119.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Package "lmerTest." R. Retrieved from http://cran.uib.no/web/packages/lmerTest/lmerTest.pdf

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in Specific Language Impairment: A meta-analysis. *Journal of Speech, Language and Hearing Research*, *60*(December), 1–13. https://doi.org/10.1044/2017_JSLHR-L-16-0439

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, *12*, 119–131.

Lee, W.-S. (2005). A phonetic study of the "er-hua" rimes in Beijing Mandarin. In *INTERSPEECH* (pp. 1093–1096). Lisbon, Portugal. Retrieved from http://www.isca-speech.org/archive/archive_papers/interspeech_2005/i05_1093.pdf

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1). https://doi.org/10.1017/S0025100303001208

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38-75. https://doi.org/10.1017/S0140525X99001776

Li, M.-F., Gao, X.-Y., Chou, T.-L., & Wu, J.-T. (2017). Neighborhood Frequency Effect in Chinese Word Recognition: Evidence from Naming and Lexical Decision. *Journal of Psycholinguistic Research*, *46*, 227–245. https://doi.org/10.1007/s10936-016-9431-5

Li, M.-F., Lin, W.-C., Chou, T.-L., Yang, F.-L., & Wu, J.-T. (2015). The Role of Orthographic Neighborhood Size Effects in Chinese Word Recognition. *Journal of Psycholinguisic Research*, *44*, 219–236. https://doi.org/10.1007/s10936-014-9340-4

Li, Q.-L., Bi, H.-Y., Wei, T.-Q., & Chen, B.-G. (2011). Orthographic neighborhood size effect in Chinese character naming: Orthographic and phonological activations. *Acta Psychologica*, *136*(1), 35–41. https://doi.org/10.1016/j.actpsy.2010.09.012

Li, Q.-L., Bi, H.-Y., & Zhang, J. X. (2010). Neural correlates of the orthographic neighborhood size effect in Chinese. *European Journal of Neuroscience*, *32*(5), 1–7. https://doi.org/10.1111/j.1460-9568.2010.07327.x

Li, Y., Zhang, L., Xia, Z., Yang, J., Shu, H., & Li, P. (2017). The Relationship between Intrinsic Couplings of the Visual Word Form Area with Spoken Language Network and Reading Ability in Children and Adults. *Frontiers in Human Neuroscience*, *11*(June). https://doi.org/10.3389/fnhum.2017.00327

Lin, Y. H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology*. University of Texas at Austin.

Lin, Y. H. (2007). *The Sounds of Chinese with Audio CD*. Cambridge University Press.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*(3), 615–625. https://doi.org/10.3758/BF03212113

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, *16*(5–6), 565–581. https://doi.org/10.1080/01690960143000137

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spken Words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/MPG.0b013e3181a15ae8.Screening

MacKay, D. G., & Burke, D. M. (1990). Cognition and Aging: a theory of new learning and the use of old connections. In T. H. Hess (Ed.), *Aging and Cognition: Knowledge organization and utilization* (pp. 213–263). North-Holland: Elsevier Science Publishers B.V. https://doi.org/10.1016/S0166-4115(08)60159-4

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The Dynamics of Lexical Competition During Spoken Word Recognition, *31*, 133–156.

Malins, J. G., Gao, D., Tao, R., Booth, J. R., Shu, H., Joanisse, M. F., … Desroches, A. S. (2014). Developmental differences in the influence of phonological similarity on spoken word processing in Mandarin Chinese. *Brain and Language*, *138*, 38–50. https://doi.org/10.1016/j.bandl.2014.09.002

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407–420. https://doi.org/10.1016/j.jml.2010.02.004

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*(8), 2032–2043. https://doi.org/10.1016/j.neuropsychologia.2012.05.002

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). Clearpond: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0043230

Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to Phonological Similarity Within and Across Languages. *Journal of Psycholinguisic Research*, *37*(3), 141–170. https://doi.org/10.1007/s10936-007-9064-9

Marks, E. A., Moates, D. R., Bond, Z. S., & Stockmal, V. (2002). Word reconstruction and consonant features in English and Spanish. *Linguistics*, *40*(2), 421–438. https://doi.org/10.1515/ling.2002.018

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing Spoken Words : The Importance of Word Onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576–585.

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

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

Meyer, A. S. (1990). The Time Course of Phonological Encoding in Language Production: Phonological Encoding Inside a Syllable. *Journal of Memory and Language*, *30*, 69–89.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

Mok, P. P. K. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*, *2*, 148–154. https://doi.org/10.1515/iral.1964.2.1.155

Morais, J., Cary, L., Alegria, J., & Bettelson, P. (1979). Does awarenesss of speech as a sequence of phonemes arise spontaneously? *Cognition*, *7*, 323–331. Retrieved from https://s3.amazonaws.com/academia.edu.documents/44353672/Does_Awareness_of_Speech_as_a_Sequence_o20160403-31999-84u1ld.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1520506944&Signature=PrFifwBc6S2AqV1%2BzD1nB9hUo7Y%3D&response-content-disposition=inlin

Morowitz, H. J., & Singer, J. L. (1995). *The Mind, The Brain, And Complex Adaptive Systems*. Boulder, Colorado: Westview Press.

Muneaux, M., & Ziegler, J. C. (2004). Locus of orthographic effects in spoken word recognition: Novel insights from the neighbour generation task. *Language and Cognitive Processes*, *19*(5), 641–660. https://doi.org/10.1080/01690960444000052

Murray, J. D. (1988). How the Leopard Gets Its Spots. *Scientific American*, *258*(3), 80–87. Retrieved from https://pdfs.semanticscholar.org/7ea8/71b9a25b129f0c74e799e52bd063cb0bc4f5.pdf

Myers, J. (2002). An analogical approach to the Mandarin syllabary. *Journal of Chinese Phonology*, *11*, 163–190.

Myers, J., & Tsay, J. (2005). The processing of phonological acceptability judgments. *Proceedings of Symposium on 90-92 NSC Projects*, 26–45.

Neergaard, K. D., Xu, H., & Huang, C.-R. (2016). Database of Mandarin Neighborhood Statistics Database of Mandarin Neighborhood Statistics. In *Language Resources and Evaluation Conference 2016* (pp. 2270–2277). Portoroz, Slovenia. Retrieved from https://www.researchgate.net/publication/303288164_Database_of_Mandarin_Neighborhood_Statistics

Newman, M. E. J. (2002). Assortative Mixing in Networks. *Physical Review Letters*, *89*(20), 1–5. https://doi.org/10.1103/PhysRevLett.89.208701

Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, *69*(3), 1–9. https://doi.org/10.1103/PhysRevE.68.036122

Newman, R. S., & Bernstein Ratner, N. (2007). The Role of Selected Lexical Factors on Confrontation Naming Accuracy, Speed, and Fluency in Adults Who Do and Do Not Stutter. *Journal of Speech, Language, and Hearing Research*, *50*(February), 196–213. Retrieved from http://jslhr.pubs.asha.org/article.aspx?articleid=1765440

Newman, R. S., Sawusch, J. R., & Luce, P. A. (2005). Do postonset segments define a lexical neighborhood ? *Memory & Cognition*, *33*(6), 941–960.

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

Norris, D. (1994). Shortlist - a Connectionist Model of Continuous Speech Recognition. *Cognition*, *52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Nuzzo, R. (2014). Statistical Errors. *Nature News*, *506*, 150–152. https://doi.org/10.1038/506150a

O'Séaghdha, P. G., & Chen, J.-Y. (2009). Toward a Language-General Account of Word Production : The Proximate Units Principle. In *CogSci... Annual Conference of the Cognitive Science Society* (Vol. July-Augus, pp. 68–73). Amsterdam, Netherlands: Cognitive Science Society. Retrieved from http://csjarchive.cogsci.rpi.edu/Proceedings/2009/papers/13/paper13.pdf

O'Séaghdha, P. G., Chen, J.-Y., & Chen, T. (2010). Proximate Units in Word Production: Phonological Encoding Begins with Syllables in Mandarin Chinese but with Segments in English. *Cognition*, *115*(2), 282–302. https://doi.org/10.1016/j.cognition.2010.01.001.Proximate

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, *37*(3), 346–357. https://doi.org/10.3758/MC.37.3.346

Onnis, L., Frank, S. L., Yun, H., & Lou-Magnuson, M. (2016). Statistical learning bias predicts second-language reading efficiency. In *CogSci... Annual Conference of the Cognitive Science Society*. Austin, Texas: Cognitive Science Society. https://doi.org/10.1063/1.3033202

Onnis, L., Lou-Magnuson, M., Yun, H., & Thiessen, E. (2015). Is statistical learning trainable? In *Cogsci* (pp. 1781–1786).

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, (June), 1047–1052.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing Luca. *Journal of Memory and Language*, *53*(2), 225–237. Retrieved from http://www.annualreviews.org/doi/10.1146/annurev-psych-122216-011805

Pattamadilok, C., Knierim, I. N., Kawabata Duncan, K. J., & Devlin, J. T. (2010). How Does Learning to Read Affect Speech Perception? *Journal of Neuroscience*, *30*(25), 8435–8444. https://doi.org/10.1523/JNEUROSCI.5791-09.2010

Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language and Cognitive Processes*, *22*(5), 700–726. https://doi.org/10.1080/01690960601049628

Pattamadilok, C., Perre, L., Dufau, S., & Ziegler, J. C. (2009). On-line Orthographic Influences on Spoken Language in a Semantic Task. *Journal of Cognitive Neuroscience*, *21*(1), 169–179. https://doi.org/10.1162/jocn.2009.21014

Pattamadilok, C., Perre, L., & Ziegler, J. C. (2011). Beyond rhyme or reason: ERPs reveal task-specific activation of orthography on spoken language. *Brain and Language*, *116*, 116–124. https://doi.org/10.1016/j.bandl.2010.12.002

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 767–779. https://doi.org/10.1037/0096-1523.24.3.767

Pérez, M. A. (2007). Age of acquisition persists as the main factor in picture naming when cumulative word frequency and frequency trajectory are controlled. *The Quarterly Journal of Experimental Psychology*, *60*(1), 32–42. https://doi.org/10.1080/17470210600577423

Perre, L., Midgley, K., & Ziegler, J. C. (2009). When beef primes reef more than leaf : Orthographic information affects phonological priming in spoken word recognition. *Psychophysiology*, *46*, 739–746. https://doi.org/10.1111/j.1469-8986.2009.00813.x

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research*, *1275*, 73–80. https://doi.org/10.1016/j.brainres.2009.04.018

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198. https://doi.org/10.1037/h0049234

Pike, K. L. (1947). Grammatical Prerequisites to Phonemic Analysis. *Word*, *3*(November), 155–172. https://doi.org/10.1080/00437956.1947.11659314

Psychological Software Tools, I. (2012). E-Prime 2.0. Pitsburgh, PA: Psychological Software Tools, Inc.

Quillian, M. R. (1962). A revised design for an understanding machine. *Mechanical Translation*, *7*, 17–29.

Rae, P. J. L., & Perfect, T. J. (2014). Visual distraction during word-list retrieval does not consistently disrupt memory. *Frontiers in Psychology*, *5*(April), 362. https://doi.org/10.3389/fpsyg.2014.00362

Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography Influences the Perception and Production of Speech. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(6), 1588–1594. https://doi.org/10.1037/a0024833

Reali, F., & Griffiths, T. L. (2008). The evolution of frequency distributions : Relating regularization to inductive biases through iterated learning. In *Proceedings of the Annual Meeting of the Cognitive Science* (Vol. 30, pp. 229–234). Washington, DC: Cognitive Science Society. Retrieved from https://cloudfront.escholarship.org/dist/prd/content/qt6td4749z/qt6td4749z.pdf

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019–1034. https://doi.org/10.1037//0278-7393.28.6.1019

Roodenrys, S., & Miller, L. M. (2008). A constrained Rasch model of trace redintegration in serial recall. *Memory & Cognition*, *36*(3), 578–587. https://doi.org/Doi 10.3758/Mc.36.3.578

Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33.58. https://doi.org/10.1016/j.cogpsych.2013.10.001

Saffran, J. R. (2003). Statistical language learning. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Salthouse, T. A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychological Bulletin*, *137*(5), 753–784. https://doi.org/10.1037/a0023262.Neuroanatomical

Schiller, N. O. (1998). The Effect of Visually Masked Syllable Primes on the Naming Latencies of Words and Pictures. *Journal of Memory and Language*, *39*, 484–507.

Schiller, N. O. (1999). Masked Syllable Priming of English Nouns. *Brain and Language*, *68*, 300–305.

Schiller, N. O. (2000). Single Word Production in English: The Role of Subsyllabic Units During Phonological Encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 512–528. https://doi.org/10.10371/0278-7393.26.2.512

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., & Chen, H.-C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, *17*(1), 1–12. https://doi.org/10.1162/0898929052880057

Schmid, M. S. (2013). First Language Attrition. *Linguistic Approaches to Bilingualism*, *3*(1), 94–115. https://doi.org/10.1017/CBO9781107415324.004

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 419–425. https://doi.org/10.1037/0278-7393.12.3.419

Sears, C. R., Campbell, C. R., & Lupker, S. J. (2006). Is there a neighborhood frequency effect in english? evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1040–1062. https://doi.org/10.1037/0096-1523.32.4.1040

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood frequency and neighborhood size effects in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, *21*(4), 876–900.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic Effects on Rhyme Monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 546–554.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*(3), 383–404. https://doi.org/10.1016/S0022-5371(84)90270-6

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shen, J. (1993). Slips of the tongue and the syllable structure of Mandarin Chinese. In S.-C. Yau (Ed.), *Essays on the Chinese language by contemporary Chinese scholars* (pp. 139–161). Paris, France: Centre de Recherdches Linguistiques sur l'Asie Orientale—Ecole des Hautes Etudes en Sciences Sociales.

Shih, C., & Ao, B. (1997). Duration Study for the Bell Laboratories Mandarin Test-to-Speech System. In J. van Santen, R. Sproat, J. Olive, & J. Hirschberg (Eds.), *Progress in Speech Synthesis* (pp. 383–399). New York: Springer-Verlag. Retrieved from http://books.google.com/books?hl=en&lr=&id=rc3nG34LtvkC&oi=fnd&pg=PR5&dq=Progress+in+Speech+Synthesis&ots=6xqai30F-6&sig=YmbCEbUMjJ-IzcqNauSlg3KasXc

Shuai, L., & Malins, J. G. (2016). Encoding lexical tones in jTRACE : a simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods*. https://doi.org/10.3758/s13428-015-0690-0

Siok, W. T., Niu, Z., Jin, Z., Perfetti, C. A., & Tan, L. H. (2008). A structural-functional basis for dyslexia in the cortex of Chinese readers. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(14), 5561–5566. https://doi.org/10.1073/pnas.0801750105

Siok, W. T., Perfetti, C. A., Jin, Z., & Tan, L. H. (2004). Biological abnormality of impaired reading is constrained by culture. *Nature*, *431*(7004), 71–76. https://doi.org/Doi 10.1038/Nature02865

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449. https://doi.org/10.1016/j.cognition.2010.06.004

Sommers, M. S., & Lewis, B. P. (1999). Who Really Lives Next Door: Creating False Memories with Phonological Neighbors. *Journal of Memory and Language*, *40*(1), 83–108. https://doi.org/10.1006/jmla.1998.2614

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory & Cognition*, *6*(4), 410–415. https://doi.org/10.3758/BF03197473

Stemberger, J. P. (2004). Neighbourhood effects on error rates in speech production. *Brain and Language*, *90*(1–3), 413–422. https://doi.org/10.1016/S0093-934X(03)00452-8

Storkel, H. L. (2001). Learning new words: Phonotactic probability in Language Development. *Journal of Speech, Language, and Hearing Research*, *44*, 1321–1337. https://doi.org/10.1044/1092-4388(2003/102)

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1312–23. https://doi.org/10.1044/1092-4388(2003/102)

Storkel, H. L. (2004). Methods for Minimizing the Confounding Effects of Word Length in the Analysis of Phonotactic Probability and Neighborhood Density. *Journal of Speech, Language, and Hearing Research*, *47*(December), 1454–1468. https://doi.org/1092-4388/04/4706-1454

Storkel, H. L., Armbruster, J., & Hogan, T. P. (2006). Differentiating Phonotactic Probability and Neighborhood Density in Adult Word Learning. *Journal of Speech, Language, and Hearing Research*, *49*(December), 1175–1192. https://doi.org/10.1044/1092-4388(2006/085)

Storkel, H. L., & Morrisette, M. L. (2002). The Lexicon and Phonology: Interactions in Language Acquisition. *Language, Speech, and Hearing Services in Schools*, *33*, 24–37. https://doi.org/10.1044/0161-1461(2002/003)

Strand, J., & Liben-Nowell, D. (2016). Making Long-Distance Relationships Work: Quantifying Lexical Competition with Hidden Markov Models. *Journal of Memory and Language*, *90*, 88–102. Retrieved from http://cs.carleton.edu/faculty/dlibenno/papers/ldr-confusions/ldr.pdf

Suárez, L., Tan, S. H., Yap, M. J., & Goh, W. D. (2011). Observing neighborhood effects without neighbors. *Psychonomic Bulletin & Review*, *18*, 605–611. https://doi.org/10.3758/s13423-011-0078-9

Swadesh, M. (1934). The phonemic principle. *Language*, *2*, 117–129.

Tabak, W., Schreuder, R., & Baayen, R. H. (2010). Producing in inflected verbs: A picture naming study. *The Mental Lexicon*, *5*(1), 22–46. Retrieved from http://www.sfs.uni-tuebingen.de/~hbaayen/publications/TabakSchreuderBaayen2010.pdf

Tamaoka, K., & Makioka, S. (2009). Japanese mental syllabary and effects of mora, syllable, bi-mora and word frequencies on Japanese speech production. *Language and Speech*, *52*(1), 79–112. https://doi.org/10.1177/0023830908099884

Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human Brain Mapping*, *25*(1), 83–91. https://doi.org/10.1002/hbm.20134

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

Tan, L. H., Spinks, J. A., Gao, J. H., Liu, H. L., Perfetti, C. A., Xiong, J., … Fox, P. T. (2000). Brain activation in the processing of Chinese characters and words: A functional MRI study. *Human Brain Mapping*, *10*(1), 16–27. https://doi.org/10.1002/(SICI)1097-0193(200005)10:1<16::AID-HBM30>3.0.CO;2-M

Tolan, G. A., & Tehan, G. (1999). Determinants of short-term forgetting: decay, retroactive interference or proactive interference? *International Journal of Psychology*, *34*(5/6), 285–292.

Tsai, J.-L., Lee, C.-Y., Lin, Y.-C., Tzeng, O. J. L., & Hung, D. L. (2006). Neighborhood Size Effects of Chinese Words in Lexical Decision and Reading. *Language and Linguistics*, *7*(3), 659–675. Retrieved from http://search.proquest.com/docview/85639553?accountid=13042

Tsai, P.-T. (2007). *The effects of phonological neighborhoods on spoken word recognition in Mandarin Chinese*. Universty of Maryland, College Park.

Tukey, J. W. (1977). *Exploratory data analysis* (1st ed.). Pearson.

van Ooijen, B. (1996). Vowel mutability and lexical selection in English: evidence from a word reconstruction task. *Memory & Cognition*, *24*(5), 573–583. https://doi.org/10.3758/BF03201084

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, *19*(1), 57–95. https://doi.org/10.1080/01690960344000134

Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W., & Schiller, N. O. (2011). The functional unit of Japanese word naming: Evidence from masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1458–1473. https://doi.org/10.1037/a0024491

Verdonschot, R. G., Nakayama, M., Zhang, Q., Tamaoka, K., & Schiller, N. O. (2013). The Proximate Phonological Unit of Chinese-English Bilinguals: Proficiency Matters. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061454

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735–747. https://doi.org/10.1037/0278-7393.28.4.735

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*, *51*, 408–422.

Vitevitch, M. S., Armbruster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production:effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514–529. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14979821

Vitevitch, M. S., Chan, K. Y., & Goldstein, R. (2014). Insights into failed lexical retrieval from network science. *Cognitive Psychology*, *68*(1), 1–32. https://doi.org/10.1016/j.cogpsych.2013.10.002

Vitevitch, M. S., Goldstein, R., & Johnson, E. (2016). Path-Length and the Misperception of Speech: Insights from Network Science and Psycholinguistics. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Eds.), *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (pp. 29–45). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-47238-5

Vitevitch, M. S., & Luce, P. A. (1998). When words complete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408. https://doi.org/10.1006/jmla.1998.2618

Vitevitch, M. S., Luce, P. A., Charles-Luce, J. A. N., & Kemmerer, D. (1997). Phonotactics and Syllable Stress: Implications for the Processing of Spoken Nonsense Words. *Language and Speech*, *40*(1), 47–62.

Vitevitch, M. S., & Rodríguez, E. (2004). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, *3*(1), 64–73. https://doi.org/10.1080/14769670400027332.Neighborhood

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4), 491–504. https://doi.org/10.1016/j.pestbp.2011.02.012.Investigations

Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, *21*(6), 760–770. https://doi.org/10.1080/01690960500287196

Vitevitch, M. S., & Stamer, M. K. (2009). *The influence of neighborhood density (and neighborhood frequency) in Spanish speech production: A follow-up report*. Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/5500/SLL_Technical_Report_1.pdf?sequence=1

Vitevitch, M. S., Stamer, M. K., & Sereno, J. A. (2008). Word length and lexical competition: Longer is the same as shorter. *Language and Speech*, *51*(4), 361–383. https://doi.org/10.1177/0023830908099070

Wan, I.-P. (2006). A Psycholinguistic Study of Postnuclear Glides and Coda Nasals in Mandarin. *Journal of Language and Linguistics*, *5*(2), 158–176.

Wan, I. P. (1996). *Tone errors in Mandarin Chinese*. State University of New York at Buffalo.

Wang, H. S. (1998). An experimental study on the phonetic constraints of Mandarin Chinese. In B. K. Tsou (Ed.), *Studia Linguistica Serica* (pp. 259–268). Hong Kong: City University of Hong Kong Language Information Sciences Research Center.

Wang, Q., & Zhang, J. (2011). N400 solution effect of Chinese character fragments: An orthographic neighborhood size effect. *Brain Research Bulletin*, *86*(3–4), 179–188. https://doi.org/10.1016/j.brainresbull.2011.07.006

Wang, W., Li, X., Ning, N., & Zhang, J. X. (2012). The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters*, *516*(1), 67–71. https://doi.org/10.1016/j.neulet.2012.03.059

Waters, G. S., Seidenberg, M. S., & Bruck, M. (1984). Children's and adults' use of spelling-sound information in three reading tasks. *Memory & Cognition*, *12*(3), 293–305. https://doi.org/10.3758/BF03197678

Watts, D. J., & Strogatz, S. H. (1998). Collectivedynamics of "small-world" networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Wen, W. (1980). Cong Yingwen de tongxingci lai kan hanyu pinyin wenzi de tongyinci [A study of Chinese homophones from the view of English homographs]. *Yuwen Xiandaihua [Modernizing Our Language]*, *2*, 120–124. Retrieved from https://www.scopus.com/record/display.uri?eid=2-s2.0-84958968714&origin=inward&txGid=fc57d8a12539b83b0db763a7af56a970

Wiener, S., & Turnbull, R. (2015). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech*, *59*(1), 1–24. https://doi.org/10.1177/0023830915578000

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

Wong, A. W.-K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review*, *16*(5), 888–892. https://doi.org/10.3758/PBR.16.5.888

Wong, A. W.-K., Huang, J., & Chen, H.-C. (2012). Phonological Units in Spoken Word Production: Insights from Cantonese. *PLoS ONE*, *7*(11), e48776. https://doi.org/10.1371/journal.pone.0048776

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, *1459*(August), 1–45. https://doi.org/10.1080/01621459.2016.1180986

Wood, S. N., & Scheipl, F. (2016). Generalized Additive Mixed Models using “mgcv” and “lme4.” Retrieved from http://cran.r-project.org/package=gamm4

Wood, S. N., Scheipl, F., & Faraway, J. J. (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, *23*(3), 341–360. https://doi.org/10.1007/s11222-012-9314-z

Wu, D. (2017). *Cross-regional word duration patterns in Mandarin*. University of Illinois at Urbana-Champaign.

Wu, F., & Kenstowicz, M. (2015). Duration Reflexes of Syllable Structure in Mandarin. *Lingua*, *164*, 87–99. Retrieved from http://lingphil.mit.edu/papers/kenstowicz/Wu-Kenstowicz.pdf

Wu, J.-T., Yang, F.-L., & Jin, W.-C. (2013). Beyond phonology matters in character recognition. *Chinese Journal of Psychology*, *55*(3), 289–318. https://doi.org/10.6129/CJP.20130608

Wu, J., Lu, J., Zhang, H., Zhang, J., Yao, C., Zhuang, D., … Zhou, L. (2015). Direct evidence from intraoperative electrocortical stimulation indicates shared and distinct speech production center between Chinese and English languages. *Human Brain Mapping*, *36*(12), 4972–4985. https://doi.org/10.1002/hbm.22991

Wurm, L. H. (2007). Danger and usefulness: an alternative framework for understanding rapid evaluation effects in perception? *Psychonomic Bulletin & Review*, *14*(6), 1218–1225. https://doi.org/10.3758/BF03193116

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*(1), 37–48. https://doi.org/10.1016/j.jml.2013.12.003

Xu, S. (1980). *Putonghua yuyin zhishi [Phonology of Standard Chinese]*. Beijing: Wenzi Gaige Chubans.

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11*(3), 452–457. https://doi.org/10.3758/BF03196594

Yip, M. (1982). Reduplication and C-V Skeleta in Chinese Secret Language. *Linguistic Inquiry*, *13*(4), 637–660. Retrieved from http://www.jstor.org/stable/4178299?seq=1#page_scan_tab_contents

You, R., Qian, N., & Gao, Z. (1980). Lun Putonghua de yinwei xitong [On the phonemic system of Standard Chinese]. *Zhongguo Yuwen*, *5*(158), 328–334.

You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked Syllable Priming Effects in Word and Picture Naming in Chinese. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0046595

Yudes, C., Macizo, P., & Bajo, T. (2011). The influence of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in Psychology*, *2*(Article 309). https://doi.org/10.3389/fpsyg.2011.00309

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*, 153–201. https://doi.org/10.1017/S0952675710000060

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*, 1761–1770. https://doi.org/10.1016/j.neuropsychologia.2011.02.054

Zhao, X., & Li, P. (2009). An online database of phonological representations for Mandarin Chinese. *Behavior Research Methods*, *41*(2), 575–583. https://doi.org/10.3758/BRM.41.2.575

Zhou, X., & Marslen-Wilson, W. (1999). Phonology, Orthography, and Semantic Activation in Reading Chinese. *Journal of Memory and Language*, *41*(4), 579–606. https://doi.org/10.1006/jmla.1999.2663

Zhou, Y. (2003). *The Historical Evolution of Chinese Languages and Scripts. Pathways to Advanced Skills Series, Volume 8.* Columbus, Ohio: Ohio State University National East Asian Languages Resource Center.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review*, *5*(4), 683–689. https://doi.org/10.3758/BF03208845

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory and Cognition*, *32*(5), 732–741. https://doi.org/10.3758/BF03195863

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*(4), 779–793. https://doi.org/10.1016/S0749-596X(03)00006-8

Zipf, G. K. (1934). *The psycho-biology of language*. Oxford, England: Houghton, Mifflin.

绯色toy. (2015). “成龙”我的洗发液. Retrieved January 9, 2016, from http://www.bilibili.com/video/av2023391/