Title

Investigation of Cognitive Mechanisms and Factors in Bilingual Lexical Entrainment

ISBN

9798293845507

Author

Song, Yongjia

Publication Date

2025-09-17

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Investigation of Cognitive Mechanisms and Factors in Bilingual Lexical Entrainment

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Language Science

by

Yongjia Song

Dissertation Committee:
Associate Professor Gregory Scontras, Chair
Associate Professor Richard Futrell
Assistant Professor Xin Xie
Associate Professor Iva M. Ivanova (University of Texas at El Paso)

2025

DEDICATION

To my parents, my best friend in life Yifan, and my sweet Lantern.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to many people in my life for their support on this journey. As I look back on ten years of study and life in the United States from my undergraduate years through my doctoral studies, I am filled with gratitude for all the people and experiences that have shaped this path.

First, my deepest gratitude goes to my committee chair and advisor, Greg. Since my undergraduate years, you have guided me into the world of research and supported my growth step by step. On this academic journey, your guidance, trust, support, and encouragement have given me the strength to persevere and move forward with confidence, even in the face of difficulties.

To all my committee members, Richard Futrell, who provided invaluable support on the modeling work, Xin Xie and Iva Ivanova, who have provided feedback and advice since the beginning of these projects, thank you so much for your support and help with these projects, as well as for your advice on this dissertation.

I would especially like to thank Judith Kroll. This story of bilingual lexical entrainment began with you — you handed me the seed of this idea and the journey began. With your support and guidance in my first three years, you deepened my passion for bilingualism and motivated me to continue exploring this field.

To all my research assistants — Jiachen, Jian, Kathlyn, and Yiqun — thank you for your trust and support in working with me, even when I was a new team leader. Danni, Huanyu, Noah, and Yeiza, I am very grateful for your invaluable efforts in helping with data cleaning on the final project during such a rushed period.

I also want to thank my cohort in the Language Science department for their friendship, feedback, and support. I truly enjoyed our discussions about projects and beyond. To Yuting, my former research assistant, my roommate, and my important friend - our friendship began with my very first project, and I believe it was a special serendipity that connected us. To Yiming, your support and patience during my struggles gave me the strength to keep going.

To my boyfriend Run, I feel so fortunate to have met you and to have someone with whom I can share all my emotions and thoughts. To my family and my best friend Yifan, all of my courage comes from your unwavering support.

VITA

Yongjia Song

EDUCATION

Doctor of Philosophy in Language Science **2025**
University of California, Irvine *Irvine, CA*

Bachelor of Arts in Psychology **2020**
University of California, Irvine *Irvine, CA*

ABSTRACT OF THE DISSERTATION

Investigation of Cognitive Mechanisms and Factors in Bilingual Lexical Entrainment

By

Yongjia Song

Doctor of Philosophy in Language Science

University of California, Irvine, 2025

Associate Professor Gregory Scontras, Chair

This dissertation investigates how bilinguals behave in lexical entrainment, a phenomenon in which speakers align their word choices with those of their conversational partners. By examining individual differences and production modality in shaping this behavior, this dissertation discusses variability in cognitive processing during lexical selection. In particular, I focus on bilinguals' lexical entrainment behavior to explore the individual differences related to language proficiency in this behavior. I find that bilinguals who are less proficient/dominant in English exhibit more entrainment, challenging prior assumptions that speakers' proficiency has a limited influence on their entrainment behavior. To investigate cognitive mechanisms underlying lexical entrainment, I develop a computational model within a Rate-Distortion of Control (RDC) framework, which integrates both automatic priming and rational control processes, and implement reinforcement learning to simulate the empirical findings. The model successfully predicts the empirical results and supports the hypothesis that both automatic priming and rational control processes are engaged in lexical entrainment. I also explore the effect of production modality in lexical entrainment. Empirical results demonstrate that bilinguals entrain more in typing compared to speaking, though no interaction effect with proficiency was observed. These findings contribute to our understanding of the selection stage in language production in bilinguals and provide insight into how cognitive processes may vary across modalities.

Chapter 1

Introduction

In conversation, we usually find that the language use of two interlocutors becomes similar as the conversation unfolds. The alignment at the lexical level is called lexical entrainment (Brennan and Clark, 1996), which means that people tend to repeat words from their conversational partners. Underlying this common behavior, the mechanisms and individual differences are questions that have long been studied to gain a better understanding of language production in interactive scenarios. In this dissertation, I focus on understanding lexical entrainment in bilingual populations, a group of people who have diverse language use, with an investigation into factors such as proficiency and production modality. Moreover, I explore the ways of understanding mechanisms in lexical entrainment at a computational level.

Starting from Chapter 2, I introduce basic theories of lexical production and lexical entrainment, discussing the relevant literature on experimental evidence of different mechanisms in lexical entrainment. Building on these theories, I discuss the variation among bilingual individuals and emphasize the possible factors that affect bilingual lexical entrainment.

Then Chapter 3 describes a series of experiments investigating proficiency in affecting bilingual lexical entrainment. The experiments examined the entrainment rate in English for bilinguals speaking English as one of their languages, and the entrainment rate in Mandarin for Mandarin-English speakers. I found that bilinguals who are less dominant in English—less proficient in English relative to their other language—entrain more. This result emphasizes the importance of proficiency in influencing lexical entrainment not only when bilingual as a listener but also as a speaker in conversation. It challenges the conclusions from previous research that assume speaker proficiency is less crucial in entrainment behavior.

What mechanisms engage in lexical entrainment is also a core question. One major theory from Pickering and Garrod (2006) states that a speaker and a listener align their words with each other due to automatic priming. On the other hand, Brennan and Clark (1996) proposed the establishment of conceptual pacts in conversation, which result in entrainment behaviors. The process of establishing conceptual pacts requires the involvement of cognitive control, so lexical entrainment is argued to be a strategic behavior with the engagement of cognitive control. In Chapter 4, I implement a Rate-Distortion of Control (RDC) model with policy gradient, a reinforcement learning method to predict the learning process occurring in entrainment, and develop a computational model to simulate the behavioral results of the study discussed in Chapter 3. I hypothesize that both automatic and rational processes from the two theories are applied in lexical entrainment. The model results successfully predict the empirical results and justify the coordination of both mechanisms in lexical entrainment.

Then, Chapter 5 investigates the effect of production modality on bilingual lexical entrainment, with the exploration of an interaction with proficiency. Typing, typically considered the same modality as writing, is gaining more and more attention in research. Although it employs a different encoding process compared to speaking, it is argued to share a common process of conceptualization in lexical production. However, it also brings differences in bilingual language production. Some studies state that speaking produces more anxiety than typing/writing for bilinguals, so different words or sentence structures are selected for different modalities. My study focuses on how different production modalities, in particular typing versus speaking, affect bilingual lexical entrainment, and whether production modality interacts with proficiency to affect the entrainment behaviors. This study proposes that different cognitive loads are applied in different production modalities, resulting in discrepancies in entrainment behaviors between speaking and typing, and an interaction effect occurs in affecting bilinguals' entrainment rates. The results presented the main effect of production modality, where bilinguals in typing entrained more than speaking, but the interaction effect was not observed. This result emphasizes the importance of modality in research on language production and also suggests how cognitive resources are allocated and cognitive mechanisms are managed differently across modalities.

Finally, Chapter 6 summarizes the findings regarding proficiency and production modality in bilingual lexical entrainment in experimental studies. It also discusses the possible manipulations of the factors within the model, presenting a more coherent picture of bilingual lexical entrainment with a clearer understanding of cognitive mechanisms involved in this special case of language production.

Chapter 2

Background

This chapter introduces the basic theories of lexical production and lexical entrainment, along with empirical evidence on how the partner's proficiency affects lexical entrainment. Moreover, it discusses the features of bilingual language production, highlighting gaps in research on bilingual lexical entrainment. Studying bilingual lexical entrainment with the manipulation of different factors would not only provide a more thorough understanding of individual differences in lexical entrainment but also offer insights into bilingual language production in the special case of interactive scenarios.

2.1 Lexical Production

The process of lexical production is usually separated into three stages: 1) conceptualization, 2) formulation, and 3) articulation (Dell, 1986; Schriefers et al., 1990; Levelt, 1993). In conceptualization, a speaker conceptualizes and selects the meanings they want to express. Then, the speaker formulates the words by accessing the semantic features and the phonological components, and finally articulates the sounds. A key question in this process is how a speaker accesses the word they intend to produce in the formulation stage. Most studies agree on the two steps in the process of lexical access in formulation: 1) semantic access, where the speaker accesses the meaning of the target word, and 2) phonological encoding, where the speaker accesses the phonological components of the target word. The controversial point is how these two steps connect to each other. There are two models hypothesizing the connections differently.

2.1.1 Interactive Two-step Model

One model of lexical access is the interactive two-step model developed by Dell et al. (1997) (Figure 2.1). This model is built on two assumptions: the spreading activation of lemma units and phonological units, and an interactive connection between the two steps. In the model, semantic features, lemmas, and phonological units are treated as nodes. Nodes of semantic features connect to nodes of lemmas, and nodes of lemmas connect to nodes of phonological units. When a node of semantic features is activated, it spreads the activation to all other connected lemma nodes with various activation levels, and these lemma nodes spread the activation to phonological units (Collins and Loftus, 1975; Dell, 1986). Due to the interactive connections, the activation of phonological units can also activate other lemmas connected to the phonological units. In the interactive two-step model, Dell et al. (1997) proposed that multiple lemmas are activated based on the semantic features that speakers wish to express. For example, when "cat" is the target concept to express, the lemma CAT is activated, along with other lemmas connected to shared semantic features, such as DOG. The lemma CAT then activates the linked phonological units /k/, /æ/, and /t/. According to the interactive connection assumption, these three phonological units affect the lemma stage and activate other connected lemmas with weaker connections, such as RAT. Eventually, in the step of articulation, the three phonological units with the strongest activation are articulated, and the word CAT is produced.

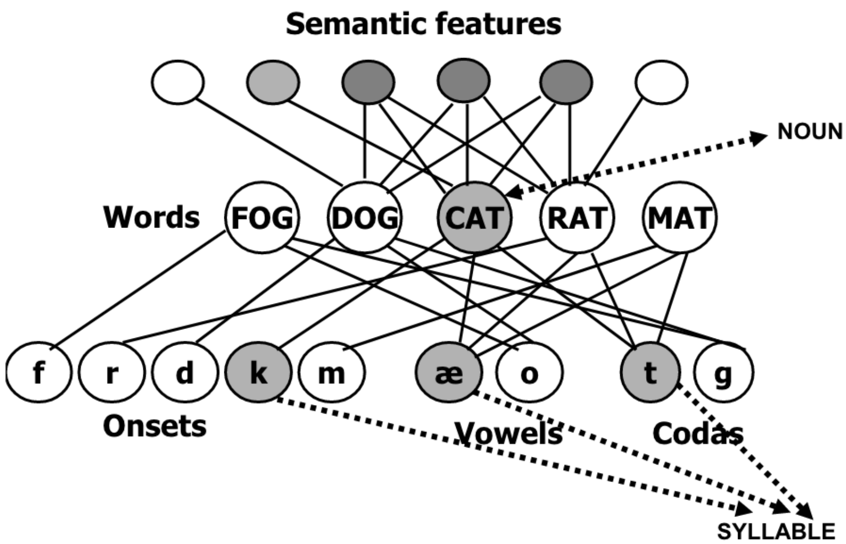


Figure 2.1: The interactive Two-step Model (Dell et al., 1997), the picture is from Dell (2014)

The occurrence of speech errors was argued to be evidence supporting this model (Dell et al., 1997). The mistake is caused by the activation of other nodes with shared semantic features or phonological components. For example, /kæg/ may be mistakenly pronounced instead of /kæt/ because DOG is activated by the shared semantic features, and DOG activates the related phonological unit /g/, which interferes with the production.

2.1.2 WEAVER++

Levelt et al. (1999) proposed another theory and developed a computational model, WEAVER++, to predict the process of lexical production (Figure 2.2). The shared idea between the two models is the two distinct steps of lemma access and phonological encoding. However, Levelt et al. (1999) proposed a different way of connecting these two steps. First, there is an additional step of lexical selection between lemma access and phonological encoding. In this model, the semantic features activate multiple lemmas with various strengths of activation.

The lemma with the strongest connection is selected. Then, the selected lemma activates its connected phonological units. The model also argues that the direction from lemma access to lexical selection and phonological encoding is feed-forward, meaning that the direction is one-way. The activation at the lemma level determines the selection and affects the activation of phonological units, but phonological nodes cannot affect the previous step of lemma selection. The additional step of lexical selection and the feature of feed-forward connection ensure the accuracy of lexical production.

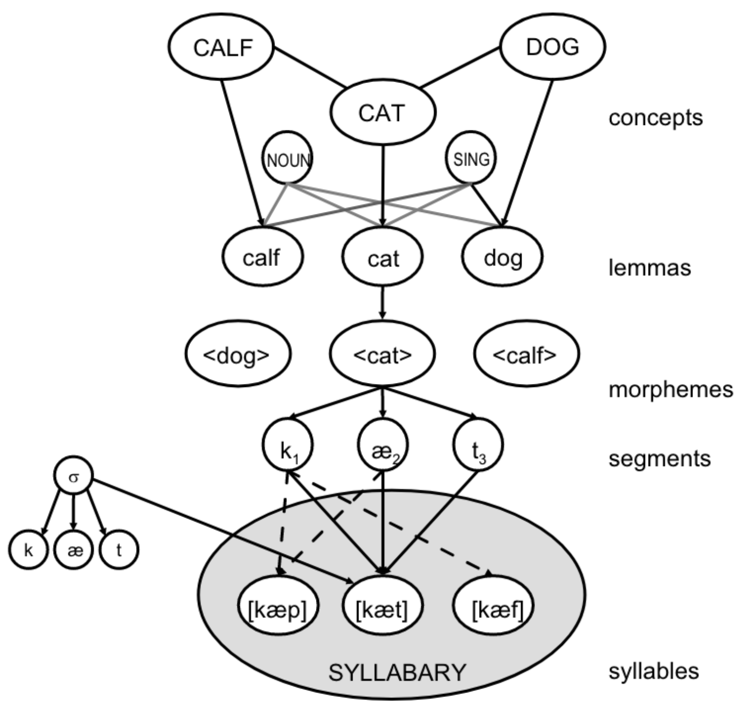


Figure 2.2: The model WEAVER++ (Levelt et al., 1999), the picture is from Dell (2014)

This model emphasizes accurate word selection and production in speech, as we have far more correct productions compared to speech errors (Levelt et al., 1999). Instead of analyzing speech errors, Levelt et al. (1999) stated that time latency is an effective measurement to observe the process of accurate word production. For example, in implicit priming (Meyer, 1991), a speaker would produce a word faster if they have been exposed to other phonologically similar words previously. The previous exposure facilitates the activation of phonological units and makes the production time latency shorter. Priming can also bring interference in the step of lexical selection and cause longer time latency in production. If a speaker is primed by similar semantic features, it increases the activation of other competitors in the step of selecting the target word. Then, it takes a longer time for the speaker to select the appropriate word.

Overall, these two models provide a general framework for lexical production. The interactive two-step model explains lexical production from the aspect of speech errors, while the WEAVER++ model focuses more on accurate speech production. Therefore, this model is more prevalent to be applied to explain common language effects, such as implicit priming, as I mentioned above.

The interactive two-step model does not treat lexical selection as a discrete step in production, but it is emphasized as an important step in Levelt et al. (1999), and we can somewhat observe the process of selection in language use, such as using different names to describe the same object. There are many questions regarding lexical selection. For example, what kinds of words are easier to select, and why are they selected? What information or cues does a speaker rely on to select a word? To answer these questions, one way is to delve into a typical language phenomenon where we can clearly observe a change in lexical selection. In the next section, I discuss a phenomenon called lexical entrainment. In lexical entrainment, a speaker tends to adjust their word selection under the influence of the person they are talking to. By investigating this behavior, we would have the opportunity to examine how people manage and select a word from all candidates, and what factors determine their word choices.

2.2 Lexical Entrainment

Lexical entrainment (Garrod and Anderson, 1987) is a phenomenon where people tend to reuse the words previously used by conversational partners. In lexical entrainment, a speaker may change their word choices during the conversation. For example, if we want to describe a computer, the most frequent word is *computer*. Other words, like *laptop*, can also be used to describe a portable computer. When two people are talking about a portable computer, even if the speaker has chosen to use *computer* at first, they are likely to change the word choice to *laptop* after hearing *laptop* many times from their conversational partner. Many studies investigate the mechanisms underlying lexical entrainment and the factors related to this behavior. In this section, I introduce two prevalent theories of cognitive mechanisms underlying lexical entrainment. Then, the role of a conversational partner in lexical entrainment is discussed. Recent studies have investigated how this behavior is affected by different partners. The partner's proficiency is stated as one of the important factors in lexical entrainment.

2.2.1 Theories of Cognitive Mechanisms

Interactive Priming

In the discussion of Levelt et al. (1999), the priming effect is argued to be evidence of lexical selection because time latency is affected by previous semantic and phonological exposures (Neely, 1976; Meyer, 1991, 1996). It was first observed in a lexical decision task (Meyer and Schvaneveldt, 1971). In the task, participants were asked to identify actual words (e.g. *doctor*) in contrast to nonsense words (e.g. *domter*). Meyer and Schvaneveldt (1971) found that exposure to semantically related words facilitated participants' responses. For example, participants identified *doctor* faster if they were previously exposed to *nurse* compared to being exposed to other unrelated words, such as *bread*. This indicates that the activation of the semantic features of *nurse* might also activate other words that connect to the same semantic features, which include *doctor*. Thus, *doctor* becomes easier for participants to activate and identify. Thus, in Pickering and Garrod (2004, 2006), priming is argued to be a fundamental mechanism by which speakers in a conversation choose the appropriate word, and an interactive alignment model was developed to explain the occurrence of lexical alignment (i.e., lexical entrainment). When a speaker first hears a word from a conversational partner, the semantic features of the word and its lemma representation are activated. Thus, the speaker is primed by the word heard from the conversational partner. When the speaker produces a word to express the same concept, the activated lemma representation is easier to access and produce compared to other competitors. In the model, the alignment is argued to be an automatic priming behavior because the speaker is inevitably primed by the words from the conversational partner. Then, once the speaker produces the same word, it is heard by the partner again and primes the partner to produce the same word. Through this process of repeatedly using the same word, two interlocutors align with each other's words and eventually converge to similar expressions. This interactive alignment helps to achieve communicative purposes. Moreover, the initial speaker in a conversation changes the role to a listener and then back to a speaker, so Pickering and Garrod (2004) also argued that the roles of two interlocutors in a conversation are equivalent. That means both interlocutors have the same amount of effect on each other to complete the alignment.

Conceptual Pacts

On the other hand, lexical entrainment is described as a more rational behavior with greater control involved compared to the theory from Pickering and Garrod (2004). Lexical entrainment was first defined by Garrod and Anderson (1987). They stated that people who are involved in a conversation should have a common conceptual and semantic system to

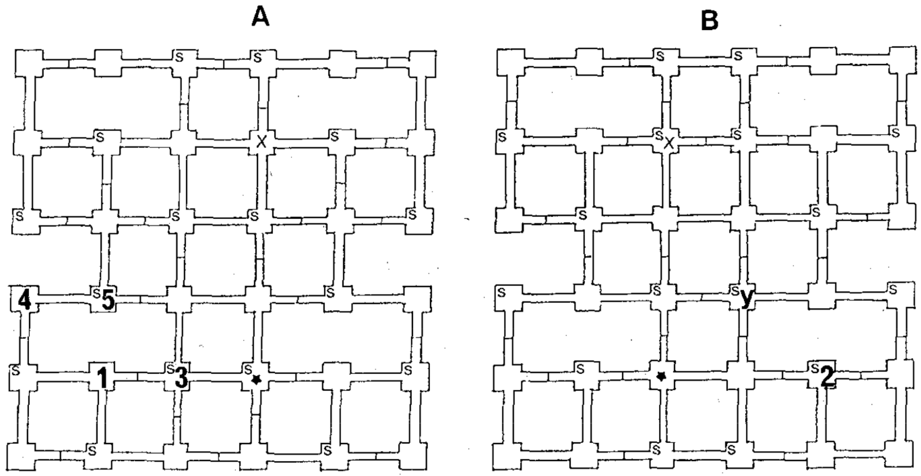


Figure 2.3: Examples of the two mazes for players A and B from Garrod and Anderson (1987)

achieve a successful conversation. Garrod and Anderson (1987) discussed the idea of coordination in accomplishing a communicative goal. People are interdependent in conversation and expect each other's behavior so that they can make a joint decision. In this study, researchers observed dialogues during the process of completing a maze game (Figure 2.3). In the maze game, the two players need to cooperate with each other to reach the goal positions, and their spontaneous dialogues were recorded during the process of figuring out the mazes. They found that participants in the same dialogue were likely to use the same types of descriptions, a phenomenon they named linguistic entrainment. They also found that the entrainment behavior is progressive; participants entrained to each other more after engaging in more dialogues. This was explained as the result of participants coordinating with each other and establishing agreements on concepts to solve the maze game.

Based on this idea, Brennan and Clark (1996) proposed the establishment of a conceptual pact that accounts for lexical entrainment. The conceptual pact is a temporary agreement about how people conceptualize an object, and the consequence of repeatedly using the utterances in the conceptual pact is lexical entrainment. Similar to Garrod and Anderson (1987), the conceptual pact is established jointly by a speaker and a conversational partner. What

Brennan and Clark (1996) added to this idea is that the conceptual pact allows for modifications during the process of establishment. During this process, hedges may occur if multiple alternative conceptualizations are possible and the speaker is not confident about the one they provide. Because of these hedges, the conceptual pact is not established immediately. The speaker and the partner need to frequently use the conceptual pact on the path to the agreement, and thus, the frequency of use matters. A stronger conceptual pact is established if people spend more time in this process and use the same words more frequently. Moreover, the conceptual pact should be accessible to both people in a conversation. That means, if the partner switches the role to a speaker who leads the conversation, they should also be able to use the established conceptual pact. Lastly, unlike Pickering and Garrod (2004), who argued that two interlocutors swap roles and have the same amount of effect on each other, Brennan and Clark (1996) stated that a speaker takes on the role of a leader to direct the establishment of conceptual pacts, and the other partners are followers. Conceptual pacts are also stated to be partner-specific. If the speaker begins to speak with a new partner, the old conceptual pacts established with the previous partner no longer apply in the new conversation. The speaker has to start over to achieve new agreements on conceptualization with the new partner and create new conceptual pacts. Therefore, who the partner is plays an important role in determining the entrainment behavior.

2.2.2 Conversational Partners

Besides studying the mechanisms in lexical entrainment, many studies also focus on the factors related to this behavior. For example, partner-specificity is one important feature mentioned in Brennan and Clark (1996). Thus, many studies manipulate conversational partners to explore their effect on lexical entrainment. If a conversational partner matters in affecting this behavior, the theory from Brennan and Clark (1996) regarding lexical entrainment may be supported.

In Branigan et al. (2011), the study examined the frequency of lexical entrainment in conversations between humans and computers. In the study, they conducted a picture matching and naming task (Figure 2.4) to measure entrainment behavior and compared the rate of entrainment in groups of conversations between humans and humans vs. humans and computers.

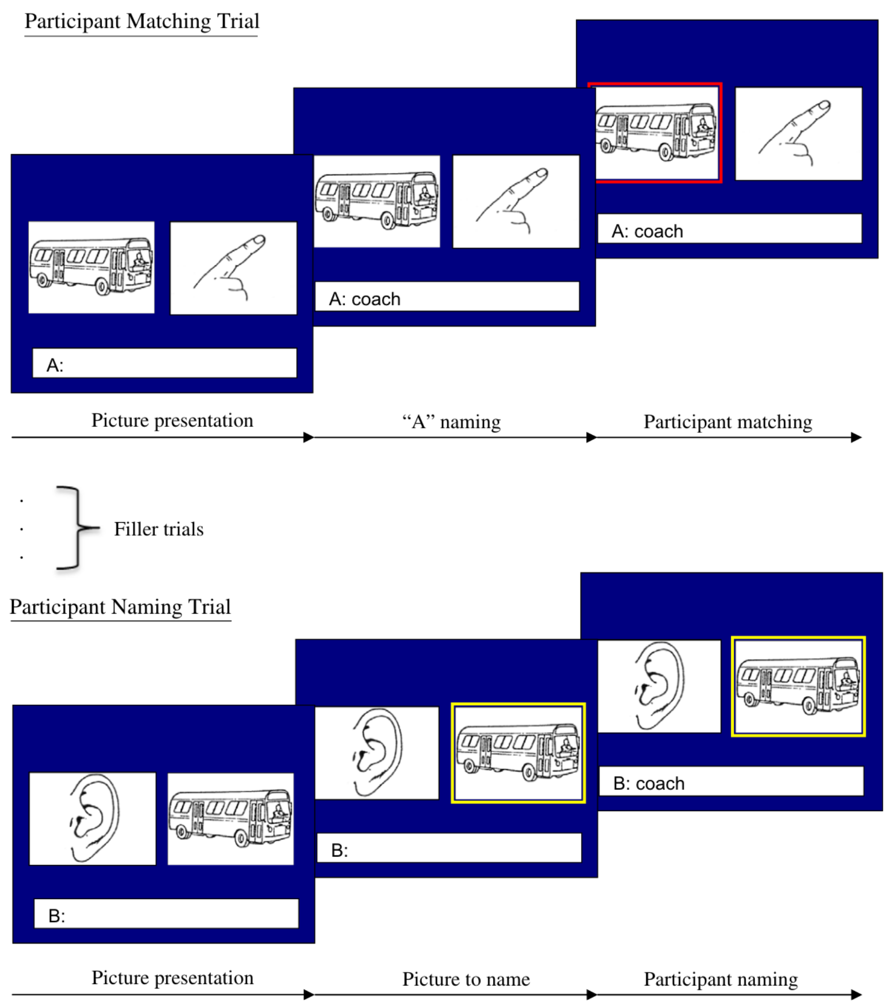


Figure 2.4: Screen shots of the trials picture naming and matching task in Branigan et al. (2011)

In the experiment, participants were informed that they would be paired with either a human partner or a computer partner. The picture matching and naming task began with a matching trial, where participants saw two pictures and needed to select the one that matched the word given by their partner (e.g., matching *coach* with the picture of a coach in Figure 2.4). The word provided was a less frequent name for the object. After several fillers, participants encountered the same picture again in a naming trial. In the naming trial, participants typed the name of the picture they had previously seen. If their response was the same as the word they received from their partner, lexical entrainment was thought to have happened. First, the authors observed lexical entrainment in both conditions. However, they found that participants tended to entrain more when they believed that they were talking to a computer than when they believed that they were talking to a human. Moreover, researchers told half of the participants that they were interacting with a less capable computer, while the other half were told they were interacting with an advanced computer. In this way, participants would have different beliefs about the linguistic capability of their computer partners. They found that participants were more likely to entrain when they believed they were interacting with a less capable computer. These results demonstrate that people exhibit different entrainment behaviors depending on their beliefs about their partner's ability to produce language. It also suggests that lexical entrainment is a more strategic behavior, as speakers adjust their behavior depending on who the partner is, which supports the theory presented by Brennan and Clark (1996). Speakers perceive their partners' ability to produce language and adjust their alignments to ensure successful communication. The empirical results also highlight the effect of a partner's linguistic capability on lexical entrainment, or the linguistic capability from the speaker's perspective. If the speakers perceive that their partners have difficulties understanding the conversation, they are more likely to entrain with the partner to ensure successful communication.

The effect of linguistic ability on lexical entrainment was also explored in conversations between adults and children, as well as between native speakers and non-native speakers (Cai et al., 2021; Suffill et al., 2021). Cai et al. (2021) studied the lexical entrainment frequency of adults when interacting with children, who are perceived as less capable of interpreting conversations due to their limited vocabulary. This study also conducted a picture matching and naming task in Mandarin and in a spoken format. Researchers found that adult participants entrained more when interacting with children than with adults. They also found more entrainment in the conversations when the partner was a non-native partner compared to conversations between two native speakers. These findings were consistent with the conclusion regarding the effect of linguistic ability presented in Branigan et al. (2011). The tendency of speakers to entrain to their partners' words increases when they interact with people who are less capable of producing a language, such as children and non-native speakers. This adaptability makes bilingual speakers an ideal group for studying variations in lexical entrainment.

Bilingual people are a population that can help us to investigate the variations happening in speakers because they employ different strategies to manage multiple languages. They are also very diverse even within groups. Understanding bilinguals and their language use can provide us with an opportunity to investigate the diversity of lexical use. In the next section, I introduce bilingual lexical production and bilinguals' performance in lexical entrainment.

2.3 Bilingual Lexical Production

To understand bilingual lexical production, the first question we need to consider is whether the lexical production process for bilinguals is the same as for monolinguals. As I discussed in previous sections, monolinguals access semantic features, activate related lemmas, and then select the target word with the activation of its phonological components. Because bilinguals are able to control two languages, an early question regarding bilingual lexical production is whether it is language-specific or non-specific. Language-specific selection means that bilinguals treat two languages as two independent systems, and the process of lexical production for each language is the same as that of monolinguals. In this theory, bilinguals' two languages are assumed not to affect each other. Thus, during the lexical access process, only competitors in the target language are activated and compete for selection. On the other hand, bilingual lexical production is argued to be language non-specific, meaning that both languages are involved in this process (Costa et al., 1999; Kroll and Groot, 2005). For example, in the step of lexical access and selection, related lemmas in both the target and non-target languages are activated as competitors. Bilinguals must select the target word from competitors not only in their target language but also in their non-target language.

Evidence supporting the language non-specific theory includes the asymmetry of switching costs between L1 (the first language) and L2 (the second language). In Meuter and Allport (1999), researchers conducted an experiment in which participants, either L1 English speakers or L2 English speakers, were asked to name numerals in two languages indicated by different colors. They measured the response time for both switching trials, where the language of the target trials differed from the language used in the preceding trials, and non-switching trials, where the same language was used consistently. Firstly, they found that there is a language switching cost. Bilinguals responded more slowly when switching from one language to another, indicating that activating a language takes more time if another language was previously used. This result suggests that the two languages for bilinguals are not independent but influence each other. Moreover, they found that switching from L2 to L1 involved longer latencies than switching from L1 to L2, indicating that it is more costly for bilinguals to switch from a weaker language to a more dominant language. Similar results were found in Costa and Santesteban (2004), but they also observed that the asymmetry became smaller when the proficiency levels of the two languages were equally high. The asymmetry of switching costs is explained by the inhibitory control mechanism (Kroll et al., 2008; Hoshino and Thierry, 2011; Misra et al., 2012). In inhibitory control, bilinguals' two languages are co-activated, and bilingual speakers have to inhibit the non-target language to successfully produce the target language. In switching from L2 to L1, bilinguals put more effort into inhibiting their dominant language, usually their L1, so it also takes more effort for them to reactivate L1, hence the asymmetric switching cost occurs. The co-activation of two languages and the inhibitory control mechanism in language switching support the language non-specific theory. In the process of lexical selection for bilinguals, competitors include not only words from the target language but also from the non-target language. Thus, bilinguals must manage their languages and make appropriate selections.

Next, I introduce some studies about bilingual lexical entrainment. By studying bilingual lexical entrainment, we gain a more comprehensive understanding of lexical entrainment and lexical selection. It allows us to observe how various factors affect lexical selection in communication. Furthermore, we have the opportunity to look at bilingual lexical use and study how they manage two languages and employ strategies to help them achieve communicative goals in different languages.

2.3.1 Bilingual Lexical Entrainment

Bilingual as a Conversational Partner

In research on bilingual lexical entrainment, the effect of a partner's language proficiency has received most of the attention. As previously mentioned, language competence is a factor related to a speaker's lexical entrainment. Bilinguals who are not native speakers (i.e., L2 speakers) are often considered less competent to produce a language compared to native speakers, so many studies have examined the entrainment of native speakers when they interact with other non-native partners (Cai et al., 2021; Suffill et al., 2021). These studies found that native speakers tend to entrain more when interacting with non-native partners, who are assumed to have lower language proficiency.

A more thorough investigation of the lexical entrainment between native speakers and non-native partners was conducted in Ivanova et al. (2021). They stated three hypotheses regarding the entrainment behavior of native speakers when interacting with both native partners and non-native partners. The first hypothesis suggests that native speakers would entrain more when interacting with non-native partners compared to native partners. They also argued that more entrainment occurs when there is a lack of feedback on comprehension success from the partners. The third hypothesis states that the entrainment of native speakers is modulated by their perceptions of whether their partner is competent to complete a successful communication. They conducted two experiments to test the entrainment rates of native British English speakers and native Castilian Spanish speakers interacting with both native partners and non-native partners. A picture matching and naming task was conducted in person, where participants and their partners alternated between naming and matching in both experiments. Participants were either assigned to interact first with a native partner, followed by a non-native partner, or the sequence was reversed. The first experiment was conducted in English. In the first experiment, they manipulated whether communicative success feedback was provided to participants (e.g., letting participants know whether they were correct in matching names with pictures). The results of this experiment showed that participants performed more entrainment when interacting with a non-native partner, and the rate was consistently higher when no feedback was given. It supports the important role of conversational partners in native speakers' entrainment behavior. The second experiment was similar but conducted in Castilian Spanish, with an additional variable of whether participants were presented with grammar errors from the partner, to test if they entrain more with the occurrence of grammar errors. They observed a similar effect of non-native partners with no feedback on enhancing the entrainment of native speakers, but the occurrence of grammar errors did not increase the entrainment rate. They explained that native speakers might intentionally avoid using the same words as their partner for the purpose of correcting the grammar errors. Communicative success is not the only purpose of the conversation, but the speaker also wants to present correct language use to the non-native partner when grammar errors occur. With the desire to show the partner correct language use, the speaker performs less alignment to improve the non-native partner's ability to use the language.

Bilingual as a Speaker

While many studies have supported the role of conversational partners in lexical entrainment and demonstrated the effect of the partner's language capability, or proficiency, how lexical entrainment is affected from a speaker's perspective is less discussed. Findings of how a speaker's proficiency affects lexical entrainment are also mixed.

Figure 2.5: Example of a map for route-giving task from Suffill et al. (2021).

In Suffill et al. (2021), researchers examined the lexical entrainment behaviors of native English speakers compared to non-native speakers during interactions with both native and non-native partners. This study conducted a route-giving task to measure entrainment rates in natural conversations. Different from the picture matching and naming task where participants respond depending on the information trial by trail, the route-giving task lets participants engage in a natural conversation with a paired partner. In the task, participants and their paired partner were given a map (Figure 2.5) and asked to take turns directing each other and delivering a set of objects from one point on the map to another. The names of the objects were disfavored names, so researchers observed whether participants used the disfavored names they heard from the partner in a more natural conversation. They found that both native speakers and non-native speakers entrained more when interacting with non-native partners. It supports previous conclusions on the effect of the partner's proficiency (Branigan et al., 2011; Cai et al., 2021). Furthermore, they argued that because native speakers and non-native speakers exhibited similar enhanced entrainment rates with non-native partners, non-native speakers are equally sensitive to the partner's language proficiency. However, their own proficiency appeared to have much less impact on lexical entrainment.

Zhang and Nicol (2022) also got a similar conclusion that a speaker's proficiency has a lesser impact on their lexical entrainment. They tested lexical entrainment rates of Mandarin-English bilinguals across different proficiency levels. They grouped participants into lower-intermediate and higher-intermediate groups based on their proficiency in an English as a foreign language class. They examined the participants' entrainment rates while varying the language proficiency of their conversational partners (native speakers vs. L2 learners). The findings revealed that Mandarin-English bilinguals demonstrated a higher degree of lexical entrainment when they believed they were talking to a native speaker as opposed to L2 learners. This result again supported the essential role of the partner's proficiency in entrainment behavior. The authors suggested that non-native speakers might be actively engaged in the language learning process, which facilitates entrainment. However, when comparing the lower-intermediate and higher-intermediate groups, no significant difference was observed in their entrainment based on the speakers' proficiency levels. Therefore, they concluded that a speaker's language proficiency has a relatively smaller influence on their entrainment behavior.

However, several recent studies, such as Ivanova et al. (2025) and our study that I will present in detail in Chapter 3, found a different result of how a speaker's proficiency affects lexical entrainment. In Ivanova et al. (2025), the authors conducted two picture matching and naming experiments with Spanish-English bilinguals at different levels of Spanish proficiency.

In the first experiment, they asked participants to complete a picture matching and naming task with a confederate who was highly proficient in Spanish. In the matching task, the participant took turns with the confederate to select pictures based on the names produced by the confederate, who used dispreferred names; in the naming task, the participant produced the names and the confederate selected the pictures. The rate of re-using dispreferred names was calculated as the entrainment rate. In this experiment, the authors found that bilinguals with low proficiency entrained more.

The mixed results of bilingual lexical entrainment may be due to many reasons. Costa et al. (2008) discussed several hypotheses on whether bilinguals, especially L2 speakers, tend to entrain or not. Unlike the model from Pickering and Garrod (2004), which suggests that speakers entrain to their partners' words primarily due to interactive alignment, Costa et al. (2008) proposed that there might be other non-automatic mechanisms or purposes involved in bilingual lexical entrainment. These non-automatic mechanisms may reduce the dominance of automatic priming in determining a bilingual's entrainment in their L2 (Costa et al., 2008). For instance, L2 speakers might not entrain as much because they need to put more effort into monitoring their speech production, or they may lack sufficient knowledge about the language. On the other hand, with the purpose of learning a language, L2 speakers may use more words from their partners when talking to native speakers, as supported by Zhang and Nicol (2022).

Individual differences may also contribute to the mixed conclusions observed in bilingual lexical entrainment studies. In Tobar-Henríquez et al. (2020), the authors stated that the tendency to entrain varies across individuals. They compared the entrainment frequency of participants who are native British English speakers in their native language and found that some participants were consistently more likely to entrain while others were not. These findings highlight the importance of individual differences across participants. Speakers may have relatively robust preferences on whether to entrain or not. It appears to be more like a personal trait because the preferences were consistent even in the retest a week later.

Furthermore, the methodologies used to measure proficiency and entrainment rates can also relate to people's performance. For instance, Suffill et al. (2021) used the LEAP-Q, a self-reported questionnaire, to assess proficiency, while Zhang and Nicol (2022) categorized participants into two groups based on their different levels of proficiency in the EFL class. These methods may overlook individual differences within bilingual populations. The studies discussed above also apply various tasks to examine lexical entrainment. For example, a picture matching and naming task is widely used to observe the entrainment behavior of a single word, but tasks such as the route-giving task are also used to observe lexical entrainment in natural conversations (Suffill et al., 2021).

Overall, the models of lexical production provide a relatively clear framework for understanding the lexical production process. However, the stage of lexical selection is crucial to discuss, especially when we take into account the variations of communicative goals, the involvement of listeners, and individual differences in speakers. Lexical entrainment allows us to observe the change of lexical selection with the manipulation of different variables. In the discussion of mechanisms in lexical entrainment, it is difficult to reach a definite conclusion about how lexical entrainment occurs and distinguish between the two mechanisms, so it is necessary to develop a method to clearly define them.

In terms of lexical entrainment in different populations, bilingual people contain a large diversity in language use due to many factors, such as proficiency (Costa and Santesteban, 2004; Chaouch-Orozco et al., 2021) and language environment (Beatty-Martínez and Dussias, 2017; Gullifer et al., 2018; Beatty-Martínez et al., 2020). Thus, studying bilingual lexical entrainment not only sheds light on the factors involved in lexical selection but also illustrates how bilinguals navigate these factors to achieve successful communication. There are still many aspects that have not been well-studied in lexical production and bilingual language use. In the following chapters of this dissertation, I present a series of studies on bilingual lexical entrainment. These studies apply both empirical and computational methods to investigate the factors that influence bilinguals' entrainment behavior, as well as the underlying mechanisms that drive this process. Together, they contribute to a more comprehensive understanding of bilingual lexical production and how entrainment leads humans to achieve successful communication.

Chapter 3

Proficiency in Bilingual Lexical Entrainment

This chapter is based on the following paper: Song, Y., K. Canales, Y. Gu, J. Jin, J. Meng, Y. Zhang, J. F. Kroll, and G. Scontras (2023). Lexical Entrainment in Bilingual Language Use, in the proceedings of 45th Annual Meeting of the Cognitive Science Society. Extra results and modifications have been made for inclusion in this dissertation.

In research on factors influencing lexical entrainment, bilingual proficiency plays an important role in determining the entrainment behavior of speakers who talk to bilingual individuals. Speakers who talk to less proficient bilinguals tend to entrain more (Branigan et al., 2011; Suffill et al., 2021; Cai et al., 2021; Ivanova et al., 2021). Although the investigation of how proficiency affects bilinguals themselves in entrainment is still relatively new, there are some insights about its effect from Suffill et al. (2021) and Zhang and Nicol (2022). These studies discussed whether proficiency also affects bilingual word selection in lexical entrainment and argued that a speaker's language proficiency has a less effect on their entrainment behavior.

Nevertheless, bilingual populations who were raised in different language backgrounds, speak their languages at different levels of proficiency, and adapt their languages across different environments, develop individual differences in language use (Beatty-Martínez et al., 2020; Polinsky and Scontras, 2020). The individual differences were not the primary concern in previous research on bilingual lexical entrainment, but may result in different entrainment behaviors. Therefore, in the studies of this chapter, bilinguals' proficiencies in both languages were objectively measured and calculated as a relative proficiency/dominance to better present their diversity with respect to proficiency.

Moreover, language switching is a unique feature in bilingual language use. Research on language switching (Kroll et al., 2008; Hoshino and Thierry, 2011; Misra et al., 2012) demonstrates that both languages are co-activated when bilinguals speak, but one language is inhibited from being produced in the context of the other language. The inhibition of a dominant language—a language in which bilinguals are more proficient relative to the other—is more difficult and requires more effort to then re-activate it. This asymmetric cognitive cost in language switching may also influence lexical entrainment. I hypothesize that switching also makes bilinguals entrain differently. When switching from a less dominant language to a more dominant language, because the dominant language is more difficult to activate, bilinguals may rely more on entrainment for their ease of accessing the words they have encountered from their partners, resulting in a higher entrainment rate.

In this chapter, I introduce several experiments that explore the possible effects of proficiency and switching on bilingual lexical entrainment. We targeted bilinguals who are proficient in English, including both L1 and L2 English speakers, and investigated their lexical entrainment behavior and the factors affecting their entrainment rate. Specifically, we test whether speaker proficiency affects entrainment rates and whether switching between first and second languages yields different entrainment rates.

We conducted three experiments. Experiment 1 tested the English entrainment behavior of bilinguals who are proficient in English and another language. Experiment 2 tested the Mandarin entrainment behavior of bilinguals who are proficient in Mandarin (their L1) and English (their L2). Experiment 3 added conditions of language switching, either from English to Mandarin or from Mandarin to English, to test whether language switching affects entrainment behavior in bilinguals. Participants in three experiments completed picture-matching and naming tasks to measure entrainment rate. We measured proficiency using subjective self-reports from a demographics questionnaire and a more objective verbal-fluency task that elicited names for category members.

3.1 Experiment 1: Entrainment in English

Experiment 1 investigated the entrainment rate in English of bilinguals who are proficient in English. The experiment was conducted online.

3.1.1 Method

Participants

We used our university participant pool and social media to recruit 100 bilingual individuals who are proficient in English (29 females, 70 males, and 1 unspecified gender). The average age of participants is 27, ranging from 18 to 35. After data cleaning, 67 participants were included in our analyses. These participants include 42 L1 English speakers and 25 L2 English speakers. The majority of participants are Spanish-English bilinguals (44), followed by Chinese-English bilinguals (14), as well as a small number of French (3), Cantonese (2), Vietnamese (1), Japanese (1), Korean (1), and Armenian (1) speakers. All participants were in the U.S. at the time of testing.

Object	Object in Mandarin	English	Mandarin
bathtub (75.00)	浴缸 (97.92)	tub (55.77)	澡盆 (14.58)
bike (88.46)	自行车 (95.83)	bicycle (75.00)	单车 (41.67)
boat (82.69)	帆船 (81.25)	sailboat (32.69)	船 (25.00)
bomb (98.08)	炸弹 (75.00)	explosive (25.00)	炸药 (10.42)
bucket (94.23)	水桶 (64.58)	pail (28.85)	桶 (52.08)
bus (88.46)	校车 (45.83)	school bus (55.77)	巴士 (35.42)
chicken (88.46)	鸡 (75.00)	hen (34.62)	母鸡 (35.42)
crackers (90.38)	饼干 (91.67)	saltines (25.00)	苏打饼干 (22.92)
frog (98.08)	青蛙 (95.83)	toad (44.23)	蛤蟆 (29.17)
glasses (98.08)	眼镜 (83.33)	spectacles (13.46)	近视眼镜 (8.33)
gun (90.38)	手枪 (83.33)	pistol (36.54)	枪 (39.58)
handcuffs (78.85)	手铐 (95.83)	cuffs (34.62)	镣铐 (6.25)
jacket (96.15)	外套 (77.08)	sweater (30.77)	夹克 (50.00)
lipstick (98.08)	口红 (95.83)	makeup (42.31)	唇膏 (29.17)
needle (88.46)	针 (68.75)	pin (17.31)	针线 (22.92)
peach (92.31)	桃子 (68.75)	nectarine (13.46)	水蜜桃 (12.50)
pen (94.23)	圆珠笔 (75.00)	ballpoint (13.46)	笔 (50.00)
phone (75.00)	手机 (95.83)	cellphone (25.00)	智能手机 (16.67)
pillow (100.00)	枕头 (77.08)	cushion (30.77)	抱枕 (45.83)
rabbit (94.23)	兔子 (97.92)	bunny (75.00)	野兔 (29.17)
radio (86.54)	收音机 (72.92)	stereo (17.31)	电台 (10.42)
stairs (88.46)	楼梯 (85.42)	staircase (46.15)	台阶 (20.83)
swan (92.31)	天鹅 (85.42)	bird (36.54)	鹅 (25.00)
sword (100.00)	剑 (83.33)	dagger (15.38)	宝剑 (16.67)
tape (84.62)	磁带 (87.50)	cassette (67.31)	录音带 (25.00)
utensils (84.62)	餐具 (91.67)	silverware (26.92)	刀叉 (29.17)

Note. Numbers in parentheses indicate % of use.

Table 3.1: English nouns used in experimental stimuli for Experiment 1 and Mandarin translations used in Experiment 2. Nouns in the Object column are the frequent names for the image, while nouns in the English and Mandarin columns are less-frequent names used as the entrainment targets. The percentage of use for English names was calculated from all responses provided by fifty-two English speakers in the English norming study, including both the first and second naming. The percentage of use for Mandarin names was based on responses from forty-eight Mandarin speakers in the norming study of Mandarin.

Norming We first performed a norming study to choose the stimuli for our task. Fifty-two English speakers were recruited to perform a spontaneous picture-naming task. In the task, participants encountered 69 images one at a time and typed a name for the object they saw, then they typed a second name. We used these data to measure the frequency of responses for each picture selected for the task. For our main task, we chose pictures that have at least two different words to describe them, and we used their second most frequent description as the entrainment target. Pictures and names were selected for experimental stimuli if the relative frequency of the first vs. second name provided was biased in favor of the first name, with the difference in proportional frequency between the two names ranging between 0.13 and 0.86. 26 stimuli (see Table 3.1) were chosen and used in the following picture matching and naming task.

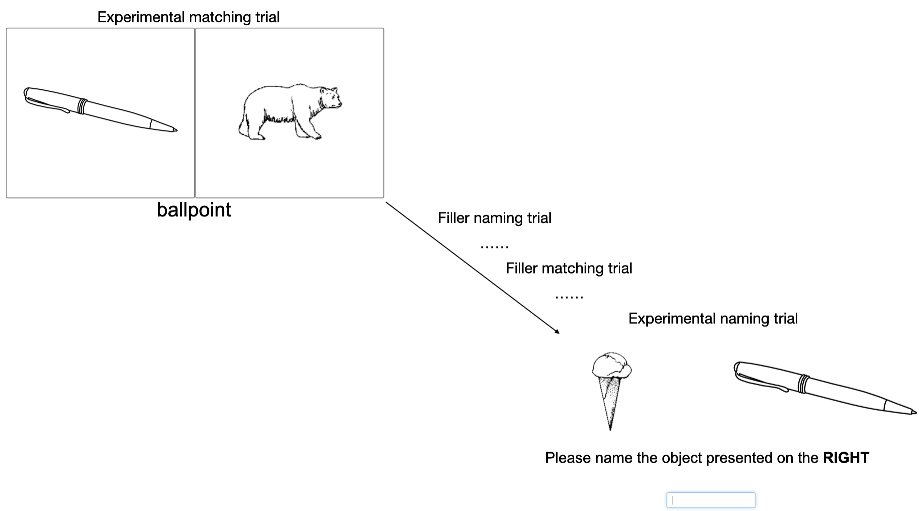


Figure 3.1: Example experimental matching and naming trails from the picture matching and naming tasks. In the example, participants saw *ballpoint* and needed to match the word with the target picture (i.e., the pen). Then, after two filler trials, they saw the target picture again in the experimental naming trial and were asked to name it. If they produced *ballpoint*, their response was coded as entrainment taking place.

Picture matching and naming task

We used a picture matching and naming task to measure the entrainment rate. The task was originally designed and used by Branigan et al. (2011). There were four trials as a sequence for each experimental item in the task, including two matching trials and two naming trials. The matching trial asked participants to match a written word with one of two pictures. The naming trial asked participants to type in the name of a pictured object. The first trial in the four-trial sequence was a matching trial where participants saw two pictures and a word on the screen (see Figure 3.1). The word in the matching trial was a disfavored word selected from the norming study. One of the pictures (i.e., the target picture) could be described with the given word, and another picture was an unrelated distractor. Participants needed to match the word with the picture by clicking the correct picture. Then, there was a filler naming trial and a filler matching trial. The filler items were selected from our norming data to be pictures that have only one high-frequency name (i.e., almost every participant in our norming study responded with the same word). In the fourth trial, participants completed an experimental naming trial. In this trial, participants encountered the target picture from the first trial in the sequence again, and they were asked to name it. If participants named the target picture with the word they had previously seen in the experimental matching trial, then we coded the response as entrainment. Participants completed a total of 104 trials (i.e., 26 four-trial sequences).

Procedure

The experiment was conducted online on Gorilla (https://gorilla.sc/). Participants joined a Zoom session, and a researcher introduced the tasks to them. Then, participants moved to a Zoom breakout room to complete the experiment individually. The experiment started with LEAP-Q to assess their language background. LEAP-Q (Marian et al., 2007) is a self-reported questionnaire to collect information about language backgrounds (e.g., acquisition of a language), self-assessment of proficiency in reading, speaking, and understanding, and language use in different environments, etc. Then, participants completed the picture matching and naming task. After completing the picture matching and naming task, participants proceeded to a verbal fluency task to assess their language dominance/proficiency. In the verbal fluency task, participants were given a series of category names (e.g., *vegetables*) and they spoke as many names for members of the category as possible in 30 seconds. The verbal fluency task was divided into two parts, with four category names in each part. Participants completed the first part in English, and completed the second part in a language that they felt most comfortable using other than English. Participants were instructed that the language should be either their native language (L1) or their second language (L2). Participants responded to different categories in each language (four chosen from a total of eight), but which categories they responded to in which language was counterbalanced.

3.1.2 Results

We transcribed the data from the verbal fluency task and excluded participants who completed the task incorrectly (i.e., did not follow the instructions or missed more than two categories in one session) or provided invalid responses. After the data cleaning, 67 participants were included in the analysis. For the analysis, we calculated language dominance both from the self-reported proficiency questions (LEAP-Q) and the results of the verbal fluency task (VF). In the questionnaire, participants were asked to assess their proficiency in reading, understanding, and speaking in English and their other language. For each question, responses were coded on a scale from 0 to 10. To calculate mean proficiency, we first averaged the answers of the three questions for each language. Then, we subtracted the mean proficiency of the other language from the mean English proficiency to get our self-reported English dominance score (see Table 3.2 for details on participant LEAP-Q report).

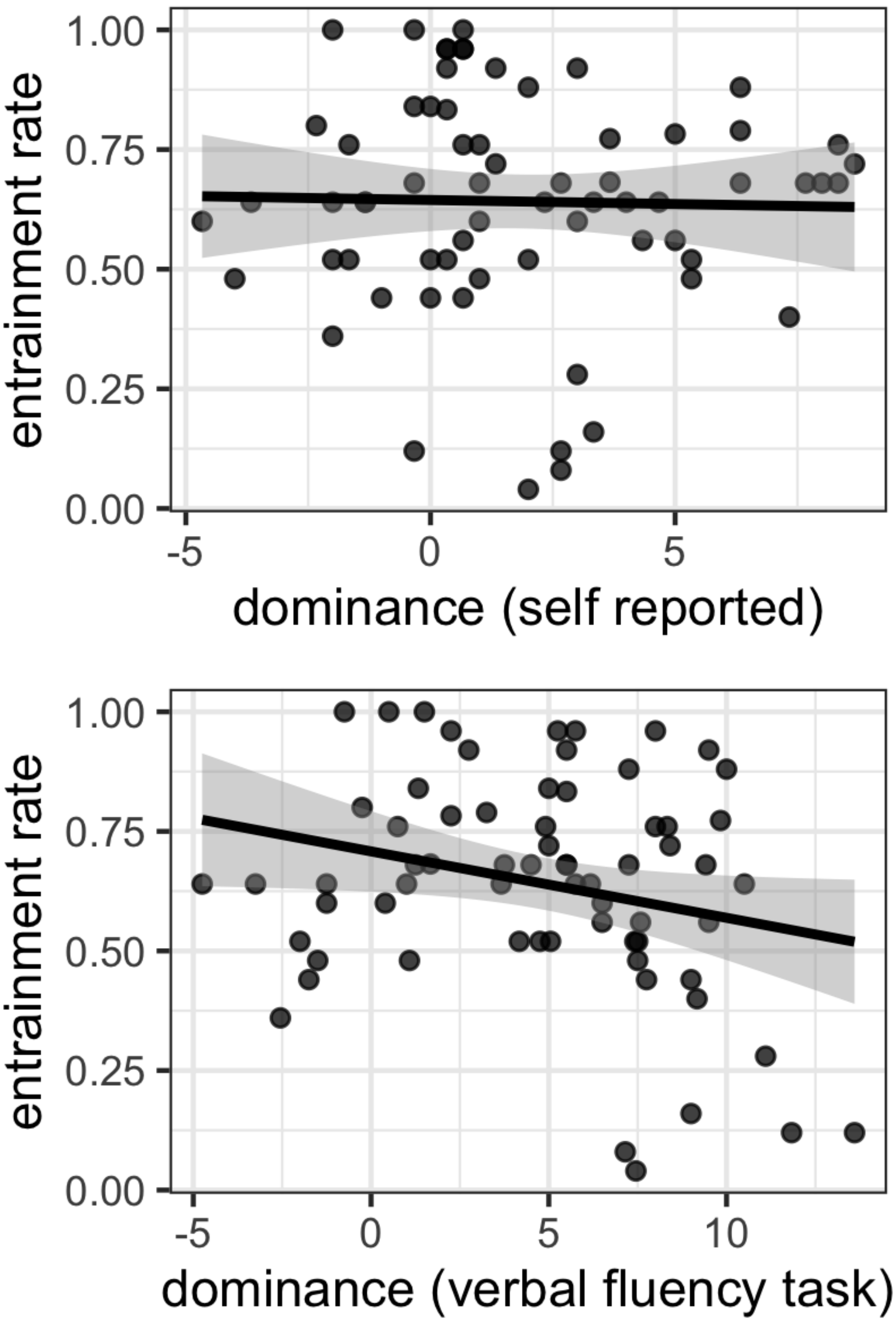


Figure 3.2: Results of Experiment 1 (entrainment in English): entrainment rate correlated with self-reported language dominance (top), and entrainment rate correlated with verbal fluency (VF) language dominance (bottom). Dominance was calculated as English dominance (values above 0 indicate that participants are more English dominant).

Table 3.2: Self-reported LEAP-Q for participants in Experiment 1

	English			Other Language		
Measures	M	SD	Range (0–10)	M	SD	Range (0–10)
Self-reported proficiency						
Understanding	8.91	1.39	3.00–10.00	7.46	2.91	0.00–10.00
Speaking	8.97	1.41	4.00–10.00	7.25	2.74	1.00–10.00
Reading	9.24	0.89	7.00–10.00	6.96	3.13	0.00–10.00
Age of Acquisition (years)						
Started learning	4.36	4.11	0.00–23.00	9.03	8.76	0.00–28.00
Attained fluency	8.51	5.82	0.00–33.00	13.91	8.78	1.00–32.00
Started reading	6.84	3.65	3.00–28.00	12.83	9.15	0.00–33.00
Attained reading fluency	10.45	5.34	5.00–33.00	15.89	8.49	3.00–33.00
Immersion duration (years)						
In a country	14.88	10.18	0.00–30.00	11.90	9.45	0.00–32.00
In a family	13.51	10.89	0.00–29.00	14.80	8.78	1.00–34.00
In a school	13.54	9.54	0.00–29.00	9.60	8.08	0.00–32.00

In the verbal fluency task, we calculated the mean of the number of responses across the four categories for each language. These means served as the language-specific proficiency score; we calculated English VF dominance by subtracting the other language's mean from the English mean. We used relative language proficiency, indicated as language dominance, instead of language-specific proficiency, as a way to investigate bilinguals who may be highly proficient in both languages but differ in dominance of language use. After calculating dominance scores, we compared the scores with participants' entrainment rates (see Figure 3.2). We ran logistic mixed-effects models predicting entrainment with respect to each language dominance measure, including random intercepts for items and participants. Although self-reported dominance is significantly correlated with VF dominance ($cor = 0.50$, $p < .001$), self-reported dominance is not a reliable predictor of entrainment behavior ($\beta = -0.02$, $p = 0.64$). However, VF dominance is a reliable predictor: English bilinguals who are more dominant/proficient in English tended to entrain less ($\beta = -0.08$, $p < 0.05$)(see Table 3.6 for details).

Spanish-English bilinguals

As the majority of our sample, the performance of Spanish-English bilinguals may direct the overall results. Therefore, we also analyzed the results for the subgroup of Spanish-English bilinguals from this experiment. 44 Spanish-English participants were included.

Spanish–English bilinguals showed a pattern similar to the full sample (Figure 3.3). A significant negative correlation was observed between English dominance (as measured by VF dominance) and the entrainment rate: participants who were more dominant/proficient in English entrained less ($\beta = -0.20$, $p < 0.01$). Consistent with the results using self-reported dominance, this relationship was not significant ($\beta = -0.09$, $p = 0.27$) for the same subgroup (Table 3.6).

3.1.3 Discussion

Our results suggest that speaker proficiency as operationalized by language dominance does relate to entrainment rates for English bilinguals, but only some measures of dominance are able to pick up on this relationship. The self-reported dominance scores showed no effect, in line with previous results using these scores (Suffill et al., 2021). However, the verbal-fluency measure showed that dominance negatively correlated with bilingual entrainment rates: bilinguals who are less dominant in English entrained more in English.

3.2 Experiment 2: Entrainment in Mandarin

Experiment 2 investigated the entrainment behavior of Mandarin-English bilinguals in their L1 (Mandarin). We expected that the effect of proficiency should extend beyond English.

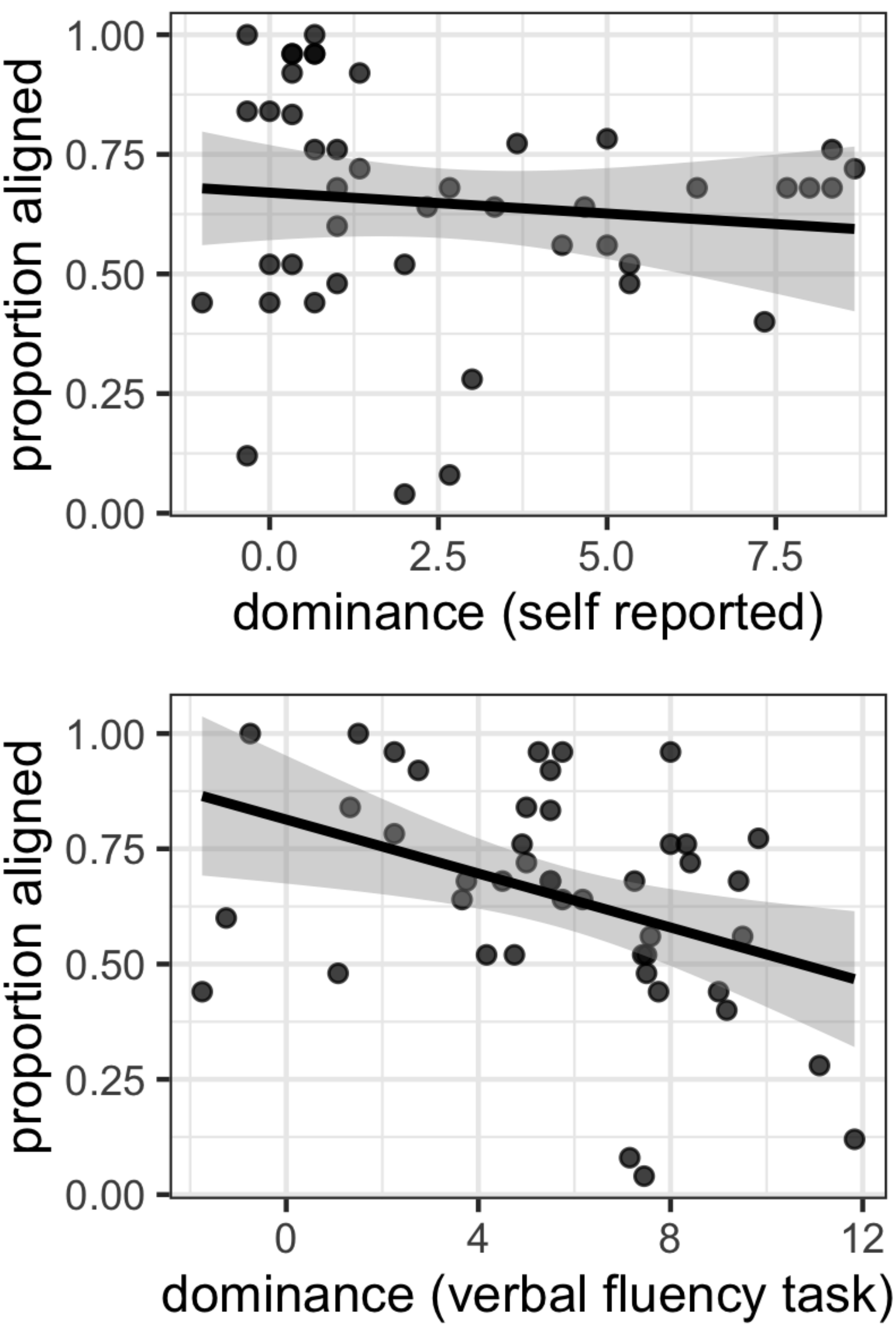


Figure 3.3: Results of Experiment 1 for only Spanish-English participants: entrainment rate correlated with self-reported language dominance (top), and entrainment rate correlated with verbal fluency (VF) language dominance (bottom).

3.2.1 Method

Twenty-seven Mandarin-English bilinguals were recruited for the study, with an average age of 23 (ranging from 18 to 34). These participants were international students currently studying in the U.S. The group consisted of 21 females and 6 males. All participants had Mandarin as their L1 and English as their L2. The tasks in Experiment 2 were Mandarin translations of the tasks in Experiment 1. We used the same pictures from Experiment 1 and performed a separate norming study in Mandarin (n = 48) to select the appropriate Mandarin words (cf. Table 3.1).

3.2.2 Results

As with Experiment 1, here we look at Mandarin dominance from both the LEAP-Q and the verbal-fluency task and their relationships with entrainment behavior. We excluded participants who had issues with the instructions of the verbal-fluency task. Data from twenty-one Mandarin-English bilinguals are included in the analysis (see Table 3.3 for details on participant LEAP-Q report). Using the same analyses (Table 3.6) as in Experiment 1, here we did not find significant effects of dominance/proficiency on the entrainment rate in Mandarin from either the self-reported dominance ($\beta = 0.13$, $p = 0.43$) or VF dominance ($\beta = 0.10$, $p = 0.09$) (see Figure 3.4).

Comparison of Experiments 1 and 2

In addition to looking at the effect of dominance within the participants of Experiment 2, we also compared the entrainment rate of Mandarin-English bilinguals in Mandarin (Experiment 2) with the entrainment rate of the 13 Mandarin-English bilinguals who specified Mandarin as their L1 who took part in the English entrainment task from Experiment 1 (see Figure

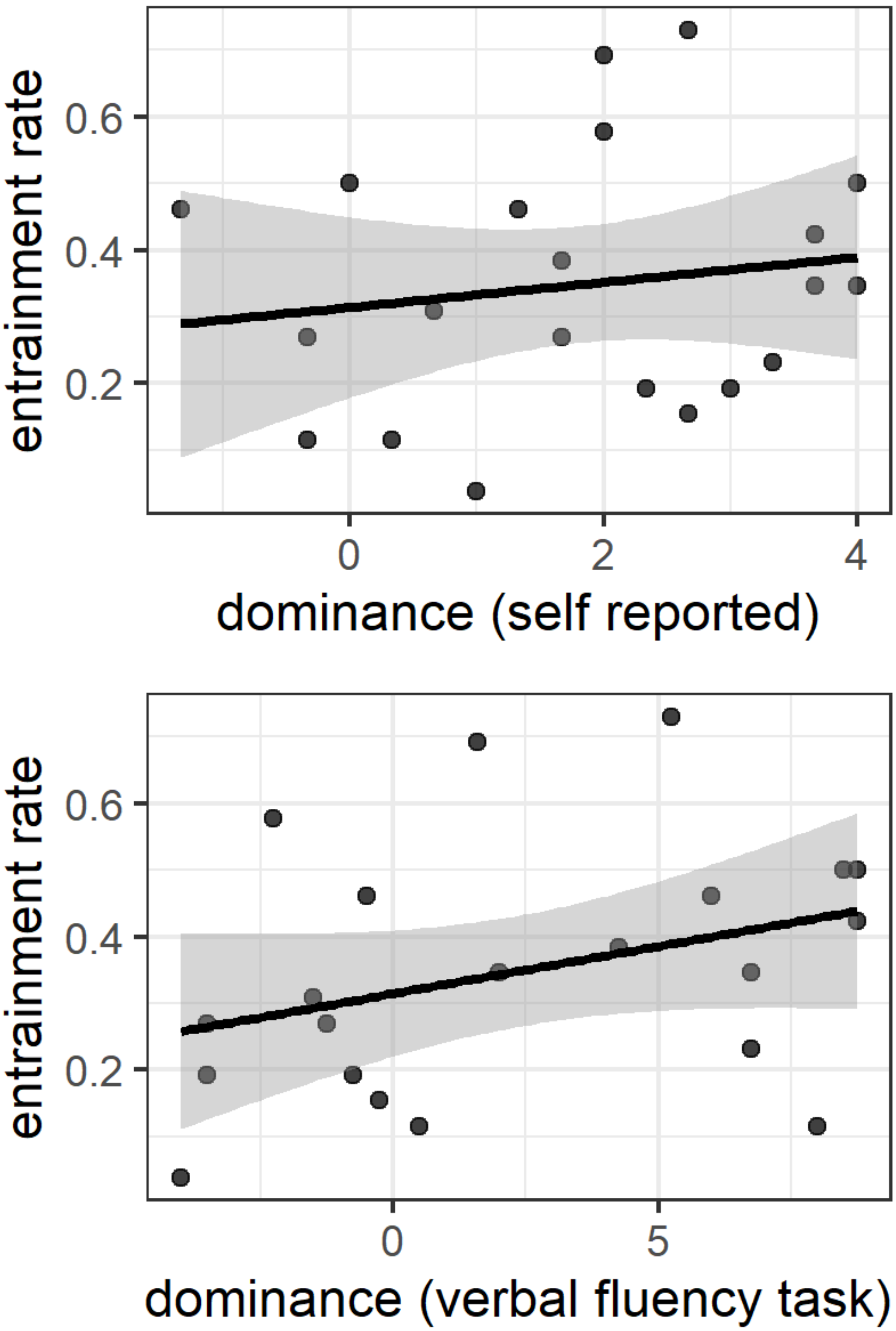


Figure 3.4: Results of Experiment 2 (entrainment in Mandarin): entrainment rate correlated with self-reported language dominance (top), and entrainment rate correlated with verbal fluency (VF) language dominance (bottom). Dominance was calculated as Mandarin dominance (values above 0 indicate that participants are more Mandarin dominant).

Table 3.3: Self-reported LEAP-Q for participants in Experiment 2

	English			Mandarin		
Measures	M	SD	Range (0–10)	M	SD	Range (0–10)
Self-reported proficiency						
Understanding	7.71	1.68	4.00–10.00	9.62	0.74	8.00–10.00
Speaking	7.52	1.78	4.00–10.00	9.62	0.86	7.00–10.00
Reading	7.76	1.37	4.00–10.00	9.19	1.25	6.00–10.00
Age of Acquisition (years)						
Started learning	5.67	3.51	0.00–17.00	0.71	1.82	0.00–6.00
Attained fluency	13.52	5.53	4.00–27.00	4.52	1.44	3.00–8.00
Started reading	9.90	3.83	5.00–19.00	5.29	1.42	2.00–7.00
Attained reading fluency	16.65	4.48	6.00–27.00	7.95	2.04	4.00–12.00
Immersion duration (years)						
In a country	6.15	6.39	0.50–25.67	18.40	5.27	1.42–26.83
In a family	2.70	7.18	0.00–25.67	19.86	5.41	1.42–27.00
In a school	5.34	3.91	0.00–17.67	14.57	6.46	0.25–26.33

3.5).

In this comparison, the entrainment rate in Experiment 1 represents their entrainment in L2, and the entrainment rate in Experiment 2 represents their entrainment in their L1. Running a logistic mixed effects model (Table 3.6) predicting entrainment across the two experiments with random intercepts for items and participants, we found that Mandarin-English speakers entrain more in their L2 (English) than in L1 (Mandarin) ($\beta = -1.44$, $p < 0.01$).

3.2.3 Discussion

While we failed to replicate the VF dominance effect in Mandarin entrainment, the comparison of the Mandarin-English bilinguals in Experiments 1 and 2 shows that these bilinguals entrain more in their L2, which, for these speakers, is their less dominant language. While our norming studies served to create parallel materials across the two studies, there remains the possibility that the difference in materials across the two experiments (i.e., the specific

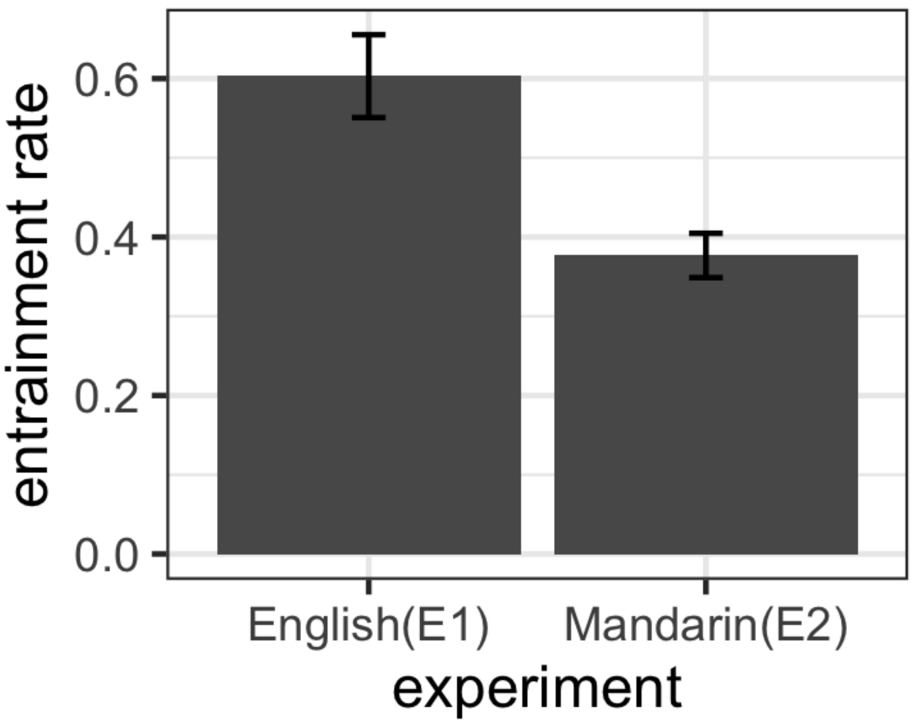


Figure 3.5: Average entrainment rate of Mandarin-English bilinguals in English (Experiment 1) and in Mandarin (Experiment 2). The mean of the entrainment rate in English is 0.60, and the mean of the entrainment rate in Mandarin is 0.35. Error bars represent bootstrapped confidence intervals drawn from 10,000 samples of the data.

English and Mandarin nouns chosen as labels for the images) might have driven this result.

Even more interesting than the comparison across the two experiments is the absence of a dominance effect in Mandarin. In the subgroup analysis of Spanish speakers in Experiment 1, a significant effect of proficiency on lexical entrainment was observed. However, no such effect was found among Mandarin–English speakers in Experiment 2. It is possible that the reduced number of participants in Experiment 2 relative to Experiment 1 (21 vs. 67 participants) led to power issues in the current analysis. Another possibility relates to differences between the two bilingual groups themselves. Experiment 1 primarily included Spanish–English bilinguals, who showed a clear effect of proficiency on entrainment behavior. In contrast, the number of Mandarin–English bilinguals included in Experiment 1 was relatively small, limiting their influence on the overall result. As a result, patterns of how Mandarin-English bilinguals behave can be overshadowed by the result of Spanish-English bilinguals. These findings suggest the potential influence of the specific language pair—or language dyad—on individual differences in bilingual lexical entrainment.

3.3 Experiment 3: Switching Languages

Experiment 3 focused on the effect of language switching on the entrainment behavior of Mandarin–English bilinguals. We hypothesized that bilinguals would entrain more when switching from their less dominant language (English) to the more dominant language (Mandarin), due to greater suppression of the dominant language when using the less dominant language.

3.3.1 Method

We recruited thirty-nine Mandarin-English bilinguals from a university participant pool and social media (with an average age of 21.3, ranging from 18 to 29). 28 of them are females, and 11 are males.

Experiment 3 added a manipulation of switching to investigate whether the entrainment behavior is affected by language switching. The experiment had a similar structure to the two previous experiments, but the picture matching and naming task was split into two parts. Participants completed the task in one of two language orders: either beginning in English and then switching to Mandarin, or beginning in Mandarin and then switching to English. Twenty-one participants were assigned to the English-to-Mandarin group, and eighteen to the Mandarin-to-English group. There are 14 trials in each session. Both sessions use the same stimuli but in different languages. These 14 stimuli (Table 3.4) were selected from stimuli in Experiment 1.

Object	**Object in Mandarin**	**English**	**Mandarin**
bike (88.46)	自行车 (95.83)	bicycle (75.00)	单车 (41.67)
bomb (98.08)	炸弹 (75.00)	explosive (25.00)	炸药 (10.42)
bucket (94.23)	水桶 (64.58)	pail (28.85)	桶 (52.08)
chicken (88.46)	鸡 (75.00)	hen (34.62)	母鸡 (35.42)
handcuffs (78.85)	手铐 (95.83)	cuffs (34.62)	镣铐 (6.25)
lipstick (98.08)	口红 (95.83)	makeup (42.31)	唇膏 (29.17)
peach (92.31)	桃子 (68.75)	nectarine (13.46)	水蜜桃 (12.50)
pen (94.23)	圆珠笔 (75.00)	ballpoint (13.46)	笔 (50.00)
phone (75.00)	手机 (95.83)	cellphone (25.00)	智能手机 (16.67)
pillow (100.00)	枕头 (77.08)	cushion (30.77)	抱枕 (45.83)
rabbit (94.23)	兔子 (97.92)	bunny (75.00)	野兔 (29.17)
swan (92.31)	天鹅 (85.42)	bird (36.54)	鹅 (25.00)
sword (100.00)	剑 (83.33)	dagger (15.38)	宝剑 (16.67)
tape (84.62)	磁带 (87.50)	cassette (67.31)	录音带 (25.00)

Note. Numbers in parentheses indicate % of use.

Table 3.4: Stimuli used in Experiment 3. English nouns were used in one session, and Mandarin translations were used in the other session. Nouns in the Object column are the frequent names for the image, while nouns in the English and Mandarin columns are less frequent names used as the entrainment targets.

3.3.2 Results

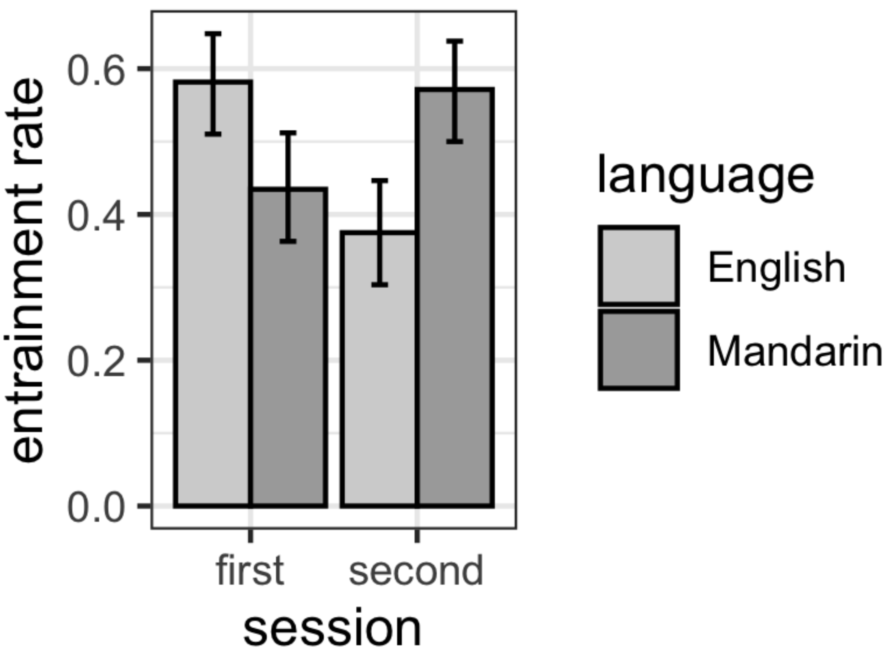


Figure 3.6: Results of Experiment 3 (switching languages): Light gray bars represent English, and dark gray bars represent Mandarin. Two bars on the left show the comparison of the entrainment rate between the first session in English (light gray)($mean = 0.58$) and the first session in Mandarin (dark gray)($mean = 0.43$). Two bars on the right show the comparison of the entrainment rate between the second session in English after switching from Mandarin (light gray) ($mean = 0.38$) and the second session in Mandarin after switching from English ($mean = 0.57$).

After data cleaning, we excluded ten participants who did not complete the VF task and three participants who reported Mandarin as their second language. Twenty-six participants were included in the analysis (see Table 3.5 for details on participant LEAP-Q report). We first compared the entrainment rates of English and Mandarin from the first session, before any language switching occurred. Entrainment rates in the first session represent the baseline entrainment behavior in English (for the English-to-Mandarin group) and in Mandarin (for the Mandarin-to-English group). We observed a trend whereby participants entrain more in English (their L2) than in Mandarin (their L1) ($\beta = -0.72$, $p = 0.15$), appearing to replicate the result from the comparison between Experiment 1 and Experiment

Table 3.5: Self-reported LEAP-Q for participants in Experiment 3

Measures	English M	English SD	English Range (0–10)	Mandarin M	Mandarin SD	Mandarin Range (0–10)
Self-reported proficiency						
Understanding	7.73	1.64	4.00–10.00	9.73	0.67	7.00–10.00
Speaking	7.04	1.51	4.00–9.00	9.73	0.53	8.00–10.00
Reading	8.15	1.22	5.00–10.00	9.77	0.43	9.00–10.00
Age of Acquisition (years)						
Started learning	6.73	2.20	3.00–12.00	0.69	0.97	0.00–3.00
Attained fluency	15.50	3.86	5.00–24.00	4.46	1.98	1.00–8.00
Started reading	9.12	3.25	3.00–19.00	4.42	1.90	0.00–8.00
Attained reading fluency	14.85	3.83	6.00–22.00	6.65	2.24	3.00–12.00
Immersion duration (years)						
In a country	3.92	3.00	0.02–11.00	16.07	6.25	1.17–22.58
In a family	0.97	2.06	0.00–7.08	17.42	6.90	1.00–28.00
In a school	3.98	2.96	0.00–11.00	13.15	6.27	0.00–22.33

2: we observed more entrainment in English than in Mandarin for these Mandarin-English bilinguals, though the result is not significant. The picture gets more interesting when we take into account the switching. We ran a logistic mixed effects model predicting entrainment by language, session order (switching), and their interaction, together with random intercepts by item and participant. The model revealed a main effect of session order (switching) ($\beta = -0.99$, $p < 0.05$), driven by the decrease in English entrainment from session 1 to session 2, as well as a significant interaction ($\beta = 1.62$, $p < 0.05$) in which Mandarin exhibits a different pattern, with higher entrainment in session 2. In the analyses for each language separately, the entrainment rate of English in the first session is significantly higher than that in the second session ($\beta = -1.06$, $p < 0.05$) after switching from Mandarin. In contrast, participants had a higher entrainment rate in Mandarin after switching from English compared to the entrainment rate in Mandarin in the first session, although we did not find a significant difference between the two Mandarin sessions ($\beta = 0.70$, $p = 0.20$) (see Table 3.6 for details). With the inclusion of proficiency, however, the results do not show the main effect of proficiency or switching for both English and Mandarin, nor the interaction of two variables (Figure 3.7).

3.3.3 Discussion

The results of this experiment provide insights into how the switching effect influences lexical entrainment. First, the results present a significant switching effect on entrainment in English: less entrainment occurs in English when participants switch from a more dominant language (Mandarin) to a less dominant one (English). A significant interaction effect was also observed, indicating that the effect of switching depends on the language. While switching led to a decrease in English entrainment, it resulted in an increase in Mandarin entrainment, which demonstrates opposite effects of switching across the two languages. Although the increase in Mandarin entrainment was not statistically significant relative to the baseline (Mandarin in the first session), the result still suggests that switching from a less dominant language to a more dominant one may lead to more entrainment behaviors. Mandarin-English speakers need to suppress their dominant language (Mandarin) more to produce English more easily in the first session, so the activation of Mandarin later becomes more difficult and takes more effort. As a result, participants may rely more on entrainment as a strategy to facilitate Mandarin production and ensure communicative success.

However, the significant decrease in English entrainment is difficult to explain. Research on language switching suggests that producing a more dominant language still involves inhibiting the less dominant language(Green, 1998; Meuter and Allport, 1999; Declerck and Philipp, 2015), so the production of the less dominant language later should also be affected by the inhibition. In our previous hypothesis, the occurrence of inhibition would lead to more entrainment behaviors. The English result, however, showed the opposite pattern. By observing the entrainment behaviors in English-to-Mandarin and Mandarin-to-English

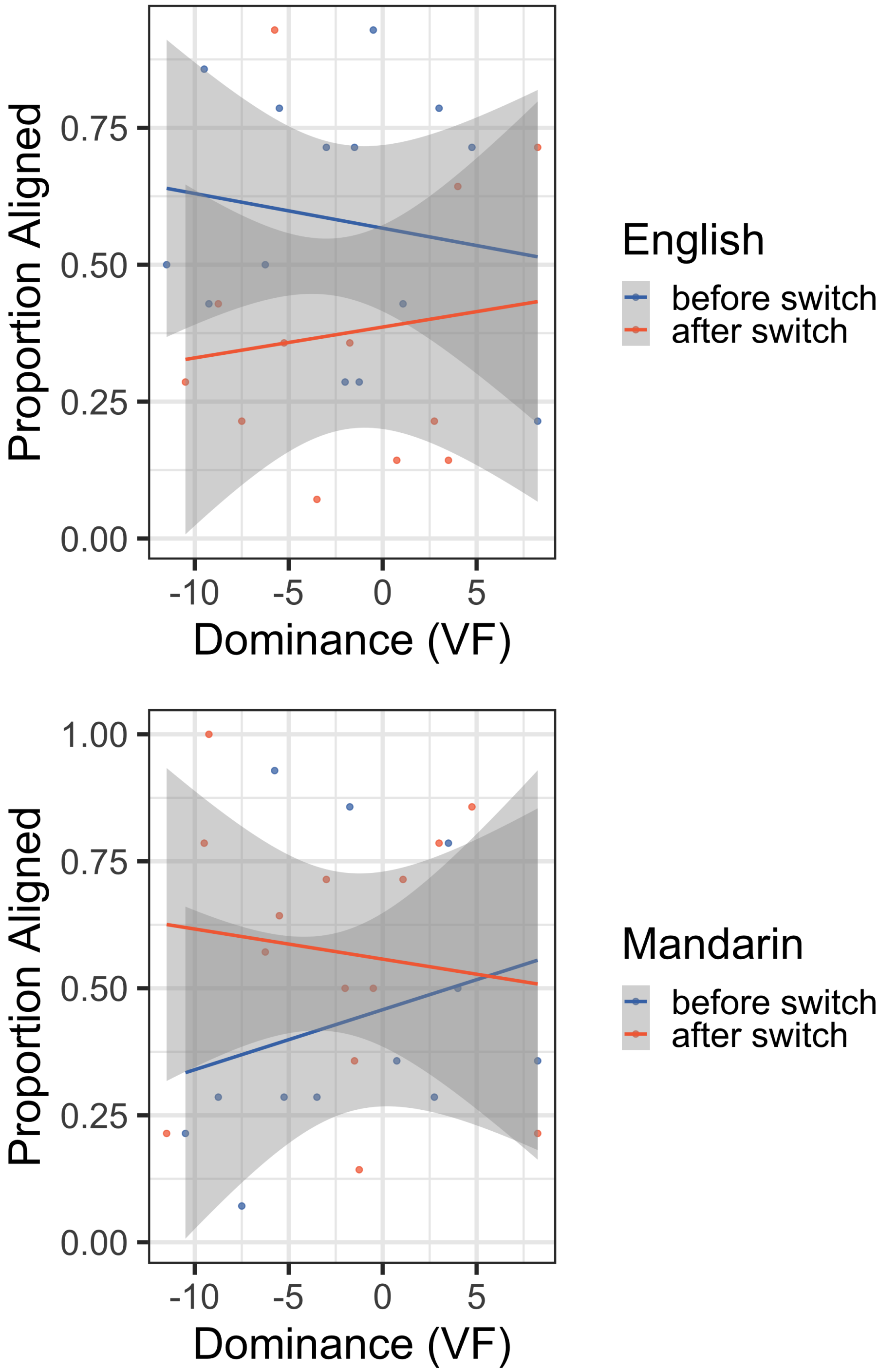


Figure 3.7: Results of proficiency and switching on entrainment in Experiment 3: Entrainment rate in English sessions correlated with switching and VF dominance (top), and entrainment rate in Mandarin sessions correlated with switching and VF dominance (bottom).

Experiment	**Model**	**Predictors**	β	**SE**	z	p
Experiment 1	all sample (67)	self-reported	-.02	.05	-.47	.64
		VF dominance	-.08	.04	-2.04	.04*
	Spanish-English (44)	self-reported	-.08	.08	-1.11	.27
		VF dominance	-.20	.07	-2.91	.004**
Experiment 2	all sample (21)	self-reported	.13	.17	0.79	.43
		VF dominance	.10	.06	1.68	.09
Exp.1 vs. Exp.2	Mandarin-English (13 in Exp.1) (21 in Exp.2)	Exp.1 vs Exp.2	-1.44	.48	-3.00	.003**
Experiment 3	all sample (26)	session order	-.99	.44	-2.25	.02*
		language	-.68	.48	-1.43	.15
		session*language	1.62	.81	2.00	.05*
	English sessions	session order	-1.06	.48	-2.20	.03*
	Mandarin sessions	session order	.70	.53	1.30	.20
	First session	Eng vs. Man	-.71	.51	-1.41	.16
	Second session	Eng vs. Man	1.06	.60	1.76	.08

Note. Session order was coded with the first session as the reference. Language was coded with English as the reference. β represents the estimate (regression coefficient) of the predictor in the model. * $p < .05$. ** $p < .01$

Table 3.6: Logistic mixed-effect model analyses of entrainment rate in relationship to different predictors in Experiment 1 (English), 2 (Mandarin), and 3 (Switching).

tasks, we found that the entrainment rates were similar across sessions within each task. Participants who began in English and entrained more tended to maintain this behavior and entrained more in Mandarin as well. Similarly, participants who began the task in Mandarin and entrained less also performed less entrainment in English in the second session. One possibility is that people become accustomed to the behavior they exhibited in the first session, so they continue it in the second session. This process may be independent of what language is used.

3.4 General Discussion

Over the course of Experiments 1 and 2, we investigated the role of speaker proficiency in driving bilingual lexical entrainment. Inspired by past work showing a reliable effect of listener proficiency but not speaker proficiency in entrainment, we explored whether different measures of proficiency—specifically those looking at the relative dominance of a bilingual's two languages—might reveal a clearer effect. Our results support the conclusion that language dominance/proficiency of speakers does play a role in lexical entrainment, but the way that proficiency gets measured matters for the results that are obtained. In Experiment 1, bilinguals who are more dominant in English according to our verbal-fluency measure—but not the LEAP-Q score—showed less entrainment. The LEAP-Q results replicate the absence of an effect found by Suffill et al. (2021). However, the verbal-fluency results indicate that English bilinguals who are less proficient in English, particularly L2 English speakers, may be more influenced by the lexical choices of their conversational partners. This result is surprising given the findings from previous studies (Suffill et al., 2021; Zhang and Nicol, 2022), and it runs counter to the reasoning from Suffill et al. (2021) that lower-proficiency speakers should entrain *less* than higher-proficiency speakers. However, it aligns with evidence from Ivanova et al. (2025), who also reported more entrainment among lower-proficient bilinguals. The result also highlights the central role of measurement in assessing language proficiency: more objective measures, such as verbal fluency, may provide a clearer window onto proficiency, at least those aspects of proficiency that are relevant to entrainment.

Although we did not observe similar results in Mandarin in Experiment 2, the comparison between Experiments 1 and 2 found that Mandarin-English bilinguals tended to entrain more in their L2 (English) than in their L1 (Mandarin)—a result consistent with the observation from Experiment 1 that lower proficiency leads to increased entrainment. It is important to note, however, that the different nouns tested in our English and Mandarin tasks may have led to the observed difference. Moreover, the non-significant dominance result in Experiment 2 may be due to power issues with insufficient participants (less than a third of the number tested in Expt 1).

In Experiment 3, we observed that the entrainment in English (L2) was lower after switching from Mandarin (L1). Conversely, we observed a trend in which Mandarin entrainment increased after switching from English. This result of Mandarin entrainment may support the idea that bilinguals tend to entrain more when switching from a less dominant language to a more dominant one due to the inhibition of the language. However, we also discussed an alternative explanation that the entrainment preference may be independent of the language used or the switching direction. Instead, it may reflect a consistent entrainment strategy across sessions, which means if a speaker engaged in more entrainment at first, they would maintain the behavior even when switching to the other language. The sample size of this experiment is relatively small. Therefore, additional evidence from future research is needed.

Taken together, our results suggest that bilinguals may rely on their conversational partner's lexical choices more when they are less proficient in the language. Bilinguals who are less proficient may be less confident in their own word choices and thus tend to rely more on the word choices of their partners. This finding contributes to research on listener proficiency in entrainment, demonstrating that the properties of both the speaker and the listener likely influence when entrainment occurs. These results suggest a possible strategic process involving cognitive control, as bilinguals rationally consider their proficiency level when entraining. However, there is no clear evidence regarding the engagement of automatic priming in this process, nor evidence for its absence. It is possible that proficiency also affects the automatic priming, such that less proficient bilinguals may be more easily primed due to weaker connections between semantic meanings and lexical options. In light of the question about the mechanisms of lexical entrainment, I turn to a computational modeling study in the next chapter.

Chapter 4

Computational Modeling of Mechanisms in Lexical Entrainment

This chapter mainly discusses the cognitive mechanisms in lexical entrainment. In the Background, I introduced two main theories of explaining the cognitive mechanisms underlying lexical entrainment. First, in Brennan and Clark (1996), lexical entrainment is explained as a process in which two interlocutors establish agreements on the concepts that people communicate about, which are called conceptual pacts. To establish conceptual pacts, a speaker has to pay attention to the partner's production and adjust their word use accordingly. In this process, the recruitment of cognitive control is emphasized because it helps to establish conceptual pacts and achieve communication goals. This theory is generally supported by many empirical studies. These studies focused on the listener-specific feature of lexical entrainment and found that people adjust their entrainment behavior when interacting with different conversational partners, including computers (Branigan et al., 2011), children (Cai et al., 2021), and non-native partners (Suffill et al., 2021; Zhang and Nicol, 2022; Ivanova et al., 2021, 2025). On the other hand, Pickering and Garrod (2004, 2006) raised priming as a fundamental mechanism in lexical entrainment. In their model, two interlocutors are primed by each other's words, resulting in alignment on linguistic representations. Although priming was discussed as one of the possible processes in lexical entrainment, its involvement is usually assumed without concrete empirical support. A study conducted by Abel and Babel (2017) examined the speech convergence across different levels of task difficulty and observed more convergences in easier tasks. In the study, participants, as a pair, were asked to collaborate on building identical LEGO constructions without seeing each other's constructions. Each member in the pair had half of the instructions for the constructions. The constructions were grouped into three levels of task difficulties. Task difficulty was also measured by completion time and error rate. As a result, higher linguistic similarity was observed in easy-level tasks, followed by the medium-level tasks, but no convergence happened in the hard-level tasks. Convergence also occurred in participant pairs who completed tasks in shorter times and fewer errors. Abel and Babel (2017) argued that people arrive at speech convergence with a lower cognitive load requirement in an easier task because convergence is an automatic process that requires the freeing up of resources to take place. However, the result might also be interpreted as evidence of deliberate, rational processes driving entrainment. In easy-level tasks, fewer cognitive resources are needed for completing the task, so more resources are available for speakers to evaluate lexical options, and then speakers deliberately align with the partner for more efficient communication. In general, there is a lack of research on the correlation between cognitive loads and automatic or rational processes.

A recent study provided some empirical evidence for the possibility of both mechanisms working together in lexical entrainment. In the study by Ivanova et al. (2025), the researchers conducted two picture matching and naming experiments with Spanish-English bilinguals and found that bilinguals with low proficiency entrained more in their first experiment. They argued that low-proficiency bilinguals use entrainment strategically, which supports the involvement of rational processes or cognitive control. In their second experiment, they attempted to eliminate the strategic motivation by removing the action of matching for participants. Instead of matching names from the confederate with pictures, participants only listened to the names without the intention of matching. In this way, participants were not motivated to pay attention to the names from the confederate and reuse them. In their results, they still observed significant entrainment behaviors in this second task, but with less entrainment compared to the first experiment. They argued that the entrainment rate in Experiment 2 was reduced because the strategic motivation that directs the rational process of lexical entrainment was removed, leaving only an automatic process, which resulted in a significant but reduced entrainment behavior.

While the conclusion of Ivanova et al. (2025) provides some support for the hypothesis that both an automatic process and a rational process work together to determine lexical entrainment, the occurrence of automatic processes is still assumed in their conclusion. They assumed that the automatic process is the process left after removing the strategic motivation, but it is also possible that people still rely on entrainment as a strategic process for communication, but to some extent reduce its use because of the lack of motivation. That means, the rational process is still the major mechanism working in entrainment observed in Experiment 2 of Ivanova et al. (2025), albeit to a reduced extent, given the reduced communicative motivation. Although the automatic process is always assumed because it is difficult for people to entirely block the influence of the information they are exposed to, it is non-trivial to be conclusively observed empirically, so its role in lexical entrainment requires additional support.

To answer these questions about mechanisms in lexical entrainment, in this chapter, I discuss my attempts to implement computational models to formalize the two mechanisms and investigating the possibilities of how they work in lexical entrainment.

4.1 Basic Model Framework

The idea of modeling the two mechanisms was developed from a recent computational model of language production (Futrell, 2023). This model proposes a framework for modeling language production that incorporates two mechanisms: automaticity and cognitive control of rewards. In this model, the two mechanisms are assumed to work together to accomplish word productions. With computational methods, the model successfully predicts the incremental process of language production under the constraints of various communication goals. It raises the possibility of a better understanding of the mechanisms involved in lexical entrainment: lexical entrainment should involve both priming, as an automatic process, and a rational process to help interlocutors intentionally manage their word choices. I discuss the basic model framework next and explain why it is a suitable foundation for modeling lexical entrainment.

4.1.1 RDC Language Production Model

Futrell (2023) developed a computational model of lexical production. The model is developed from a theory that humans do not always select the most probable behavior, but rather maximize the reward of the action under constraints from communicative goals or information in world states (Simon, 1955; Lewis et al., 2014), known as the Rate-Distortion Control Model (RDC model). For instance, when describing a little cat, although *cat* is the most frequent word to use, we may want to use *kitten* instead to emphasize the cat's age. In this case, *kitten* may bring a larger reward for the speaker, particularly when the speaker has a communicative goal of emphasizing that the age of the selected cat is young. In the model, a speaker's production is assumed to be determined by two mechanisms: (i) the automatic policy, and (ii) the cognitive control of rewards. Automatic policy means that the production action is determined by the most probable action, regardless of the communicative goal. For instance, when a speaker wants to describe a specific event (e.g., talking about a little cat), the word they select to say is an action x (e.g., saying *kitten*), and the probability of saying this word under the current state (s), regardless of the communicative goal, is defined as the automatic policy probability $P_0(x \mid s)$. There are different words that can be used, so there are also many alternative actions *($\bar{x}$)* (e.g., saying *cat*), with the probability $P_0(\bar{x} \mid s)$. The automatic policy is one factor that predicts the target action relative to other alternative actions, represented as $\ln \frac{P_0(x|s)}{P_0(\bar{x}|s)}$ in the model. In addition, cognitive control is also involved in predicting the reward of the action with respect to the goal of communication g. In this component, the speaker must consider the reward of using each word and select the one with the relatively higher reward, represented as ΔR in the model. ΔR means the reward difference between the use of two words. The calculation of rewards depends on how the reward is computationally defined for different communicative goals and the variables considered in affecting the reward. For example, the ΔR of selecting *kitten* or *cat* would relate to the discrepancy between the meanings of the two words in describing the specific target. There are various ways to operationalize meaning differences. For example, lexical meaning may represent how a listener interprets the word, where the meaning difference reflects the discrepancy in comprehension between two lexical alternatives from the listener's perspective. In this case, a listener's model capturing how words are interpreted in context can be developed to define the reward function (R). In developing the reward function, the communicative goal is one important constraint on the reward, because the target action may have a relatively low frequency of use but a high communicative reward for the specific goal. By combining two components and adjusting γ to balance the weight of the two parts in the model, the target action can be predicted as in Equation (4.1)[1].

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0(x \mid s)}{P_0(\bar{x} \mid s)} + \gamma \Delta R_g \right) \tag{4.1}$$

Futrell (2023) also discusses an incremental update to production based on the previous action. The model discussed here is simpler and more suitable for single-word production. It has successfully predicted the empirical results of lexical productions, such as those from Koranda et al. (2022), which argue that word production is determined by a trade-off between the accuracy of the word in communication and the ease of production, and from the semantic interference in a picture-word interference experiment (Lupker, 1979).

This model provides us insight into how people access and select words to achieve their communicative goals. From this model, the two mechanisms are assumed to work together to determine language production, and it provides a relatively accurate method of calculating the amount of involvement for the two mechanisms in production (e.g., via the parameter γ). Therefore, we can map the two theories of lexical entrainment into the two components in this model. First, automaticity in the interactive alignment is a crucial feature of the interactive alignment model proposed by Pickering and Garrod (2004) and is also considered as a component of directing language production, as noted by Futrell (2023). The priming in the interactive alignment may increase the probability of one word over another in the automatic component of the model, making the speaker more likely to select that word. Furthermore, cognitive control is considered as an important mechanism in entrainment based on the theory from Brennan and Clark (1996), which can also link to the cognitive control component in the model. Building on this framework, the parameter γ, which represents the weight of cognitive control of reward relative to the automatic process, can also be manipulated to reflect individual differences observed in empirical research, because speakers who behave differently in entrainment may rely on the two mechanisms to varying degrees. Therefore, in this study, I use computational methods to try to more accurately define the automatic and rational processes by implementing the RDC language production model and predicting

[1] This equation represents the predicted probability of the target action when there are two alternatives. The incremental updating process is not included.

observed entrainment behavior. By examining lexical entrainment from a computational perspective, I aim to more precisely investigate the ways in which the two mechanisms may work together. To implement the model, I explain the way of defining the two components in the model and updating them in the next section.

4.1.2 Policy Gradient

In the process of lexical entrainment, a speaker decides on a word by observing the word choice of their partner. Therefore, I assume that the speaker learns from the partner's word choice and applies the update from this learning process to each mechanism to determine the word use for the next step. Thus, a learning rule should be applied to the model to make the update. In this model, I apply the policy gradient method (Sutton and Barto, 2018) as the method for updating. Policy gradient is a reinforcement learning method that optimizes the policy by adjusting parameters in the policy to maximize the expected rewards. In policy gradient, the policy is denoted as $\pi(x \mid s, \theta)$, where x is the action, s is the state, and θ is the policy parameter. Policy here represents the probability of selecting an action x given a state s controlled by a policy parameter θ. The policy parameter θ determines how the policy behaves, and learning involves updating θ to improve the policy's performance. The update of policy happens on θ where

$$\theta_{t+1} = \theta_t + \alpha \nabla \hat{J}(\theta_t) \tag{4.2}$$

$J(\theta)$ represents the gradient of the performance measure with respect to θ_t (θ at time t). Then, the policy parameter of the next time θ_{t+1} is updated with θ_t and the estimate of the performance measure $\nabla \hat{J}(\theta_t)$ with a free parameter α.

In the general theory of reinforcement learning, an agent learns how good an action is in a particular state, which is quantified by a value function. The value function estimates the expected future rewards. However, in policy gradient, the policy is directly parameterized and optimized. Therefore, learning a value function becomes optional. The value function is primarily used to reduce variance, but it is not required for selecting actions. In this way, the function that approximates the learning process can be simpler compared to the parameterization of the value function. Policy gradient methods are well-applied for modeling cognitive sciences. For instance, Lai and Gershman (2021) applied a rate-distortion model combined with policy gradient updates to explain how cognitive resource constraints shape decision-making.

In the model from Lai and Gershman (2021), the authors argued that there is a trade-off between policy complexity and rewards. The optimized policy aims to maximize rewards with a constraint on policy complexity within the agent's cognitive capacity. This idea aligns with the Rate Distortion theory, where reward is maximized under some constraints in the environment. Therefore, they proposed an RDC model by applying policy gradient to incrementally adjust the policy complexity. They discussed the model's application across various decision-making phenomena. One example from Lai and Gershman (2021) is simulating stochasticity, a phenomenon where agents make difference choices in identical situations, which reflects the variability in decision-making. Lai and Gershman (2021) argued that agents, constrained by cognitive capacity, must be stochastic to achieve the optimal balance between reward and policy complexity. Thus, stochasticity can be modeled by their RDC model with the policy for estimating an action (x) (Equation 4.3).

$$\pi_\theta(x \mid s) \propto exp(\beta\theta_x \cdot \phi(s) + \log P(x)) \tag{4.3}$$

In the example of stochasticity, Lai and Gershman (2021) simulated the task from Collins and Frank (2012) where participants were shown a set of stimuli (states) and asked to select one of three actions. The set size (states) varied from 2 to 6 across groups, and each state within a group was associated with a different correct action. According to the hypothesis in Lai and Gershman (2021), the increasing cognitive load (larger set size), which should reduce policy complexity, will make the action selection more stochastic. Therefore, they modeled the probability of selecting stochastic actions through Equation 4.3. In this policy, θ denotes the adjustable policy parameter, and $\phi(s)$ is the set of state features (e.g., stimuli in the task), which represents the variations in different contexts for selecting the action. $P(x)$ represents the probability of the action x across all states. β is a free parameter similar to the γ in model (4.1) to balance the weight of the two components. In the task, an agent is assumed to learn from the action taken in response to the previous stimulus, so the policy should be updated based on the previous actions. Lai and Gershman (2021) defined the update rules for θ_x and $P(x)$ based on the basic framework of the policy gradient method in Equation (4.2) with different equations of performance measure. The update rule of $P(x)$ is

$$\Delta P(x) = \alpha_P(\pi_\theta(x \mid s) - P(x)) \tag{4.4}$$

where α_P is a free parameter to determine the learning rate on $P(x)$, and the performance measure is defined as the difference between the probability of action x (e.g., a stochastic action) under a specific state and the probability across all states. The equation for the update rule of θ_x is

$$\Delta\theta_x = \alpha_\theta \beta \delta (1 - \pi_\theta(x \mid s)) \tag{4.5}$$

In Equation (4.5), α_θ is the other free parameter of learning rate with respect to θ_x. The performance measure of θ_x contains a free parameter β, a prediction error measure δ, and $(1 - \pi_\theta(x \mid s))$ representing the difference of selecting the stochastic action (equal to 1) and the probability of the action in the specific state. δ is the prediction error of the critic $\hat{V}(s)$, which estimates the expected reward for the given state.

$$\delta = \beta - \log \frac{\pi_\theta(x \mid s)}{P(x)} - \hat{V}(s) \tag{4.6}$$

Lai and Gershman (2021) apply an actor-critic method here, which is a type of policy gradient method that allows the agent to learn the value function in addition to the policy, although the value function may be optional depending on the case. Their result successfully modeled the stochasticity, showing that more stochasticity occurred as the set size increased.

In my simulation of lexical entrainment, I follow the model proposed by Lai and Gershman (2021) and apply the RDC model with the update rules of policy gradient. Next, I explain how the model is implemented.

4.2 Model Implementation

I take my study of bilingual lexical entrainment in Chapter 3 as the scenario I aim to simulate. In the picture matching and naming task of my empirical study, entrainment behavior occurs when participants use the less frequent word they previously saw from the virtual partner. In this context, the less frequent word is the action x that participants aim to take, and the more frequent alternative word is the other possible action $\bar{x}$. In the scenario of naming and matching pictures, the image serves as the state s, acting as the environment for actions. Since the image for each named item is always the same for both the speaker and the partner, the state s is held constant and is assumed to have no effect on entrainment behavior in the model. As I previously introduced, the two theories of mechanisms in lexical entrainment can be mapped to the automatic policy and reward in the RDC model of language production. Therefore, the policy of selecting the target action ($\pi_\theta(x \mid s)$) comprises two components that map to the two mechanisms. Referring to Equation (4.3), the automatic priming process is defined as $\ln(\frac{P(x)}{P(\bar{x})})$, which represents the probability of the target action x relative to that of the alternative action $\bar{x}$ in general. This process is not affected by states or communicative goals. Then, the rational process (i.e., conceptual pacts) is defined as the difference between the policy parameter of the two actions $(\theta_x - \theta_{\bar{x}})$ with β to balance the weight of the two components. Since the state remains constant, the set of state features $\phi(s)$ is omitted. The model is presented below.

$$\pi_\theta(x \mid s) = \sigma\left(\ln \frac{P(x)}{P(\bar{x})} + \beta(\theta_x - \theta_{\bar{x}})\right) \tag{4.7}$$

To implement policy gradient updates, first, I need to clarify how the learning process occurs in entrainment behavior. Because entrainment behavior occurs in conversation, it is a process through which the speaker learns from the actions of the partner to update their own actions. Therefore, the speaker does not estimate the reward based on their own action, but rather observes and learns about the success in the matching trial from the action done by the partner. After observing the action of the partner, the learning rule of each component is applied to update the estimation of the target action.

Based on research questions of how the two mechanisms work in lexical entrainment, there are three possibilities for the update of the probability: (i) updating the automatic policy solely while keeping the reward the same; (ii) updating reward without any change on the automatic policy; and (iii) updating the reward in addition to the automatic policy, meaning that both mechanisms work together;

Update on Automatic Policy

In the automatic policy $(\ln \frac{P(x)}{P(\bar{x})})$, the probability of both the target word x and the alternative word $\bar{x}$ should be updated with the update rule of Equation (4.4). With the update of both actions, the new automatic policy should equal

$$\ln \frac{P^{new}(x)}{P^{new}(\bar{x})} = \ln \frac{P(x) + \Delta P(x)}{P(\bar{x}) + \Delta P(\bar{x})} \tag{4.8}$$

Because the automatic policy is the probability of an action across states, it relates to the general word frequency of the word. I use corpus word frequencies from Speer (2022) to calculate $P(x)$ and $P(\bar{x})$. In the corpus, zipf frequencies are log-transformed and scaled by adding 3. To recover the original word frequency, I subtract 3 from the zipf score and exponentiate the result. In $\Delta P(x)$, the result of $\pi_\theta(x \mid s)$ is also needed and is assumed to be the observations from the norming study in Chapter 3. $\pi_\theta(x \mid s)$ represents the probability of taking action x for a specific state s. Thus, the word frequencies of the target words from the norming results indicate the frequency of the action (i.e., selecting the target words). In the norming study, participants were asked to name each picture twice, but I used only the frequencies from the first naming to ensure that the data accurately reflected the initial word selections for the images. In the norming results, some target words were not used in the first naming. To prevent zero frequencies, I add a small constant $\epsilon = 0.001$ to $\pi_\theta(x \mid s)$ to ensure that no probability is equal to zero and that all words are included. Then, the updated probabilities of both words are calculated with the update rule applied.

Update on Policy of Reward

Before updating the policy of reward ($\beta(\theta_x - \theta_{\bar{x}})$), I first need to compute the result of $\theta_x - \theta_{\bar{x}}$, which is the difference of policy parameters between the target word and the alternative word. This value can be derived from Equation (4.7). In this equation, $\pi_\theta(x \mid s)$ before updating is known from the norming observations, and $P(x)$ before updating is defined as the general word frequency calculated from corpus data. Using these values, $\theta_x - \theta_{\bar{x}}$ is derived as follows:

$$\theta_x - \theta_{\bar{x}} = \frac{\ln \frac{\pi_\theta(x|s)}{\pi_\theta(\bar{x}|s)} - \ln \frac{P(x)}{P(\bar{x})}}{\beta} \tag{4.9}$$

Then, I apply the learning rule in Equation (4.5) and Equation (4.6) to update the policy of reward. In the model of Lai and Gershman (2021), they include the critic $\hat{V}(x)$ for the case of simulating stochasticity. In the case of modeling entrainment, it occurs in a scenario where speakers encounter the target word only once, which means only a one-time update occurs. Because the variance prior to this update is unclear, I assume that the critic $\hat{V}(x)$ is equal to 0 and therefore omit it from the model. I instead use the policy gradient method without the critic. δ is simplified to

$$\delta = \beta - \ln \frac{\pi_\theta(x \mid s)}{P(x)} \tag{4.10}$$

Then the learning rule is applied for both the target word and the alternative word with the assumptions of one observation of the target word and zero observation of the alternative word. The update finally equals

$$\Delta\theta_x - \Delta\theta_{\bar{x}} = 2\alpha_\theta \beta \delta (1 - \pi_\theta(x \mid s)) \tag{4.11}$$

4.2.1 Results

I take the empirical result in Chapter 3 to evaluate the model fitting. The results in Chapter 3 are the entrainment rate across participants, so I re-analyze the results by examining the entrainment rate across stimuli. Then, I look at the correlation between the frequencies of spontaneous use and the entrainment rates. The frequencies of spontaneous use refer to the frequencies of the target words in the first instance of naming in the norming study. The

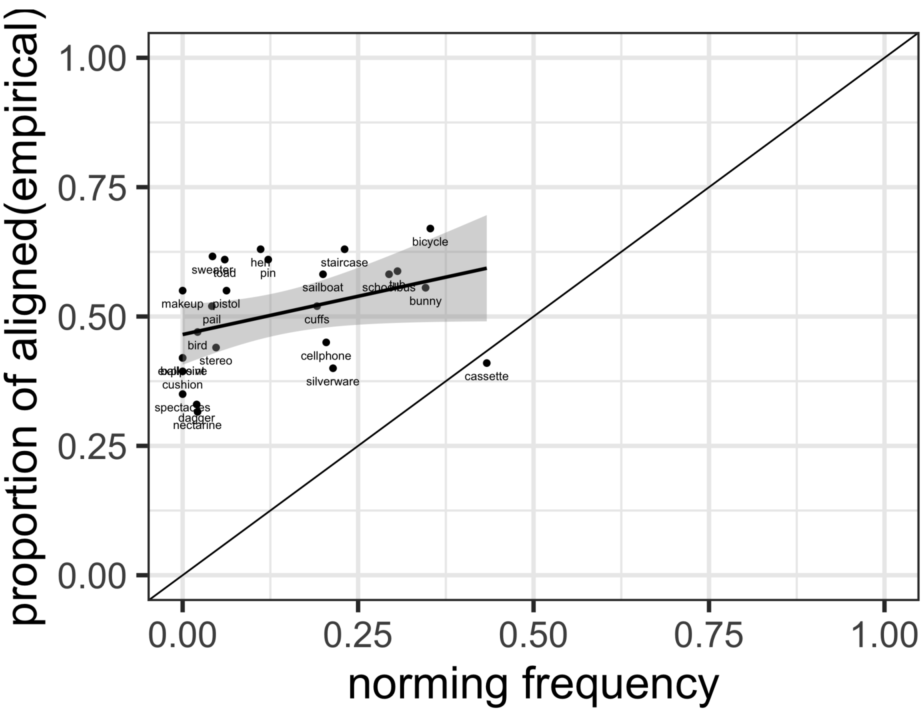


Figure 4.1: Empirical result of entrainment rate across stimuli.

result is presented in Figure 4.1. From this result, I observe that for stimuli with a larger frequency of the target word in spontaneous use—the frequency of the target word is closer to the frequency of the alternative word, and they get less boost from entrainment. This result can be explained as an inverse frequency effect in syntactic priming (Scheepers, 2003; Kaschak et al., 2011), where the less frequent construction gets a stronger priming effect. Therefore, in the comparison, a good model should fit the empirical results and present the inverse frequency effect.

First of all, I set up all free parameters (β, α_P, and α_θ) equal to 1. When the update only happens on the automatic policy, the model does not fit the empirical results ($r = 0.144$, with $p = 0.49$, Figure 4.2, top), and the update does not follow the trend observed in empirical results, where the target words with smaller frequency would be boosted more in entrainment (Figure 4.2, bottom).

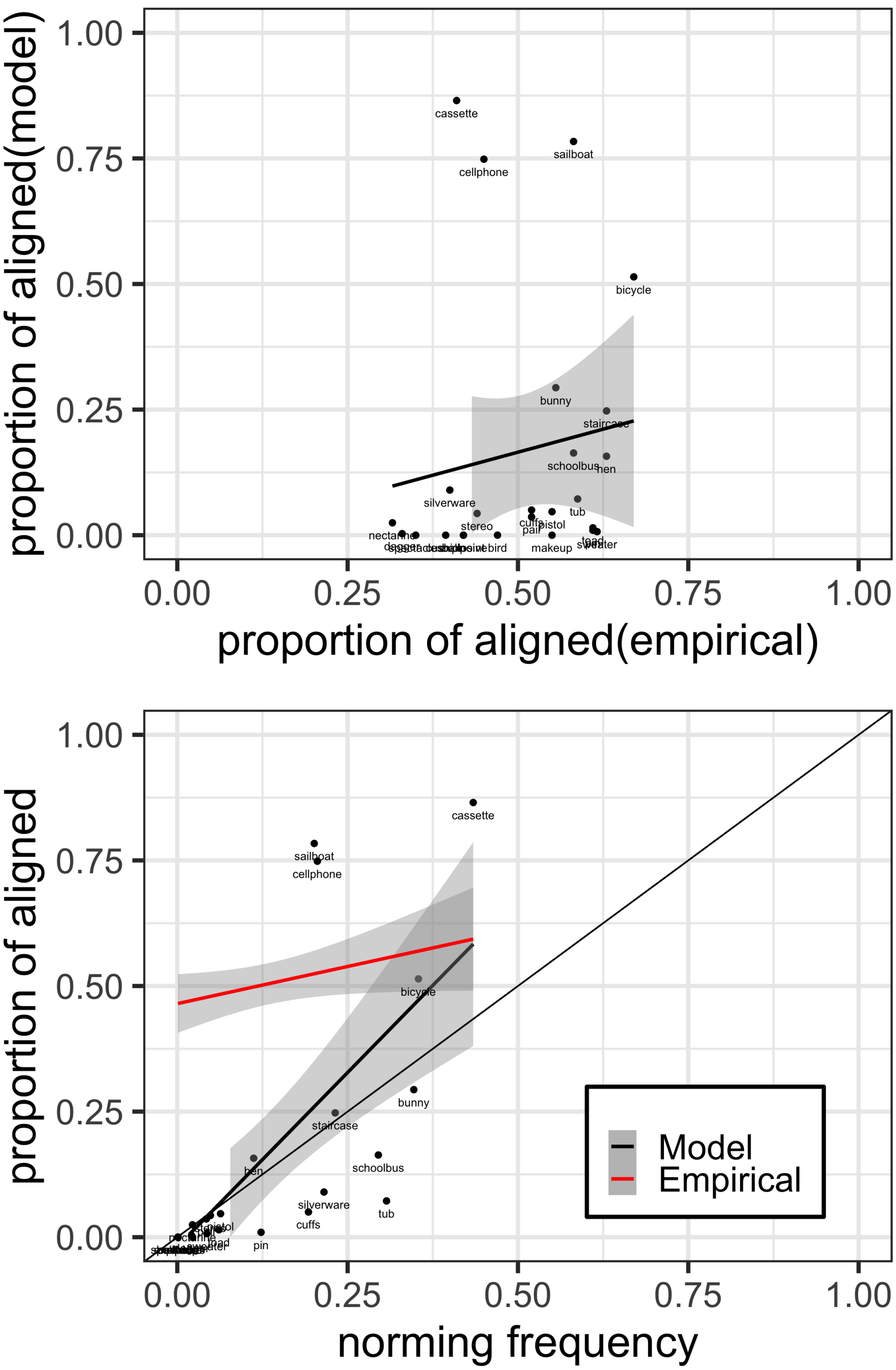


Figure 4.2: Result of RDC Model with the update only on automatic policy. The graph on the top shows the correlation between model prediction and empirical result. The result with the update only on automatic policy does not fit to empirical result. The graph at the bottom shows the entrainment boost of each item. The red line represents the entrainment boost observed from the empirical study, and the black line represents the entrainment boost predicted by the model. This model fails to predict the entrainment boost.

When only updating the reward, the model performed too much of a boost, where most stimuli resulted in model predictions above 0.75, exceeding the range of empirical results (Figure 4.3, bottom). This model result still does not fit the empirical result (Figure 4.3, top, $r = 0.001$, $p = 0.99$).

Then, updates are applied to both components to test the model. When having both components updated, the model result shows the trend of fitting to the empirical result ($r = 0.346$, with $p = 0.09$, Figure 4.4, top), and the update also shows the inverse frequency effect, where stimuli with norming frequency close to 0 are boosted more than stimuli with a larger norming frequency (Figure 4.4, bottom). Although the result is not statistically significant, it provides insight into exploring possible parameters that could improve the model's fit to the empirical result.

Parameter Manipulations

In the exploration of free parameters, I aim to find whether any parameter plays a crucial role in affecting the model, or whether multiple parameters should be manipulated together to achieve a better fit. There are three free parameters in the model: α_P, which manipulates the learning rate of the automatic policy; α_θ, which manipulates the learning rate of the reward policy; and β, which balances the weight of the two policies. I test the range of 0.1 to 2 for each parameter and investigate all valid triplets. This range is determined because the parameters must be greater than 0, and values larger than 2 would yield results close to 1 for all stimuli, which exceeds the reasonable range for the target. The valid triplets should follow the criteria that 1) the correlation between the model and the empirical result is positive; 2) the boost is positive; 3) the model significantly fits the empirical result; and 4) all stimuli are included in the model. Under the criteria, all valid parameters are mapped in Figure 4.5. From this result, we can infer that the range of α_P must be smaller than 1. Furthermore, valid α_θ changes as valid β changes. When α_θ is smaller, β becomes bigger.

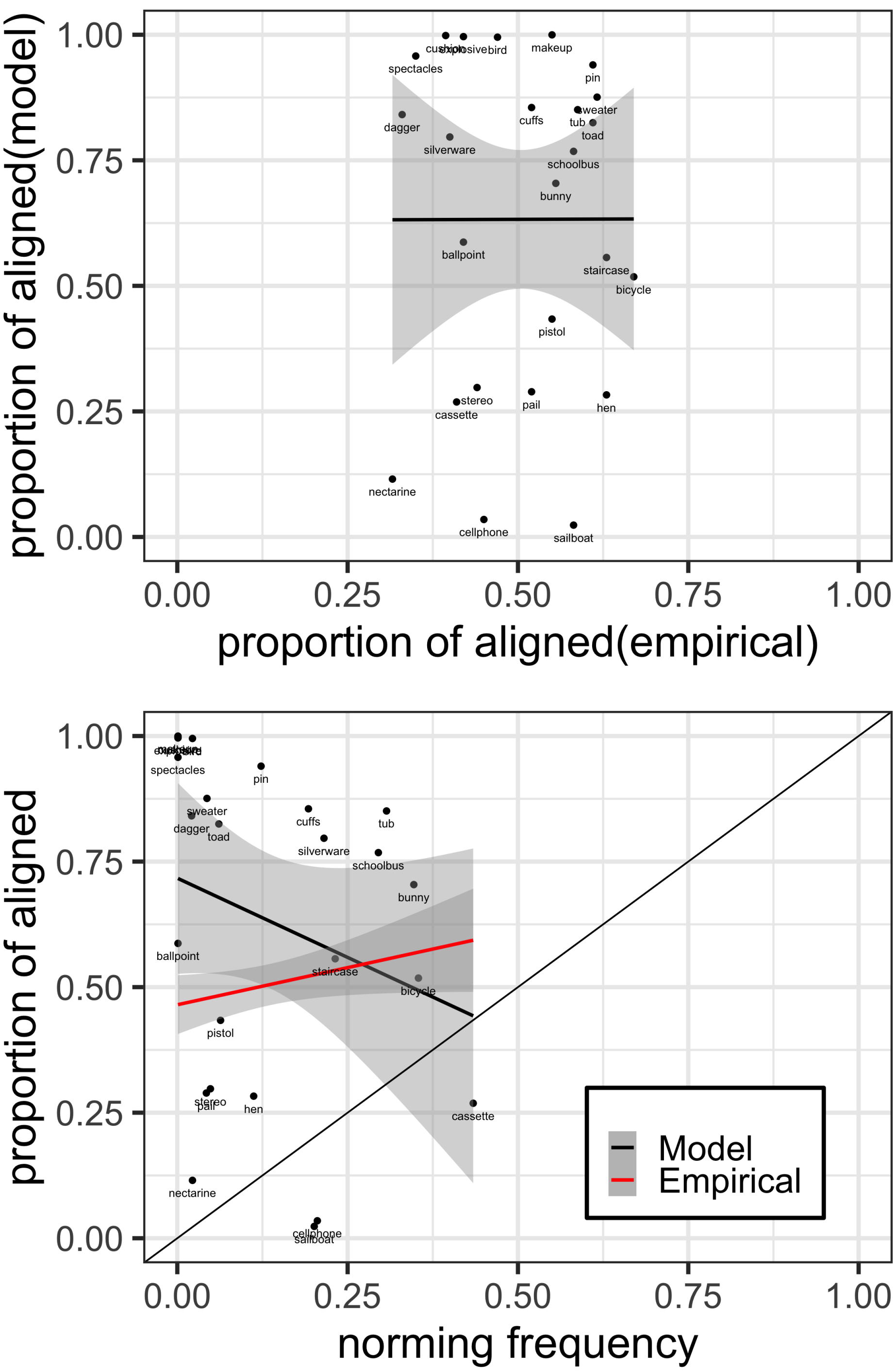


Figure 4.3: Result of Model with the update only on reward. The result with the update only on reward does not fit to empirical result, and this model fails to predict the entrainment boost.

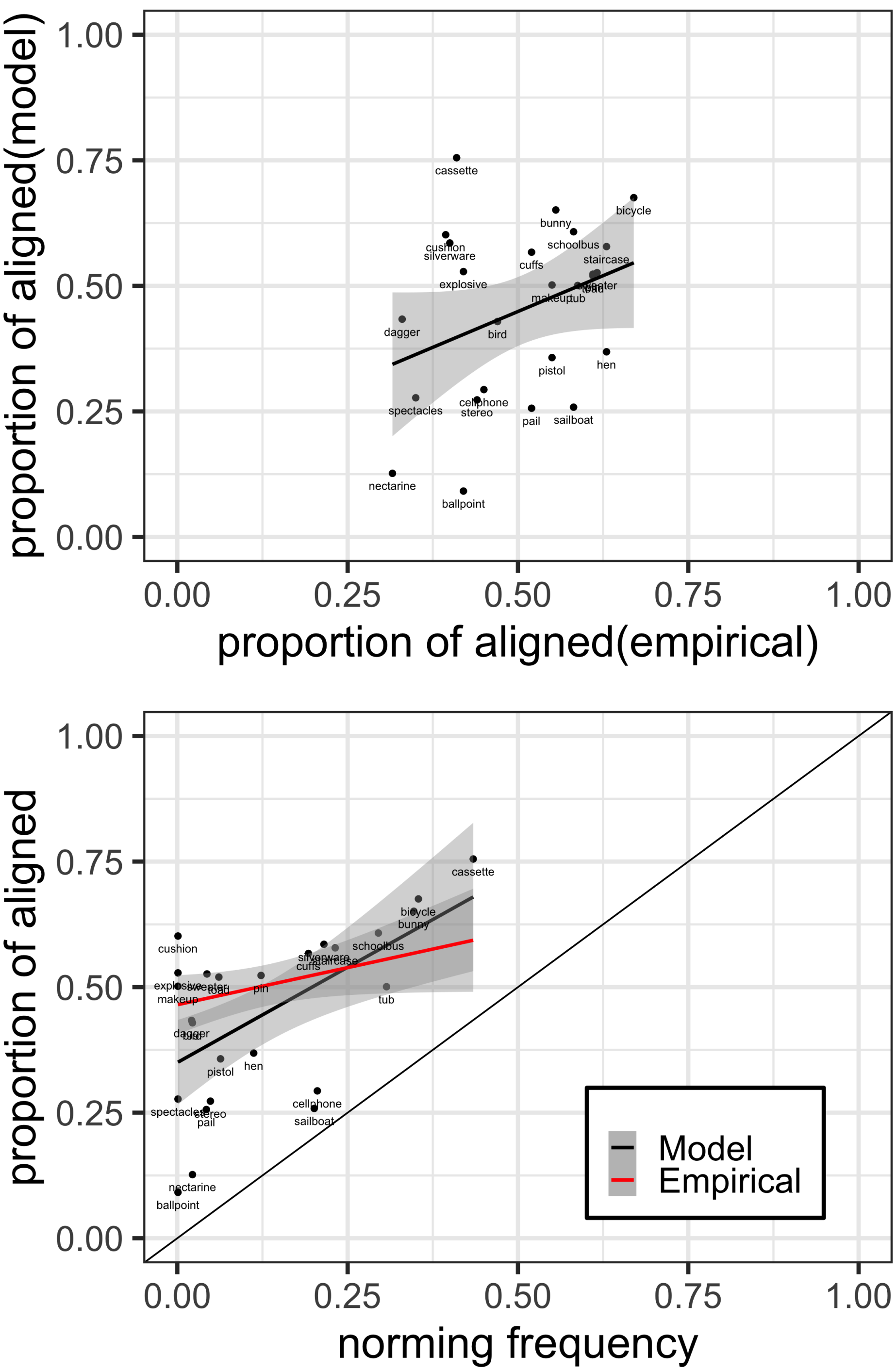


Figure 4.4: Results of Model with the update on automatic policy and reward. The correlation shows a trend of fitting to the empirical result (top), although it is not statistically significant. It also successfully predicts the entrainment boosts (bottom).

α_θ is the parameter to manipulate the learning rate of reward, so when the learning rate of reward is low, the weight of the reward part needs to be larger to make the entrainment happen.

In order to identify the best-fitting model, I also examine the RMSE of the model results, where a smaller RMSE indicates less deviation from the empirical data and a better fit. The result is shown in Figure 4.5 (bottom), with blue meaning a smaller RMSE and red meaning a larger RMSE. From these valid results, the model with $\beta = 1.5$, $\alpha_P = 1$, and $\alpha_\theta = 0.4$ is found to yield the best fit to the empirical results (RMSE = 0.16, $r = 0.49$, $p < 0.05$)(Figure 4.6).

The results of this model provide support for the hypothesis that both automatic policy and reward work together to determine the entrainment behavior. All parameters are manipulated to achieve the best fit to the empirical results, indicating that entrainment behavior is not determined solely by the learning of one policy, but rather by the learning that occurs in both processes. In the best-fit model, an α_P value of 1 supports the engagement of the automatic priming in entrainment. It suggests that participants' prior beliefs are easily updated through automatic priming. In terms of an α_θ value of 0.4 representing the learning rate in the cognitive control process, it is smaller than the learning rate in the automatic priming. It suggests that participants may be more cautious when learning through the rational process. The update through the rational process may occur more slowly than that driven by automatic priming. However, a β value of 1.5 gives a greater weight to the rational process relative to automatic priming in determining entrainment behavior. Although participants learn slowly through the rational process, they may rely more on this mechanism when entraining to others. From this best-fit model, the parameters can reflect how participants monitor each mechanism in entrainment. This model also provides insights into investigating individual differences and related factors in lexical entrainment by manipulating its parameters.

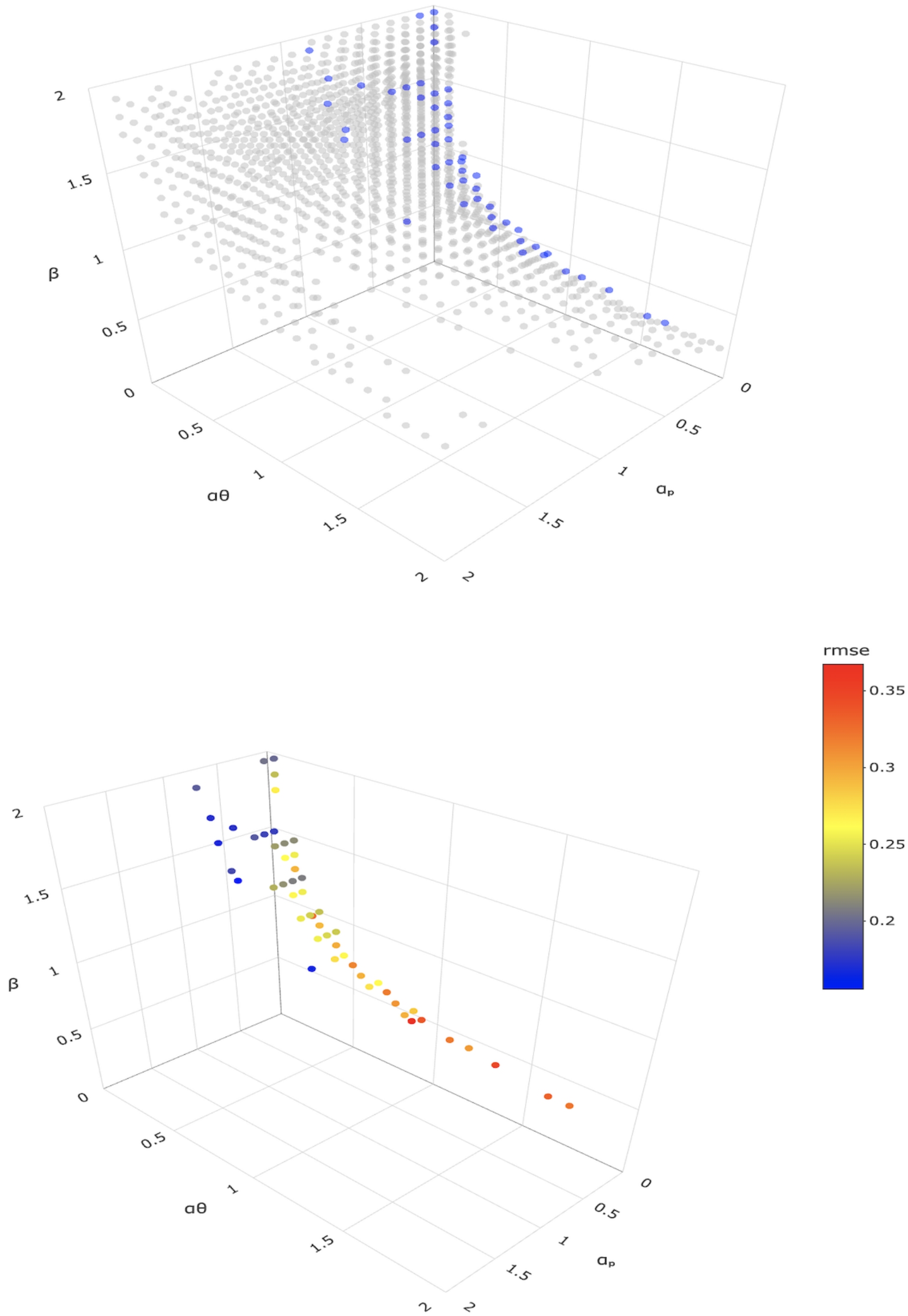


Figure 4.5: Model fitting results of all parameter triplets under the criteria. Grey points (top) are results that ran in the model but failed the criteria; blue points are valid results that met all criteria. Under the evaluation of RMSE, valid results are mapped in different colors based on the range of RMSE (bottom).

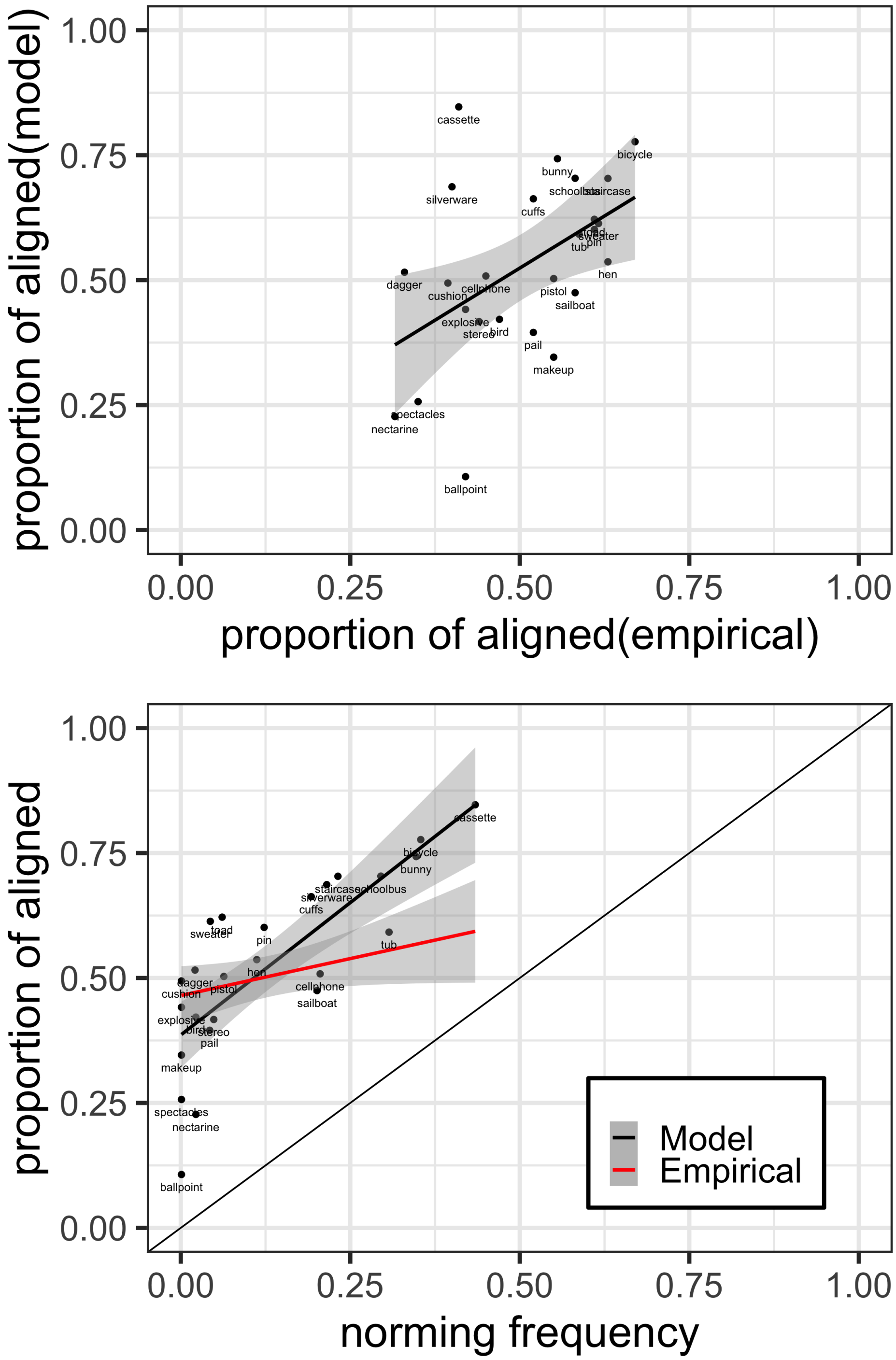


Figure 4.6: Result of Model best fitting to empirical results. The model successfully fits to the empirical results, and predicts the inverse frequency effect in entrainment boost as well.

4.3 Alternative Models

Besides the model presented above, I also made some other attempts in defining automatic policy and reward policy and modeling lexical entrainment. This section discusses the other two failed models. These two models also attempted to implement the RDC model, which contains two components representing automatic priming and cognitive control of reward. However, I explored different ways to define and update policies. In the first model, I defined the reward function as the meaning difference between two words, which was operationalized as the norming frequency difference between two words, and updated the model using a beta distribution. Although the model presents a statistically significant fit to the empirical results, this model contains conceptual errors. The other failed model followed the logic of the RDC model, but did not successfully capture the empirical patterns. I will introduce the two models next and discuss their implementations and limitations.

4.3.1 Define Reward as Frequency Difference

This model has the same basic framework as Equation (4.1), where $P_0(x)$ and $P_0(\bar{x})$ are linked to the norming frequency, and ΔR is defined as the reward differences of using the target and the distractor. I suspected that the reward should relate to how the target and the distractor match with the meaning of the picture—the meaning difference between two words. In this model, I used the difference in norming frequency between the target word and the distractor word (i.e., the frequency of the target minus the frequency of the distractor) to represent the meaning difference. In cases where norming rates are drastically different between the target and distractor, presumably this difference arises because the distractor provides a better match to the image. The reward difference ($\triangle R$) is assumed to relate to the meaning difference, and the full model is presented in Equation (4.12).

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0(x \mid s)}{P_0(\bar{x} \mid s)} + \gamma(P_0(\bar{x} \mid s) - P_0(x \mid s)) \right) \tag{4.12}$$

To update this model, I used a beta distribution with a simplifying assumption that there are only two options, the target word (e.g., *kitten*) and the distractor word (e.g., *cat*), for the speaker to choose. Two parameters α and β are used in determining the beta distribution. These two parameters differ from the parameters in the policy gradient above. In order to first find appropriate values for the parameters α and β, I explored the numbers in the range of 1 to 10 for α and in the range of 1 to 20 for β in the initial state of the beta distribution where k = 0 and n = 0 (meaning no update or exposure happens yet). I assumed that the result in the initial state should be as close to the mean of the word frequency for less frequent words (i.e., $P_0(x \mid s)$) as possible. The appropriate values were found for $\alpha = 2$ and $\beta = 13$ with the mean of target word uses (normalized) $P_0(x \mid s) = 0.134$. Then, the $P_0(x \mid s)$ was updated using the beta distribution (Equation 4.13)(Jacobs, 2015).

$$P_0^{new}(x \mid s) = \frac{\alpha + k}{\alpha + \beta + n - k} \tag{4.13}$$

After the one-time update ($k = 1$, $n = 1$), the new probability of using the less frequent word $P_0^{new}(x \mid s)$ equals 0.2, as a predicted mean of using the target words. The amount of update is 0.067 (i.e., $0.2 - 0.134$). I applied the same update on every item by adding the update to the less frequent words ($P_0^{new}(x|s)$ for each item) and subtracting the update from the more frequent word ($P_0^{new}(\bar{x}|s)$ for each item) in the automatic policy.

In the reward policy, because participants had no idea which word would be used until they received it from the partner, I assumed that the rewards for using both words were the same before the update. Then, the difference in rewards ΔR is 0 before updating. After being exposed to the target word, ΔR changes to the difference between the norming frequencies of the alternative word and the target word. Because the reward relates to the communicative goal, which necessitates using an appropriate word to help the partner select the correct item (in the picture matching and naming task), in this model, I suspected that the reward should relate to how the target and the alternative match with the meaning of the picture—the meaning difference between two words. The idea is that, in cases where norming rates are drastically different between the target and the alternative, presumably this difference arises because the alternative word provides a better match to the image. The reward difference (ΔR) should relate to the meaning difference. After exploring the parameter γ to weigh the reward component, I arrived at $\gamma = 2.67$ with the results best fitting the empirical results. In the results, the model of only having the automatic policy update is:

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0^{new}(x \mid s)}{P_0^{new}(\bar{x} \mid s)} + 0 \right) \tag{4.14}$$

Then, the model of only having the reward update is:

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0(x \mid s)}{P_0(\bar{x} \mid s)} + 2.67 * (P_0(\bar{x} \mid s) - P_0(x \mid s)) \right) \tag{4.15}$$

Finally, combining the update on reward with the update on the automatic policy, the model is:

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0^{new}(x \mid s)}{P_0^{new}(\bar{x} \mid s)} + 2.67 * (P_0(\bar{x} \mid s) - P_0(x \mid s)) \right) \tag{4.16}$$

4.3.2 Results

In the results of evaluating the models with different component updates (Equation (4.14), (4.15), and (4.16)), I also did the same comparisons between the three models with different updates and the empirical results for the aims of testing the hypothesis of whether both mechanisms are necessary in lexical entrainment.

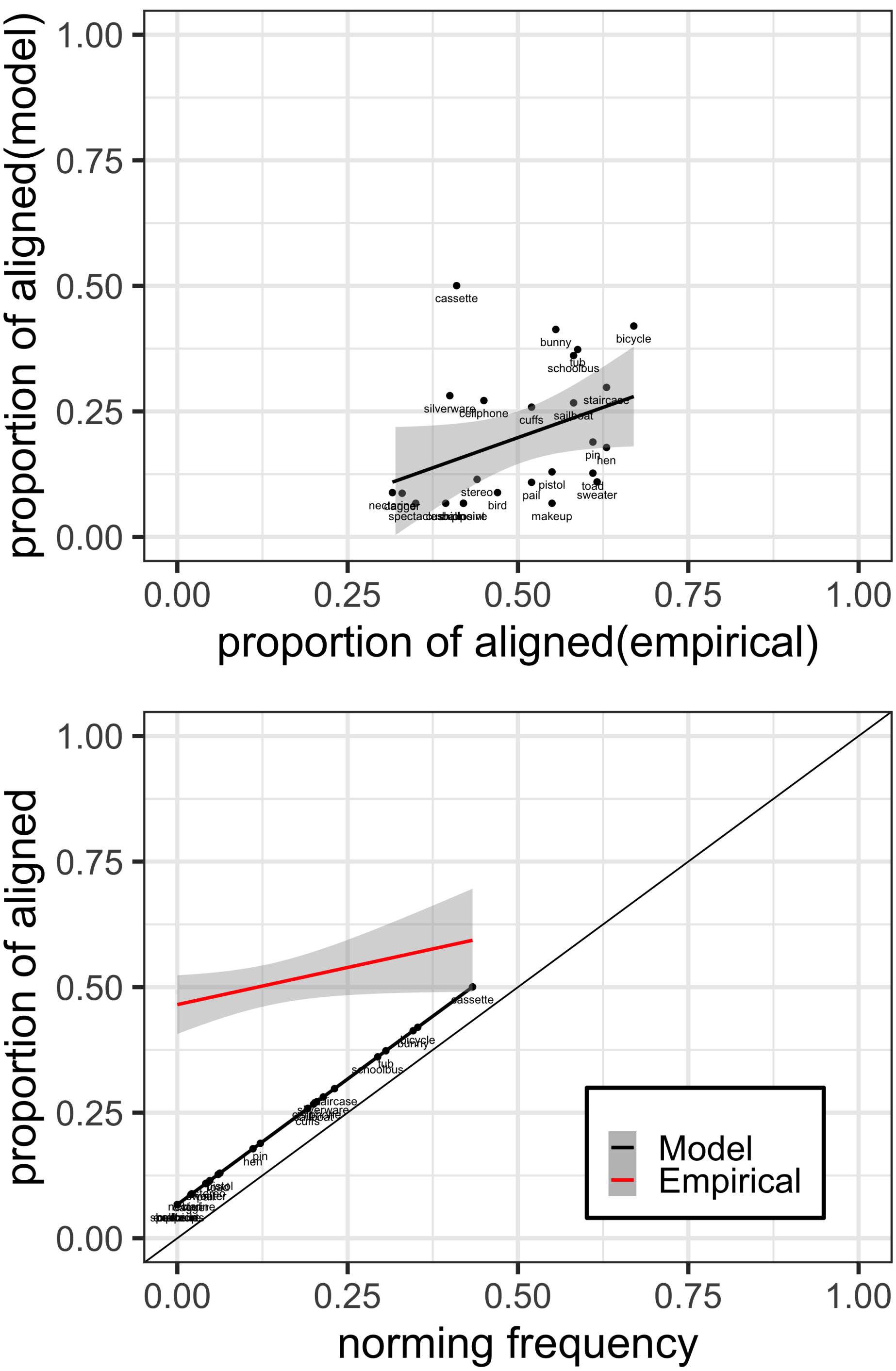


Figure 4.7: Result of the frequency-difference model with the update on only the automatic policy. The reward difference in this model is assumed to be 0. The correlation between the model results and the empirical results is $r = 0.38$, with $p = 0.06$

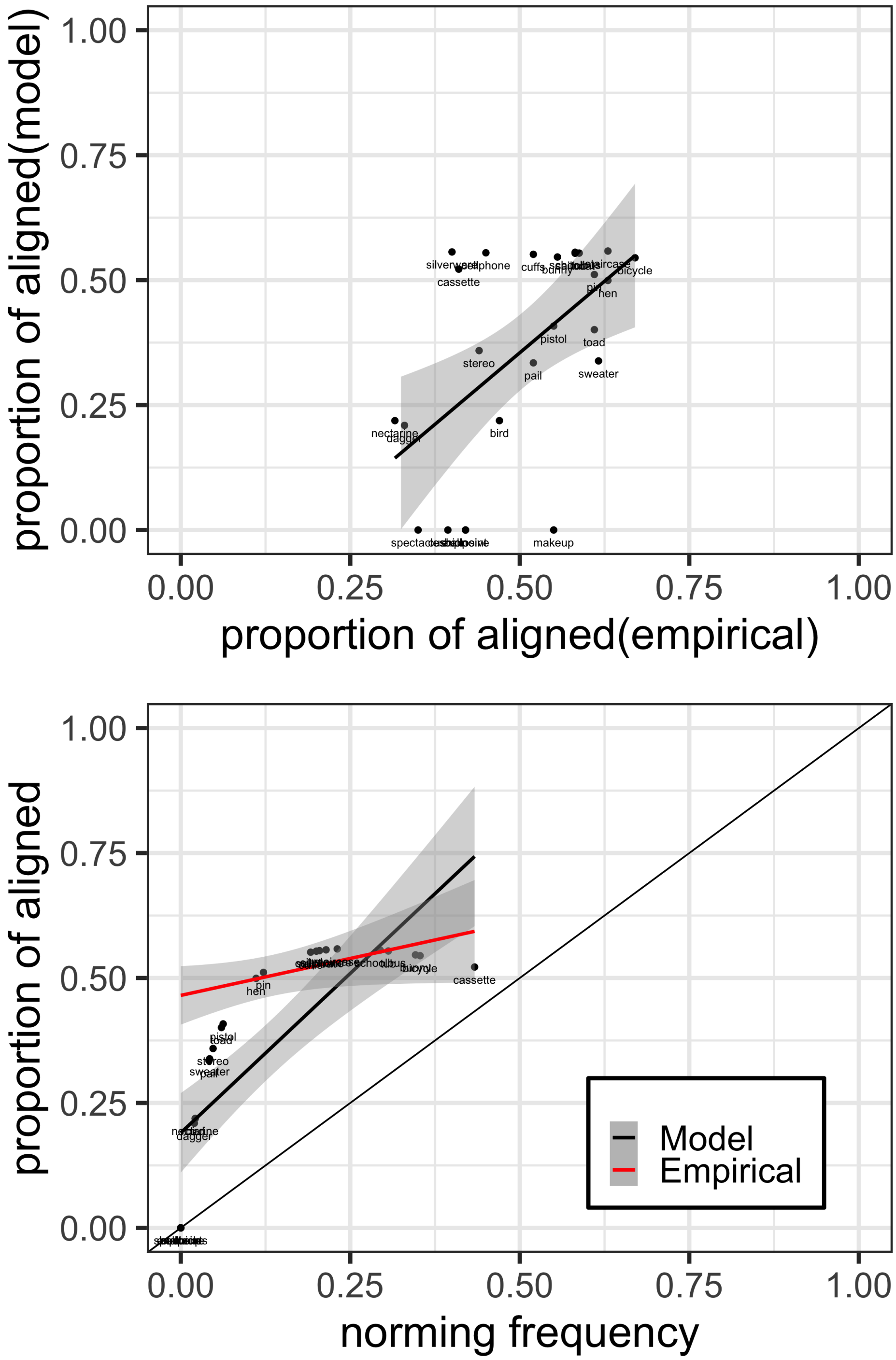


Figure 4.8: Result of frequency-difference model with the update on only reward. The model takes the reward difference as the frequency difference between the alternative word and the target word. The correlation between the model results and the empirical results is $r = 0.56$, with $p < 0.01$

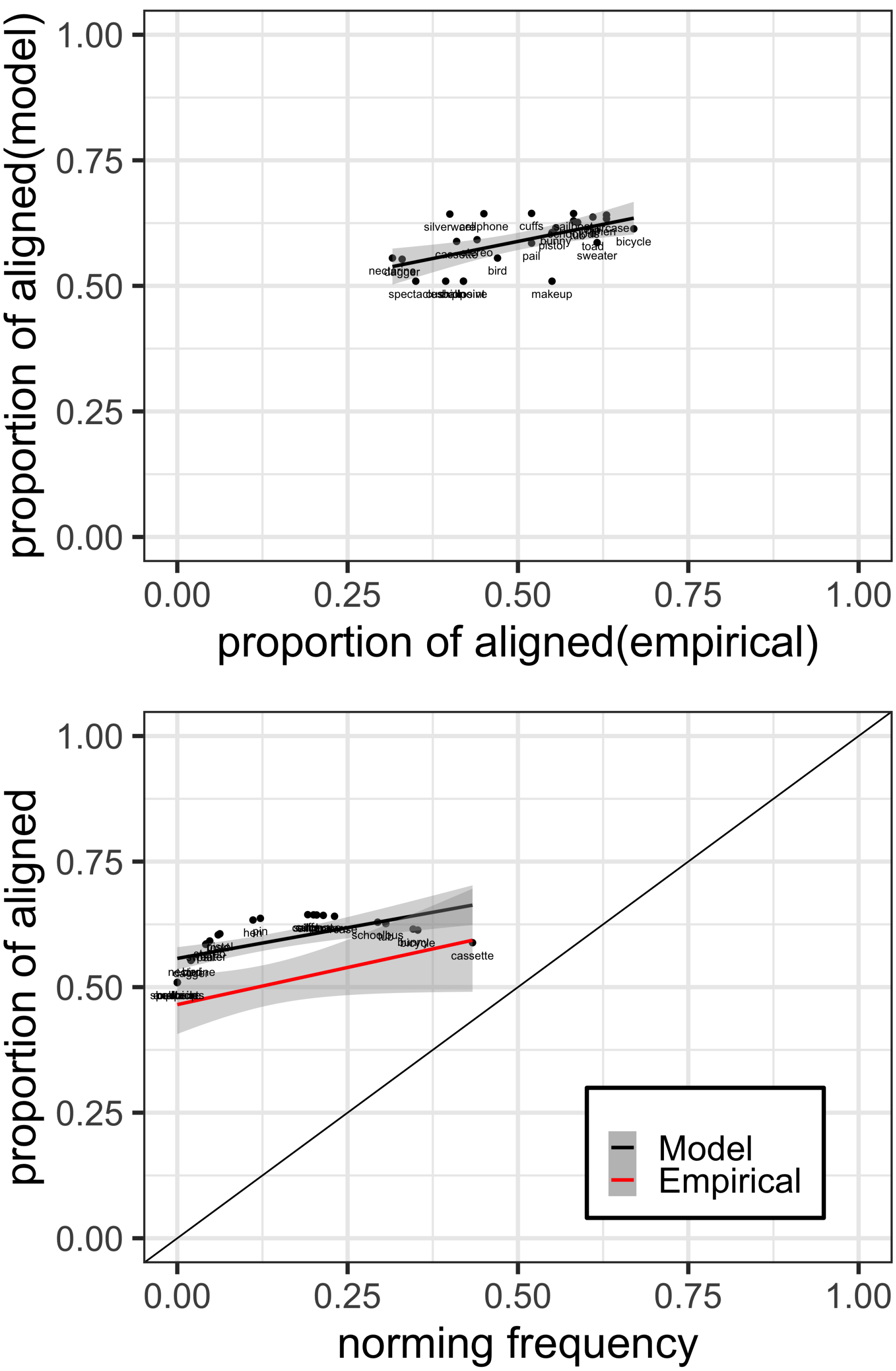


Figure 4.9: Result of frequency-difference model with the update on both automatic policy and reward. The correlation between the model results and the empirical results is $r = 0.58$, with $p < 0.01$

Similar to the results of the main model I presented in the previous section, the model with updates on both automatic policy and reward shows the best fitting ($r = 0.58$, with $p < 0.01$) to the empirical result (see Figure 4.7 for the result of update on only automatic policy; Figure 4.8 for the result of update on only reward; and Figure 4.9 for the result of updates on both.)

Although the model results of updates on both components seem to perform better with a significant correlation, there is a conceptual error in the model that does not make the model sensible. I realized that the assessment of reward policy in this model does not fit the RDC framework. In the RDC framework, the reward is evaluated under the constraints in context. For example, the model in Lai and Gershman (2021) stated the constraint of cognitive capacity in evaluating the reward. In the model from Futrell (2023), reward should also be evaluated under the constraint of the communicative goal. In defining the reward component, the function of the difference between the norming frequencies of two words only captures the relative probability between two words given the goal of naming the image, but it does not reflect the constraints imposed by the goal or evaluate the reward. Consequently, even though the fitting result is significant, the model needs a more thoughtful development of the reward policy.

4.3.3 Define Reward as Frequency Ratio

In another failed model, I instead defined the reward difference ΔR similar to the structure of the automatic policy, but applied different frequencies to different components. The automatic policy represents the frequency of using the word across all conditions, not just for a specific communicative goal, so I applied the word use from corpus data as this probability. The automatic policy is defined as $\ln \frac{P_0(x|s)}{P_0(\bar{x}|s)}$, and I used the corpus word frequency from Speer (2022) for the automatic policy in this model. The frequencies were derived from the zipf frequencies, as I did in the first model. According to Futrell (2023), rewards can be defined as a listener's model indicating the probability of achieving the communicative goal (i.e., the word is understood by the listener) given the use of different words, represented as $P_L(g \mid x)$. For instance, if a speaker wants to refer to a small cat, the reward is evaluated by a listener's model $P_L(\textit{small cat} \mid \textit{kitten})$ relative to the probability of reaching the goal with the alternative word $P_L(\textit{small cat} \mid \textit{cat})$. Therefore, the reward policy in this model is defined as

$$\Delta R = \gamma \ln \frac{P_L(g \mid x)}{P_L(g \mid \bar{x})} = \gamma \ln \frac{P(g)P_L(x \mid g)P_L(x)}{P(g)P_L(\bar{x} \mid g)P_L(\bar{x})} \tag{4.17}$$

$P(g)$ means the probability of the communicative goal and can be canceled in the equation. $P_L(x)$ represents the probability of a word used by the listener, which is evaluated by the speaker. In my scenario of observing entrainment behavior, the speaker has no prior knowledge about the listener or how the listener selects words. In this case, I assume that $\frac{P_L(x)}{P_L(\bar{x})}$ can be canceled as well. Finally, the model with both the automatic policy and reward is

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0(x \mid s)}{P_0(\bar{x} \mid s)} + \gamma \frac{P_L(x \mid g)}{P_L(\bar{x} \mid g)}\right) \tag{4.18}$$

The reward component is operated with the norming frequencies from the empirical study, where the communicative goal is to refer to the correct picture.

In the updates of this model, I applied the beta distribution as the basic update framework. Unlike the previous beta distribution update (Equation (4.13) in Model 4.3.1), where α and β are determined arbitrarily based on the mean of frequency, α and β in this model are parameterized with μ and ν, where μ represents the word frequency of the target word and ν is a parameter controlling concentration around the frequency. In other words, ν indicates how confident the frequency is. In this model, since the automatic policy and reward use different word frequencies and different stimuli have different values of μ, μ_{ap} refers to the target word frequency from the corpus. μ_R is the target word frequency from the norming study. ν is defined as the base-10 logarithm of the sum of the target and alternative word frequencies (exponentiated-unscaled zipf frequencies) from the corpus(Speer, 2022) for every stimulus (Equation (4.19)).

$$\nu = \log_{10}(10^{zipf(x)-3} + 10^{zipf(\bar{x})-3}) \tag{4.19}$$

I set the ν same for both components, indicating that the confidence in the frequencies from the corpus and the norming study is constant. Then, α_{ap} and β_{ap} for the automatic policy and α_R and β_R for the reward can be calculated (Equations (4.20))

$$\begin{aligned} \alpha_{ap} &= \mu_{ap} \cdot \nu \\ \beta_{ap} &= (1 - \mu_{ap}) \cdot \nu \\ \alpha_R &= \mu_R \cdot \nu \\ \beta_R &= (1 - \mu_R) \cdot \nu \end{aligned} \tag{4.20}$$

In this way, every stimulus has its own α and β pairs for the automatic component and the reward component. Next, with the beta distribution update similar to Equation (4.13), the new probabilities in the automatic policy $P_0^{new}(x \mid s)$ and $P_0^{new}(\bar{x} \mid s)$ and the new probabilities in reward $P_L^{new}(x \mid g)$ and $P_L^{new}(\bar{x} \mid g)$ can be calculated. I used the γ equaling 2.67, which is the same as in the previous model. To examine my hypothesis, I again generated three updated models where 1) only updates on the automatic policy

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0^{new}(x \mid s)}{P_0^{new}(\bar{x} \mid s)} + \gamma \frac{P_L(x \mid g)}{P_L(\bar{x} \mid g)} \right) \tag{4.21}$$

2) only updates on the reward

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0(x \mid s)}{P_0(\bar{x} \mid s)} + \gamma \frac{P_L^{new}(x \mid g)}{P_L^{new}(\bar{x} \mid g)} \right) \tag{4.22}$$

and 3) updates on both

$$P_g(x \mid s) = \sigma\left(\ln \frac{P_0^{new}(x \mid s)}{P_0^{new}(\bar{x} \mid s)} + \gamma \frac{P_L^{new}(x \mid g)}{P_L^{new}(\bar{x} \mid g)} \right) \tag{4.23}$$

4.3.4 Results

Similar to the comparisons in the two previous models, I had the comparisons between the results of the three models and the empirical results respectively. The results are presented in Figures 4.10, 4.11, and 4.12.

The results of this model, although showing a trend similar to the first successful model, indicate a non-significant correlation, suggesting that the model may not fit the empirical data well. Therefore, even though this model follows the framework of the RDC model, the updates with a beta distribution cannot help the model predict the empirical entrainment behaviors.

4.4 Discussion

I discuss the possibilities of how the two components work in lexical entrainment. First, perhaps the priming/automatic process is the major mechanism of lexical entrainment (Pickering and Garrod, 2004, 2006). The result of updating only the automatic policy in the successful RDC model did not predict the empirical data. Similarly, when only updating the reward policy, the model still failed to fit the empirical results. Thus, these non-fitting results suggest that both the automatic process and the rational process of evaluating rewards are necessary to determine entrainment behavior. The model (Equation 4.7), with both components updated, successfully fits the empirical results and predicts the inverse frequency effect observed in the empirical study. This result provides evidence to support the hypothesis that

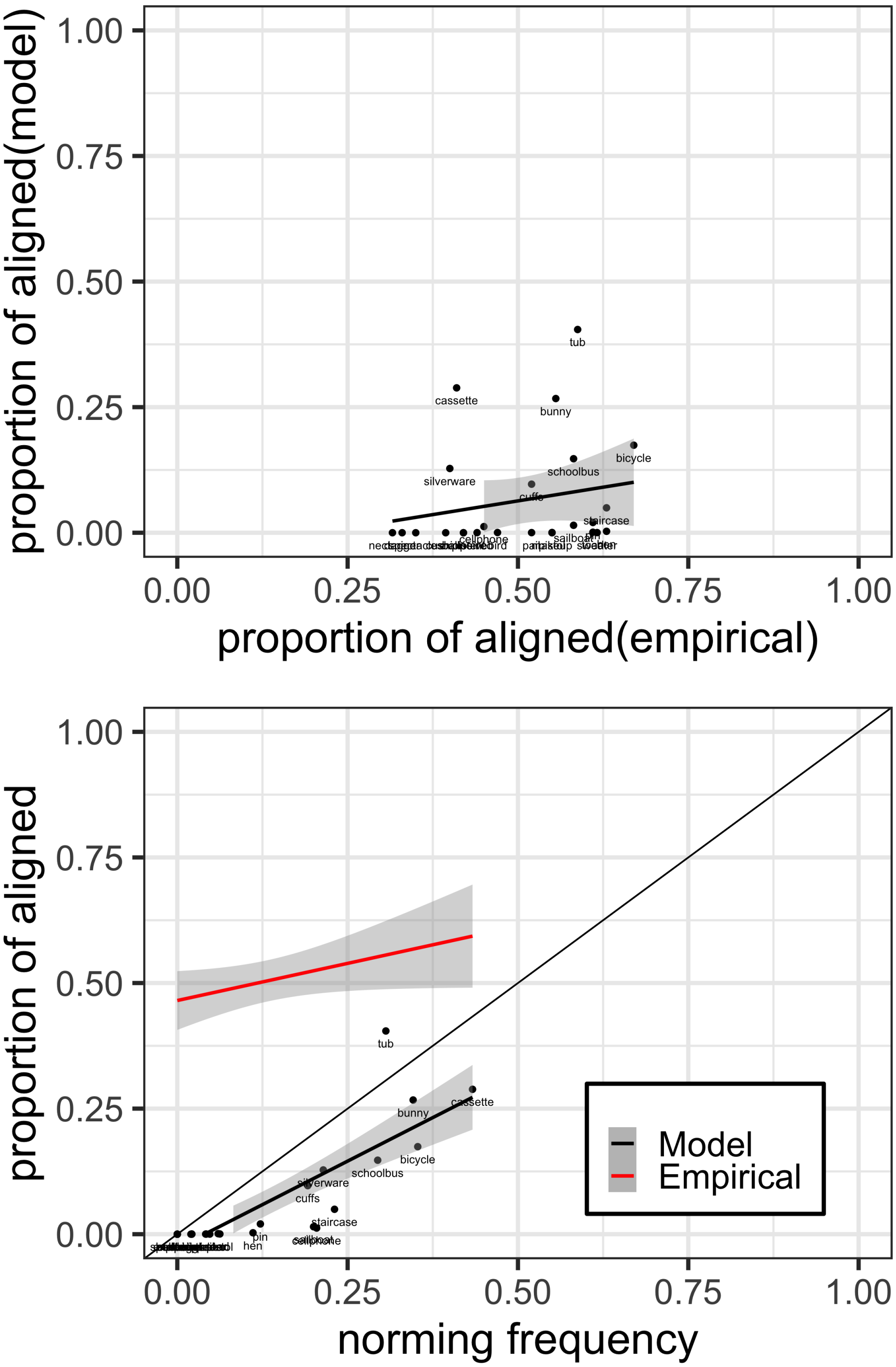


Figure 4.10: Results of Model with the update on only automatic policy. The correlation between the model results and the empirical results is $r = 0.21$, with $p = 0.32$

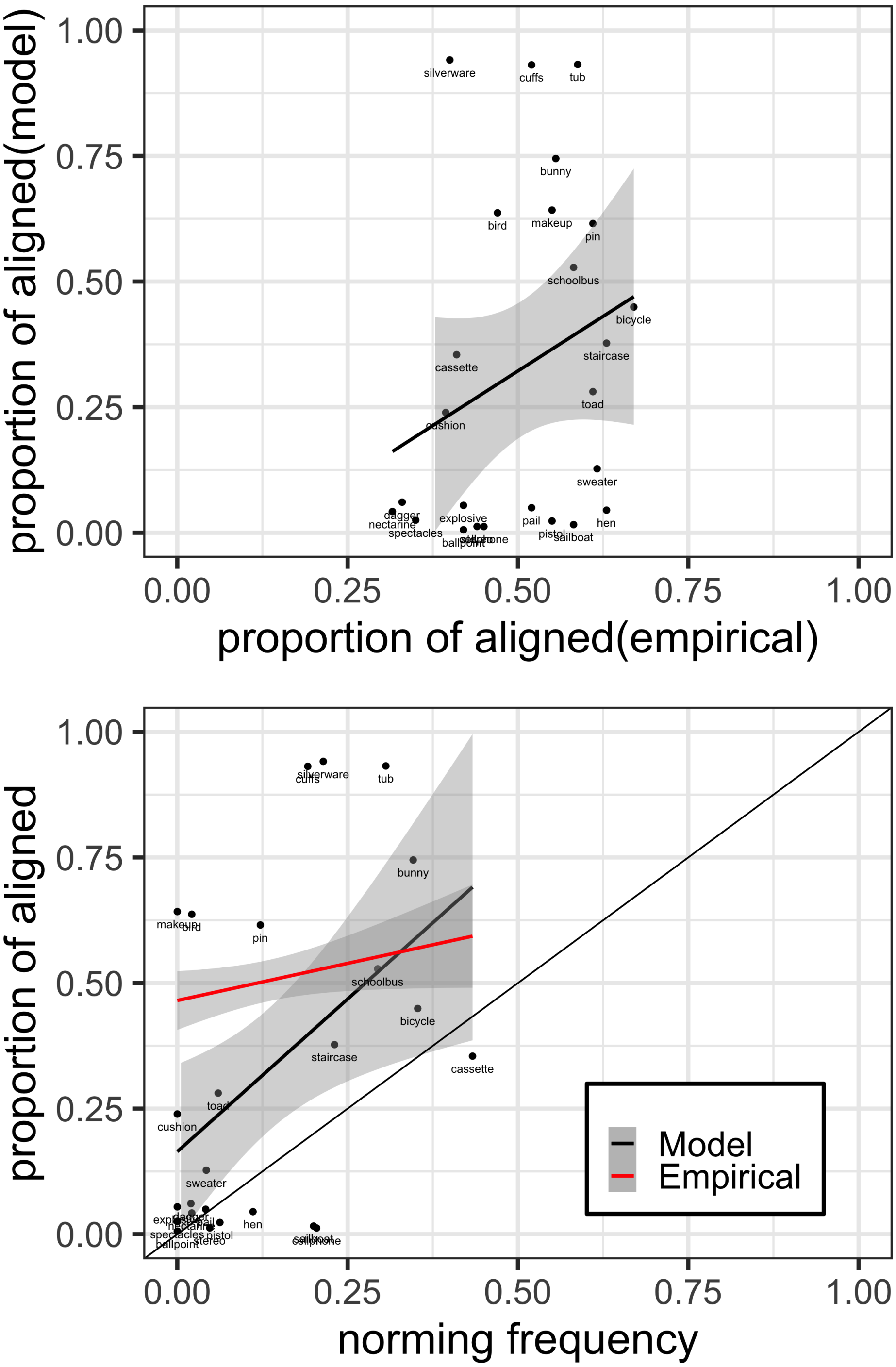
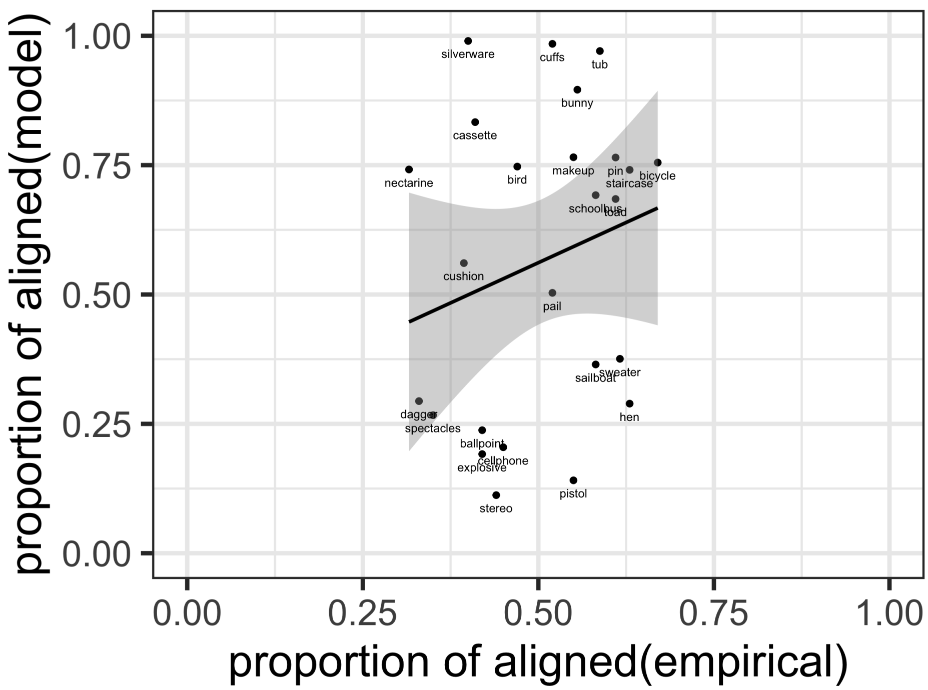


Figure 4.11: Results of Model with the update on only reward. The correlation between the model results and the empirical results is $r = 0.28$, with $p = 0.18$



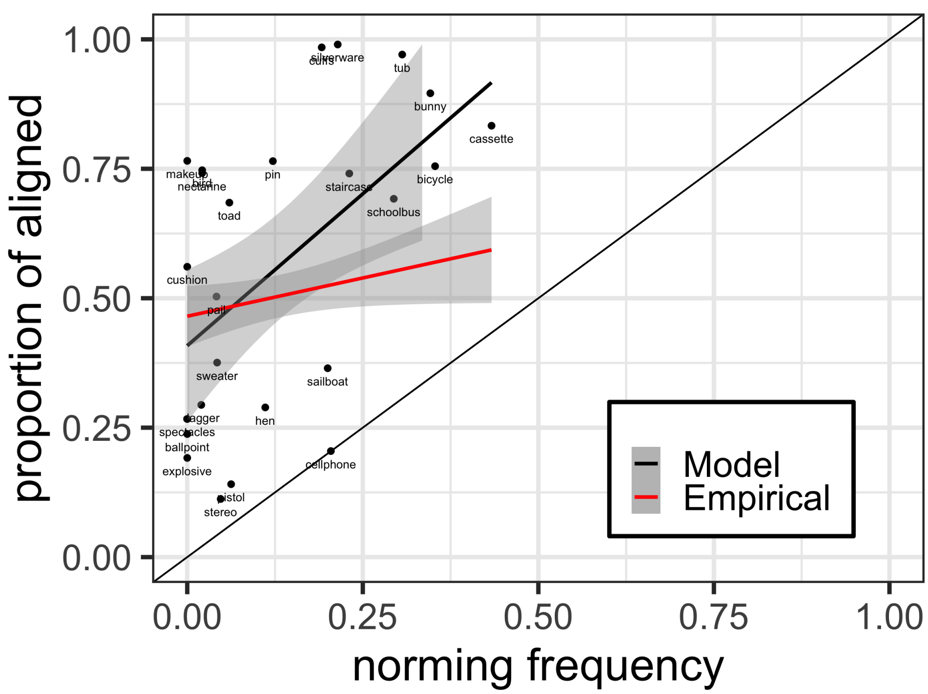


Figure 4.12: Results of Model with the update on both automatic policy and reward. The correlation between the model results and the empirical results is $r = 0.22$, with $p = 0.28$

both mechanisms are active in entrainment. In lexical entrainment, speakers are affected by previous encounters to increase the accessibility of a word. Meanwhile, they are also required to assess the benefits of communication by using the word with the engagement of a rational process. Moreover, it is also an attempt to implement the rate-distortion theory with policy gradient to computationally predict the production process in an interactive scenario.

The model successfully predicts the aggregate results of our previous experimental study, but the empirical picture is more complex. For instance, our study in Chapter 3 stated that bilinguals' proficiency should play a role in affecting their entrainment behavior. Ivanova et al. (2025) emphasize the importance of proficiency in lexical entrainment as well. The manipulation of these factors in lexical entrainment would also provide us with hints of how the two mechanisms are managed under different situations. For instance, proficiency contributes to individual differences in bilingual lexical entrainment, such that bilinguals who are less proficient in one language tend to entrain more. Building on this model, we can hypothesize that these individual differences may be linked to differences in how bilinguals monitor the two mechanisms. It is possible that proficiency affects both the automatic and reward components by manipulating the three free parameters. For highly proficient bilinguals, the semantic connection between the item and its more frequent name may be more robust compared to the connections in low proficiency bilinguals. The strong semantic connections may affect the learning rate parameter in the automatic policy. With more robust connections, it is possible that highly frequent names are always easily accessible for highly proficient bilinguals, regardless of previous exposure to low-frequency names. Thus, the learning rate parameter α_P for priming may have a lesser effect on highly proficient bilinguals, resulting in lower entrainment rates as observed in our study. In terms of the proficiency effect on the rational process, it is possible that bilinguals who are not very proficient rely more on this process for some learning purposes. For example, when bilinguals hear a new word from their conversational partner, they may prefer to use it to increase the chances of practicing and improving their language proficiency. Therefore, a larger learning rate of the policy of reward α_θ will be applied. It is also possible that bilinguals are less confident that their own word choice would be successful, so a larger α_θ would allow them to adjust more flexibly. Lastly, proficiency may also relate to the parameter β, which adjusts the weight of the two components. Highly proficient bilinguals may have a better ability to manage the relative rewards of using different words. Therefore, they are doing better on selecting the appropriate word instead of relying on their partner's responses; the control component of reward evaluation carries more weight.

Chapter 5

Production Modality in Bilingual Lexical Entrainment

From the previous chapter on computational modeling of mechanisms in lexical entrainment, we learn that the rational process may play a dominant role in determining entrainment behavior. This conclusion leads me to propose that if cognitive resources are not sufficient to support the rational process — meaning it is less engaged — entrainment should occur to a lesser extent. One way to divert resources away from the rational process is through difficulty in language production. When production becomes more difficult, more resources are allocated to solving the demanding task rather than to engaging in the rational process. Production modality drew my attention as a potential way to manipulate the difficulty of bilingual language production, and I suspect that it may influence how bilingual speakers entrain.

In research on language production, typing — often treated as the same modality as writing — and speaking are considered similar in the stage of conceptualization, but diverge in the encoding process. The shared step of conceptualization is tested by most research on production modality. Some differences between typing and speaking can also be easily observed in everyday language use. For example, spoken language tends to present greater variability in utterances, while typed or written language often involves more complex expressions (Trebits, 2016; Cho, 2018; Son, 2022). Moreover, speaking limits the chance of revision, as speakers cannot erase what they have said, whereas such revision is possible in typing and writing. Because the widespread use of messaging, which involves more typing in interactive contexts, may be changing language use habits across production modalities, the effects of production modality on entrainment behavior become an interesting area for study. From the perspective of understanding bilingual language use, production modality is a factor that has received attention in language teaching, but its effects across diverse bilingual groups remain under-researched. For example, speaking is argued to be anxiety-inducing compared to typing/writing for L2 speakers, so a worse performance in speaking is observed in L2 speakers' language use (Cheng et al., 1999; Trebits, 2016). These comparisons between speaking and typing in bilingual language use lead to my hypothesis that production modality may present different levels of difficulty when bilinguals produce their L2, with speaking being more difficult than typing. This may influence how different mechanisms are recruited in lexical selection, resulting in different entrainment behaviors. In studies about lexical entrainment I have reviewed, different studies conducted different tasks, such as a picture matching and naming task (Branigan et al., 2011; Tobar-Henríquez et al., 2020; Cai et al., 2021; Ivanova et al., 2021, 2025) and a route-giving task (Suffill et al., 2021). These tasks also collected responses in different modalities. For instance, many studies collected responses in typing (Tobar-Henríquez et al., 2020; Zhang and Nicol, 2022), while others collected speaking responses (Cai et al., 2021; Ivanova et al., 2021). For studies using the same task to measure entrainment rates, differences in production modalities operated in the task may cause different performance in entrainment. Although Branigan et al. (2011) did experiments in both speech and typing modalities and observed similar enhanced entrainment behavior in human versus computer interactions, variations may still occur in bilingual lexical entrainment because bilinguals exhibit different language use in different production modalities. In this chapter, I present my latest research on comparing bilinguals' entrainment behavior in speaking vs. typing.

5.1 Production Modality

5.1.1 Comparisons across Different Modalities

In most comparisons of production modalities, handwriting and typing are usually treated as the same modality as writing. The differences between speaking and writing, whether in terms of speaking versus handwriting or speaking versus typing, generally receive more attention. A major question in these comparisons is at what stages speech production and written production (both handwriting and typing) are similar, and at what stages they diverge. An early assumption about modality suggested that written production and spoken production share a similar process of conceptualization and lexical access, including lexical selection. This implies that there may be no difference in word selection across production modalities. Then, if all other factors are held constant, word selection behaviors, such as lexical entrainment, may follow similar patterns across modalities. To investigate conceptualization across different production modalities, Bonin and Fayol (2000) compared semantic interference effects in writing and speaking and examined whether the effect observed in speaking also occurs in writing. Semantic interference occurs when a speaker is previously exposed to words that share semantic features with the target word. This prior exposure interferes with the production of the target word. When the speaker is asked to produce the target word, they respond more slowly (Schriefers et al., 1990). For example, when a speaker intends to say *cat* to name a little cat, the prior presentation of a semantically related word, such as *dog*, will slow down the response compared to the presentation of a semantically unrelated word, like *table*. Semantic interference suggests the competition among lexical items during lexical selection. As mentioned in the background chapter, in the stage of lexical access, not only is the target lexical item activated, but other lexical items that share semantic features are also co-activated. In the semantic interference effect, prior exposure to a competing lexical item strengthens its activation, so that this competition slows down the retrieval of the target word. If production modalities share the same stage of lexical access, the semantic interference effect should be observed, and the latency of producing the target word should be similar across modalities. In Bonin and Fayol (2000), the authors replicated the same study by asking participants to write down the names of pictures while hearing distractors that were semantically related to the names. They also conducted the same study in the spoken modality for comparison. The auditory distractors were presented either 150 msec before the onset of the picture (stimulus onset asynchrony (SOA) = -150 msec) or simultaneously (SOA = 0 msec). In their results, they found that participants had longer response latencies in the -150 msec SOA condition in both written and spoken productions. This result was consistent with the finding in Schriefers et al. (1990), leading them to conclude that written production follows a similar process of lexical access to spoken production, and interference occurs similarly in both modalities.

The same conclusions were drawn in both Perret and Laganaro (2012) and Qu et al. (2016) with neuroimaging evidence from electrophysiological measurements of brain activity. In Perret and Laganaro (2012), the authors investigated the time course of processes in both handwritten and spoken production during picture naming tasks by monitoring event-related brain potentials (ERPs). They identified ERP amplitude changes in each modality during the time windows preceding and following the response, and mapped these changes to the processing stages from visual stimulus onset to response completion. They compared the ERP changes between two modalities to see if the changes follow the same pattern across modalities. They found that both written and spoken productions have identical electrophysiological activity in the time period of visual (i.e., processing visual input) (0-150 ms), semantic access (i.e., conceptualization) (150 ms - 190 ms), and lexical-semantic (i.e., formulation) (190 ms - 260 ms) processes. After these stages, the activities for the two modalities began to diverge. These results indicate that writing and speaking share processes of conceptualization and formulation, but become different in the encoding stage. In speaking, individuals need to access phonological units and plan the motor movements required for articulation. In handwriting, by contrast, they must retrieve the spelling of words, which does not always correspond directly to the phonological forms. Considering the typing process, it is also reasonable to assume that typing may share similar processes of conceptualization and formulation and differ in the encoding process, because typing requires people to plan keystrokes with finger movements besides spelling the words. Although handwriting and typing are often assumed to be the same, and few studies have compared them, the possibility of differences between handwriting and typing is raised in the results from Baus et al. (2013) and Qu et al. (2016).

In the study of Baus et al. (2013), they investigated the time windows of processes in typewriting and speaking by observing the word frequency effect across modalities. The word frequency effect refers to the phenomenon where high-frequency words are produced more quickly than low-frequency words. This effect has been examined in speaking, and its ERP effect typically arises around 200 ms (Strijkers et al., 2010). Baus et al. (2013) discussed central-cognitive processes, in which individuals conceptualize semantic meanings and access lexical items (i.e., conceptualization and formulation), and peripheral processes, which involve motor planning and execution, in language production across both typewriting and speaking. They monitored ERPs during a Spanish picture naming task in typewriting and speaking, and observed the time course of the word frequency effect in central processes and peripheral processes. According to Perret and Laganaro (2012), the time window for central processes was observed to be around 260 ms in both handwriting and speaking. Therefore, if the stages of conceptualization and formulation are consistent across modalities, the word frequency effect in typewriting would also be expected to occur around 200 ms within the central processes. In the results, Baus et al. (2013) successfully observed the word frequency effect in both typewriting and speaking, noting that lower frequency words were produced slower in both modalities. However, by observing ERPs at posterior sites, they found that the word frequency effect occurred later (at 330 ms - 430 ms) in typewriting compared to speaking (within 200 ms), which contrasts with the conclusion of previous research. They raised a possible explanation that the connections between the semantics of objects and linguistic representations for speaking (e.g., lexical phonology) may be stronger than the connections with linguistic representations for typing (e.g., orthography), because naming objects in speech is much more common than in typing. Referring back to the production models proposed by Dell (1986) and Levelt et al. (1999), the model for speaking includes connections between the lexical level and the phonological level, supporting the idea of a cascaded process between these two stages. However, Baus et al. (2013) proposed that, unlike in speaking, weaker connections between the lexical level and the level of keystroke units may slow down responses in typing. This may imply that typewriting does not fully share the same processes as speaking. The conclusion from Baus et al. (2013) was contradicted by Qu et al. (2016). In Qu et al. (2016), the authors also investigated the word frequency effect on both writing and speaking modalities with an electrophysiological measurement. Instead of testing typewriting, they performed a picture naming task in both handwriting and speaking Mandarin. They found that the electrophysiological activities began to diverge from 168 ms, similar to the onset time of lexical access in spoken production (within 200 ms). Therefore, handwriting should share the same process of lexical access as speaking. If typewriting and handwriting are treated as the same modality in contrast to speaking, they would be expected to share similar features in lexical processing. However, the mixed results from these two studies suggest that there may be variability in the language production processes across typing, handwriting, and speaking.

With the increasing amount of typing in daily language use, typing has become a modality that shares features with both speaking and handwriting. Its usage also varies depending on different scenarios. For instance, students perform typing for writing essays, which typically involves more complex forms such as complex sentence structures and formal vocabulary use. This production suggests the involvement of planning and error monitoring in handwriting (Biber, 2009). In formal writing, students can select their utterances and organize sentence structures with careful consideration before they begin writing. During the writing process, they also engage in error monitoring and continually revise their work. These processes allow for deeper cognitive engagement during production. On the other hand, typing has also become very common for daily communication through messaging, and language use in typing has become simpler with the prevalent use of abbreviations and a less formal tone. In this case, formality and complexity are not necessary, but it requires immediate responses, similar to those in speaking, which limits opportunities for thorough planning and accurate language production. More errors and simpler sentence structures may be produced, similar to what is observed in speaking. In the meantime, automatic correction features on devices also help people relieve the load of monitoring production errors. With these changes, we expect that typing may also bring variations in lexical selection. When we select a word to type, we may also consider the efficiency of typing and select the words that are easier for us to type. For example, the QWERTY effect in typing proposed by Jasmin and Casasanto (2012) states that words typed with more right-hand letters are evaluated more positive and are more likely to be preferred or used. This effect implies that typing possibly correlates to how individuals perceive word meanings, leading to different word selections. Hence, it is reasonable to also hypothesize that people may have different preferences of word uses in typing compared to speech and handwriting. Moreover, bilinguals, in particular, exhibit distinct patterns of language use across modalities. I will discuss modalities in bilingual language production next.

5.1.2 Modality in Bilingual Language Production

In research on the effect of production modality on second language production, bilinguals perform differently in writing and speaking.

For instance, Cho (2018) observed bilinguals' task performance in different levels of task complexity (easy task vs. difficult task) and across different modalities (writing (typing) vs. speaking) in bilinguals' L2. They did an argumentative task where participants were asked to defend a topic they selected either by writing an essay or giving a speech. Participants completed four tasks: a simple speaking task, a complex speaking task, a simple writing task, and a complex writing task. Simple tasks were evaluated using two criteria, while complex tasks were evaluated using five criteria. Then, they counted the number of words and errors in clauses to measure the complexity and accuracy of the production, and they counted the non-meaningful utterances, pauses, and repetitions as a measurement of fluency. While their results did not show an interaction effect between task complexity and modality, they found that speaking performance contained more words than writing performance. Meanwhile, the speaking performance contained more errors. The finding of fewer errors in writing production suggests that error monitoring may be easier in writing than in speaking. As I previously discussed regarding the difference between formal writing and speaking, participants were allowed to edit and revise their output in writing, possibly with fewer cognitive resources involved, which likely contributed to reduced error rates and improved written performance. This conclusion indicates that L2 bilinguals may experience differences in production difficulty between speaking and typing (writing), with speaking requiring more cognitive resources.

Son (2022) also studied second language production in different modalities. The author tested L2 learners at different proficiency levels to describe pictures in both speech and writing modalities within a period of time. The duration of responses should be close to the time limit. If a response was too short, the researcher also asked further questions about the picture to encourage production. The results showed that there were more syntactically and lexically complex utterances in writing (e.g., longer sentences, longer clauses, and more clauses observed), whereas spoken production contained more lexical diversity (e.g., more types of words) and more subordinate clauses. These results indicate that bilinguals focus their effort on different aspects of language production depending on the production modalities. For instance, the more syntactically complex utterances indicate that bilinguals expend more cognitive resources on organizing sentence structures. More diverse lexical use in speaking may suggest that bilinguals tend to switch lexical selections to enrich their speech.

From a psychological perspective, Trebits (2016) argued that individual differences in bilingual language production across modalities might be influenced by the different levels of anxiety in processing and producing the second language. They hypothesized that speaking is more anxiety-inducing than writing for L2 speakers due to 1) its public nature, meaning that speaking often occurs in real-time and in front of others; and 2) the lack of planning compared to writing, as discussed previously, since speaking requires immediate responses or responses in a short time limit. Trebits (2016) tested L2 learners completing narrative tasks in different levels of difficulty and in two modalities (speaking vs. writing (typing)), and assessed their anxiety in production through a questionnaire about input, processing, and output anxiety (IPOA) developed by MacIntyre and Gardner (1994). They assessed the production performance from the aspects of accuracy (i.e., the ratio of error-free clauses), lexical diversity (i.e., the ratio of word types), syntactic complexity (i.e., clause length and the ratio of subordinate clauses), and fluency of speech (i.e., speech rate). Consistent with the conclusions of language use across modalities from other studies, they observed greater lexical diversity and more use of subordinate clauses in speaking, as well as greater syntactic complexity in writing (e.g., longer clauses). In terms of accuracy, more errors were found in both writing and speaking when in a more complex task associated with a higher level of anxiety, suggesting a worse performance in both modalities. However, the negative effect of the output anxiety appeared to be greater in speaking than in writing, supporting the hypothesis that speaking induces more anxiety than writing.

From these studies, it is evident that bilinguals may exhibit different performance patterns across various modalities. It is possibly due to psychological factors, such as anxiety, and cognitive factors, such as different cognitive loads required for different modalities. Although we have observed some discrepancies in bilingual language production between modalities, we still lack clarity about the specific role of production modality in bilingual language production, particularly when considering the diversity among bilingual populations. For example, do bilinguals have different preferences for lexical selection across different modalities? How might bilinguals' proficiency levels interact with production modalities to influence their language production? Do different production modalities engage different cognitive mechanisms in language production? In this chapter, I aim to explore these questions by examining bilingual lexical entrainment across various production modalities. It enables us to determine whether lexical selection is processed differently in response to changes in production modalities. Additionally, it could provide us with insights into the diversity of bilingual language use from the perspective of production modality. We can expand our focus on diverse bilingual groups, take lexical entrainment behavior as a medium, and include different modalities of production to develop a more comprehensive understanding of lexical production and investigate how it leads humans to achieve successful communication.

5.2 Present Study

To investigate the effect of production modality on bilingual lexical entrainment, I conducted an empirical study to test entrainment behaviors in either typing or speaking modality among bilinguals proficient in English.

I first hypothesize that modality brings differences in entrainment behavior. As discussed in the background above, speaking and typing differ in terms of error monitoring. When making errors in speaking, speakers are unable to erase their incorrect production but must produce a new word as the correction, whereas individuals can freely edit their production in typing. Therefore, speaking may place more cognitive demands on error monitoring to initiate accurate production. The feature of real-time production in speaking also induces more anxiety for L2 bilinguals compared to typing/writing, as Trebits (2016) argued. Thus, speaking appears to be more difficult than typing for bilinguals, especially for L2 speakers, and speaking is expected to involve greater cognitive load, which may lead to fewer cognitive resources being allocated to other processes, such as evaluating and selecting appropriate words. In contrast to speaking, individuals have the opportunity to revise their production without penalty in typing, so they don't need to put as much effort into monitoring errors when initiating production. Typing has also been argued to induce less anxiety than speaking for L2 bilinguals (Trebits, 2016), which may free up more cognitive resources that would otherwise be used to manage anxiety. Therefore, more cognitive resources are available for the processes of evaluating and selecting words. According to the findings from Chapter 4, where the rational process plays a dominant role in affecting the entrainment behavior, a lack of resources for the rational process should result in reduced entrainment. Thus, I hypothesize that bilinguals will exhibit less entrainment in speaking than in typing due to the greater diversion of cognitive resources required for speaking.

Moreover, I also examine whether production modality interacts with bilingual proficiency to affect bilingual lexical entrainment. In Chapter 3, I discussed the effect of language proficiency on bilingual lexical entrainment, noting that bilinguals who are less proficient in one language tend to entrain more. Based on this, I first hypothesize that a similar proficiency effect occurs across both speaking and typing. However, its effect may vary depending on the production modality. In the study by Trebits (2016), speaking is argued to induce more anxiety than typing for L2 learners, so there might be more pressure for less proficient bilinguals to produce an accurate word initially while also inhibiting the interference from their L1. As a result, more cognitive resources are required for less proficient bilinguals in speaking production, and resources available for the cognitive control process become even fewer. Less proficient bilinguals have to spend fewer resources estimating the rewards between words. In contrast, for more proficient bilinguals, speaking may require less effort, allowing more cognitive resources to be available for evaluating rewards and entrainment. With respect to the proficiency effect on entrainment in typing, I assume that resources are sufficiently available for bilinguals regardless of proficiency. Thus, the proficiency effect is expected to resemble the finding in Chapter 3. However, in speaking, resources are more limited for less proficient bilinguals, which may reduce their entrainment rate. For more proficient bilinguals, since less effort is required in speaking, their entrainment behavior may increase. With the inclusion of both proficiency and production modality, I expect an interaction effect between the two variables: the difference in entrainment rates between typing and speaking should be larger for less proficient bilinguals, but smaller for more proficient bilinguals.

5.3 Method

In this study, I conducted a between-group experiment with two groups. One independent variable is the modality of production, with speaking vs. typing as the two levels. The experiment was designed with JsPsych (De Leeuw et al., 2023) and operated in our lab server.

5.3.1 Participants

I used SONA and Prolific to recruit 135 Mandarin-English bilinguals who are proficient in Mandarin and English. 35 were from SONA (30 females, 4 males, 1 unspecified gender), and 100 were from Prolific (65 females, 30 males, 5 unspecified gender). The pool of SONA participants was mainly college students at UCI. Students who completed the task were compensated with extra credit. I restricted the pool of Prolific to bilinguals who live in the US and are proficient in Mandarin. Participants on Prolific were compensated with $3. As we know, bilingual populations are diverse depending on their language backgrounds, proficiency, language environments, etc. I focus on Mandarin-English bilinguals to minimize the effect of variety in bilingual participants caused by different languages. The average age of participants from Prolific is 32, with a range of 19 to 60. The average age of participants from SONA is 21, with a range of 18 to 24. Participants from SONA are, on average, younger compared to those from Prolific, while Prolific participants span a wider age range. I excluded 5 participants from Prolific who did not report Mandarin as the other language they speak, and 33 participants who did not complete tasks correctly. After data cleaning, 97 participants were included in the analyses, comprising 29 L1 English speakers (2 from SONA and 27 from Prolific) and 68 L2 English speakers (24 from SNOA and 44 from Prolific). 47 participants completed the speaking version of the task (13 L1 English speakers and 34 L2 English speakers), and 50 participants were in the typing version of the task (16 L1 English speakers and 34 L2 English speakers).

5.3.2 Tasks

The experiment asked Mandarin-English bilinguals to complete a picture matching and naming task in English with a virtual partner. I followed a similar experiment procedure to the experiment in Chapter 3. In the experiment, participants completed a short version of

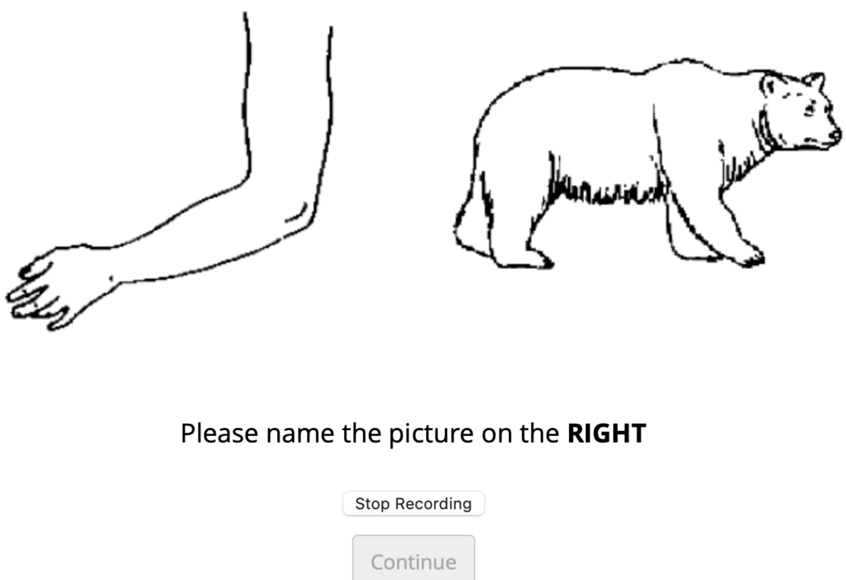


Figure 5.1: Example of naming trails from the picture matching and naming task in the speaking version.

LEAP-Q first to confirm that Mandarin was the language they spoke, other than English, and to report their self-reported proficiency. Then, I used a picture matching and naming task to present the interactive scenario and measure the entrainment rate. This task was introduced in Chapter 3, but some modifications have been made to the task in this study. The task was presented to participants in either a typing or a speaking version. The main structure of the task was similar to the task used in Chapter 3. First, participants were told to pair with a partner to complete the task. Participants saw a target picture and an irrelevant picture, with a word (target word) presented on the screen in the experimental matching trial. In the meantime, they heard the word from a virtual partner. In the trial presenting the word, both visual and auditory modalities were used simultaneously to ensure that the output modality was not influenced by the input modality, and participants would know the spelling of the word. A less frequent word was used as the target word. Participants selected the correct picture according to the word. After two filler trials, participants saw the target picture again and were asked to name it to their partner, either verbally or via typing. The example of the task in typing version appears in Figure 3.1 from Chapter 3, and the example of the task in speaking version is shown in Figure 5.1. In the naming trial of the speaking version, recording began as soon as participants entered the trial and saw the pictures. Participants produced a word out loud to name the target picture and then clicked the "stop recording" button. In both versions of the task, a response latency was recorded for every naming trial. In the typing version, response latency refers to the time elapsed between the start of the trial and the moment participants press the first keystroke. In the speaking version, response latency refers to the time elapsed from the start of the trial to the moment participants produce their first sound. The number of trials in which participants aligned with the partner's words was calculated to represent the entrainment rate. Participants completed a total of 60 trials (i.e., 15 four-trial sequences). I selected 15 objects from 26 objects used in the previous study (Table 5.1). These objects vary in the frequency of use, and the log-frequency ratio between the target word and the more frequent

Distractor	**Target**
bathtub	tub
boat	sailboat
bucket	pail
chicken	hen
frog	toad
glasses	spectacles
handcuffs	cuffs
needle	pin
pillow	cushion
rabbit	bunny
radio	stereo
stairs	staircase
sword	dagger
tape	cassette
utensils	silverware

Table 5.1: 15 selected English nouns used in experimental stimuli. Nouns in the Distractor column are the frequent names for the image, while nouns in the Target column are less frequent names used as the entrainment targets.

alternative word is larger than 0.05 to ensure that the two words have enough discrepancy in word frequency. Following the picture matching and naming task, a verbal fluency task (VF) was used to objectively assess language proficiency and dominance. In the verbal fluency task, participants named as many words as possible for a given category in each trial. There were two sessions in the VF task, and participants completed one session in English and the other in Mandarin. The order of the sessions was random.

Participants who failed to correctly name words for three or more categories in the verbal fluency task were excluded from analysis. Additionally, participants who did not complete the experiment (i.e., provided 10 or more empty responses in experimental trials, or 10 or more incorrect answers in matching trials) were also excluded from analysis.

Table 5.2: Self-reported simplified LEAP-Q for participants

Typing						
	M	SD	Range			
Years in China	10.52	8.44	0.00–28.00			
	English			Mandarin		
	M	SD	Range (0–10)	M	SD	Range (0–10)
Self-reported proficiency						
Understanding	9.16	1.42	4.00–10.00	9.10	1.25	4.00–10.00
Speaking	8.92	1.58	4.00–10.00	8.68	1.57	3.00–10.00
Reading	9.12	1.47	4.00–10.00	7.42	2.90	1.00–10.00
Speaking						
	M	SD	Range			
Years in China	9.70	8.90	0.00–28.00			
	English			Mandarin		
	M	SD	Range (0–10)	M	SD	Range (0–10)
Self-reported proficiency[a]						
Understanding	9.19	0.99	7.00–10.00	9.28	0.95	7.00–10.00
Speaking	8.85	1.32	5.00–10.00	8.81	1.51	4.00–10.00
Reading	9.06	1.15	5.00–10.00	8.11	2.38	0.00–10.00

5.4 Results

After excluding participants who were not Mandarin-English bilinguals or did not complete the experiment correctly, 97 participants were included in the analysis (see Table 5.2 for details on the participant simplified LEAP-Q report). The sample consisted of 71 participants from Prolific and 26 participants from SONA.

I first compare the proportion of aligned objects, representing the entrainment rate, between speaking vs. typing for the main effect of modality on lexical entrainment (see Figure 5.2). This result shows that participants have a significantly higher entrainment rate in typing than in speaking ($\beta = 1.07$, $p < 0.001$), which indicates that Mandarin-English bilinguals tend to entrain more when interacting with a partner and producing words in typing. This result supports the hypothesis that production modality has an effect on entrainment behavior and typing leads to more entrainment.

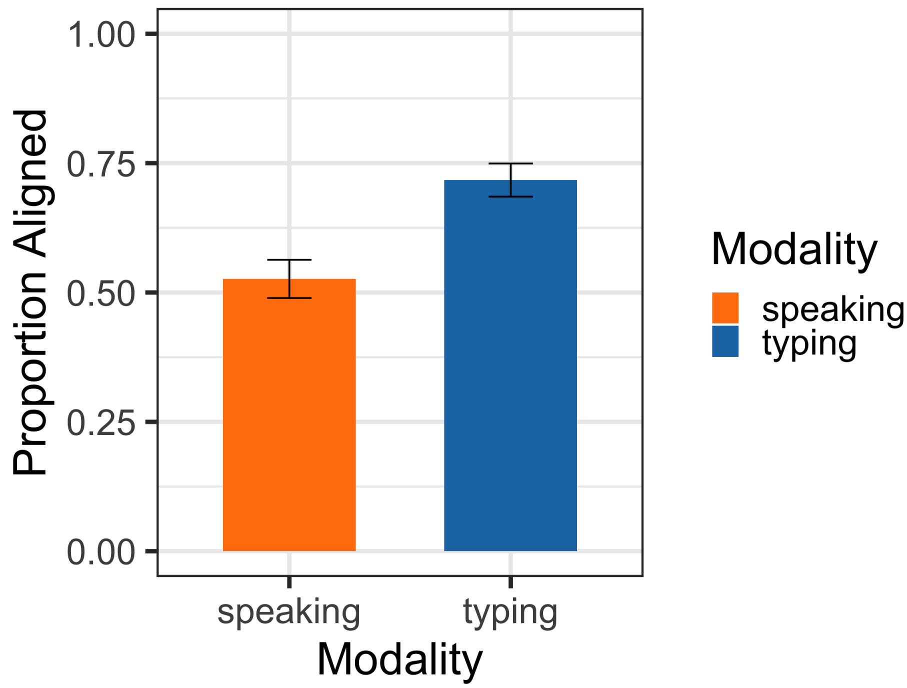


Figure 5.2: Comparison of entrainment between speaking vs. typing.

Then, I look at the main effect of proficiency and the interaction between production modality and proficiency on entrainment. I use three measurements to represent the proficiency, which are 1) a categorical variable of whether participants are L1 or L2 English speakers, 2) English dominance calculated from the self-reported proficiency, and 3) English dominance calculated from the verbal fluency task. The calculation of the two English dominance measures followed the same procedure in Chapter 3. I conducted a separate logistic mixed-effects model for each measure to predict entrainment based on modality and the proficiency measure, including random intercepts for both objects and participants.

First, in the analysis of entrainment between speaking and typing across L1 and L2 English speakers (Figure 5.3), the main effect of production modality is not significant ($\beta = 0.67$, $p = 0.10$). Although Figure 5.3 suggests that participants entrained more in typing than speaking, and L2 speakers entrained more than L1 speakers in typing, the result does not show a significant difference between L1 and L2 English speakers in entrainment ($\beta = 0.14$,

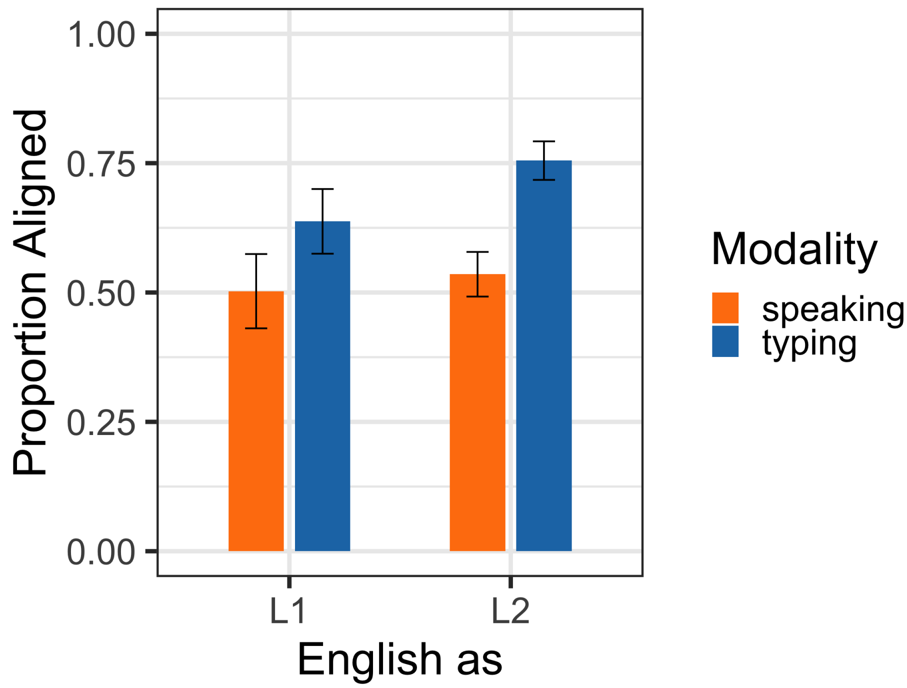


Figure 5.3: Comparison of entrainment between speaking vs. typing across Mandarin-English bilinguals speaking English as L1 and L2.

$p = 0.68$), nor the interaction effect ($\beta = 0.60$, $p = 0.22$). I also ran a logistic mixed-effects model including only the main effect of the two variables. The results show a significant difference between the two modalities ($\beta = 1.10$, $p < 0.001$), but no significant main effect of proficiency ($\beta = 0.46$, $p = 0.06$). These results indicate that a main effect of production modality exists, but its power may be limited by the unbalanced sample sizes between L1 and L2 English speakers. Although a significant interaction effect is not observed, Figure 5.3 shows a trend consistent with my hypothesis: less proficient participants (e.g., L2 English speakers) entrain more than more proficient participants (e.g., L1 English speakers) in typing, but their entrainment rate is reduced in speaking.

In the analysis of entrainment between speaking and typing with proficiency/dominance as a continuous variable, a robust main effect of production modality is observed in both the model using the self-reported proficiency measure ($\beta = 1.16$, $p < 0.001$) and the model using

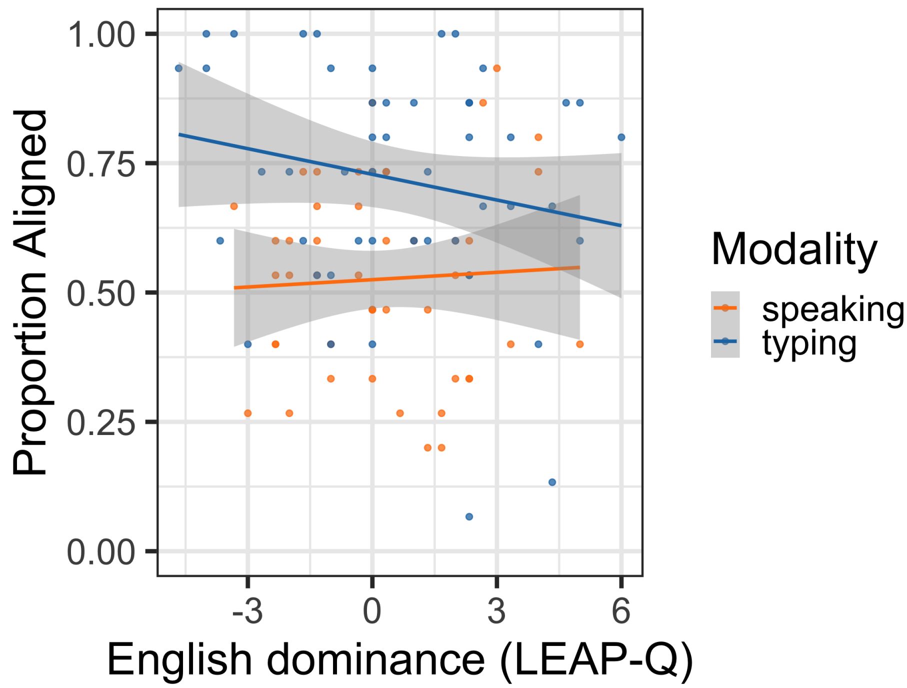


Figure 5.4: Correlation between entrainment and self-reported proficiency between speaking vs. typing

the verbal fluency measure ($\beta = 1.21$, $p < 0.001$). Figure 5.4 shows a non-significant result for the entrainment rate with self-reported proficiency as a predictor ($\beta = 0.028$, $p = 0.72$), and no interaction between production modality and self-reported proficiency ($\beta = -0.14$, $p = 0.16$), although a trend consistent with the hypothesis is observed. Similar to the results with the self-reported proficiency measure, the results with the VF measure still show that proficiency is not significantly correlated with entrainment ($\beta = -0.014$, $p = 0.63$), and no interaction is observed ($\beta = -0.04$, $p = 0.34$) (Figure 5.5)(see Table 5.3 for statistical details).

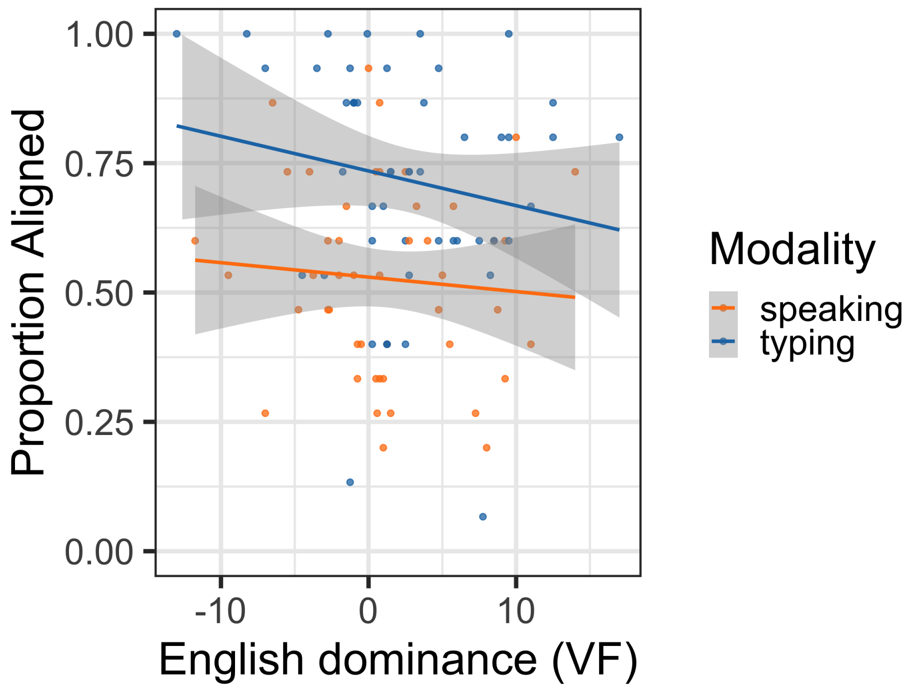


Figure 5.5: Correlation between entrainment and verbal fluency measured proficiency between speaking vs. typing

5.5 Discussion

Overall, the results show a robust effect of production modality on entrainment behavior: Mandarin-English bilinguals entrained more when typing the word to the partner compared to speaking. Although I did not observe the significant effect of proficiency on entrainment, the results of the three proficiency measures show a trend in typing similar to the conclusion from Chapter 3 that less proficient bilinguals entrain more. The statistical analysis does not reveal a significant interaction effect between production modality and proficiency on entrainment, but the figures show a trend consistent with expectations: less proficient participants entrain more than more proficient participants in typing, whereas in speaking, the entrainment rate of less proficient participants is reduced to a level similar to that of more proficient participants.

Predictors	***Estimate*** **(β)**	**SE**	z	p
modality	1.07	.23	4.63	$< .001$***
modality*English as L1 vs. L2				
modality(typing)	.67	.41	1.64	.10
English as L2	.14	.36	.41	.68
modality (typing)*English as L2	.60	.49	1.22	.22
modality*self-reported dominance				
modality (typing)	1.16	.24	4.94	$< .001$***
self-reported dominance	.03	.08	.36	.72
modality (typing)*self-reported dominance	-.14	.10	-1.42	.16
modality*VF dominance				
modality (typing)	1.21	.25	4.84	$< .001$***
VF dominance	-.014	.03	-.48	.63
modality (typing)*VF dominance	-.04	.04	-.95	.34

Note. β represents the estimate (regression coefficient) of the predictor in the model.

Table 5.3: Logistic mixed-effect model analyses of entrainment rate in relationship to modality, language proficiency/dominance in three measures, and their interactions

The results first indicate a crucial effect of production modality as I hypothesized. Mandarin English bilinguals tended to entrain more when producing words in typing. As I discussed above, production in typing may require fewer cognitive resources during the error monitoring process, as it allows for planning and revision, whereas error monitoring in speaking may be more cognitively demanding for initiating accurate production. Cognitive resources, then, are shifted toward other cognitively demanding processes, such as recruiting cognitive control to evaluate the rewards of lexical options. With this assumption, the entrainment in typing can be explained by the allocation of more cognitive resources to the control process in lexical entrainment. This explanation is also supported by the model from Chapter 4. In Chapter 4, the best-fitting model has $\beta = 1.5$, $\alpha_P = 1$, and $\alpha_\theta = 0.4$, where β controls the weight of the two mechanisms in determining entrainment, and the two α parameters represent the learning rates in each mechanism. It indicates that control of rewards takes more weight than the automatic priming process in lexical entrainment. Therefore, when producing language in typing, bilinguals may have more cognitive resources available for the control process, resulting in more entrainment. Entrainment behavior reflects changes in lexical selection, and thus, this result may also indicate the difference in the stage of lexical selection across modalities. Previous research on the comparison across modalities argues that modalities such as typing (writing) and speaking share the same processes of conceptualization and formulation, and diverge at the stage of sub-lexical processing and encoding. This conclusion illustrates that these modalities follow the same stages of language production, but there might be slight differences in the mechanisms recruited during the stage of lexical selection. While the selection processes in typing and speaking are in a similar time course, it is possible that a rational process primarily determines word selection in typing, whereas lexical selection in speaking relies more on an automatic process.

In terms of the results of proficiency, although the effect did not reach significance as observed in Chapter 3, it presents a similar trend that bilinguals who are less proficient tended to entrain more. One possible reason for this non-significant result might be that I only included Mandarin-English bilinguals in this study. In the previous study, bilingual populations vary across different languages, with the majority being Spanish-English bilinguals. The result also shows a significant proficiency effect when only Spanish-English bilinguals are included. Thus, it is possible that the significant result may only reflect the proficiency effect on entrainment among Spanish-English bilinguals, whereas the features of other bilingual populations may not be powerful enough to drive the result. The bilingual populations in the two studies differ significantly in many aspects. In terms of language dominance (i.e., English proficiency relative to the other language), participants from Experiment 1 in Chapter 3 are strongly dominant in English (mean of VF dominance = 5.21), but Mandarin-English participants from this study present a relatively weak dominance in English (mean of VF dominance = 1.96). That is because we recruited many participants who speak English as their first language (42 L1 English speakers and 25 L2 English speakers) in the previous study, whereas the majority of participants in this study speak English as a second language (29 L1 English speakers and 68 L2 English speakers). The discrepancy in English dominance and the sample size imbalance between L1 and L2 English speakers across the two studies may result in the failure to replicate the proficiency effect. Moreover, Spanish and English share many features in pronunciation and writing systems, whereas Mandarin employs a distinct writing system. These differences may cause discrepancies in other factors related to proficiency and entrainment behavior. For instance, Spanish has many cognates with English that facilitate Spanish speakers to produce English words (Rosselli et al., 2014). In addition, the task in this study included only 15 objects to control the task duration, which may have limited the variability and reduced the sensitivity of the entrainment calculation compared to the previous study.

Regarding the investigation of the interaction effect between modality and proficiency, no significant interaction effect on bilingual lexical entrainment was observed across all three English proficiency/dominance measures, but an expected trend was observed. One possible explanation for the absence of an interaction effect is that less proficient bilinguals may have even fewer cognitive resources available for the rational process in speaking. With reduced involvement of the control process in word selection, their entrainment rate decreases to a level similar to that of more proficient bilinguals, thereby eliminating the interaction effect. However, it is also possible that proficiency is independent of production modality. The entrainment behavior in typing and speaking, as well as the discrepancy between them, is not modulated by bilinguals' varying proficiency levels. There is no clear evidence supporting the relationship between production modality and bilingual proficiency; therefore, further research is needed to clarify how bilinguals of varying proficiency levels use languages across typing and speaking.

In conclusion, our results indicate that bilinguals exhibit more entrainment when producing language in typing compared to speaking. Although this study specifically focused on bilingual populations, the findings offer insight into potential differences in the mechanisms involved in lexical selection across modalities. This work represents an exploration of how production modality may influence the lexical selection process and bilingual language use.

It highlights the need for further research into the cognitive mechanisms in lexical selection and individual differences among bilingual populations with production modality taken into account.

Chapter 6

General Discussion

This dissertation aims to answer questions about how bilinguals behave in lexical entrainment. Lexical entrainment, as a special phenomenon in conversation, indicates the change of lexical selection in the process of language production. What factors bring an effect on lexical entrainment? What mechanisms control this behavior? Recent research has been investigating these questions, but the answers remain unclear, with mixed results. Language production in bilingual speakers is diverse, depending on their language backgrounds, proficiency, language environment, and other factors. Therefore, I focused on bilinguals and investigated their performance in lexical entrainment to understand individual differences in lexical entrainment. Meanwhile, how bilinguals select and produce words in conversation has also been discussed.

Proficiency is one factor that has been investigated in many studies. It is a crucial factor in lexical entrainment because it indicates the audience-specific effect in this phenomenon. When speakers talk to partners who are less proficient in the language, such as bilinguals speaking the language as a second language, the speakers adjust their word choices by aligning more with the partners' words. This result is argued to support the theory of a rational process in lexical entrainment. However, previous studies have presented mixed results on how proficiency affects lexical entrainment when bilinguals take the role of the speaker and select words. The first study of this dissertation contributes to answering this question. I tested entrainment behaviors in bilinguals at different proficiency levels by conducting picture matching and naming tasks in both English and Mandarin, and measured proficiency with both subjective and objective measures. I found that bilinguals who are less proficient in English entrained more in the task. This result demonstrates that proficiency is also crucial for bilingual speakers to determine their entrainment behavior. Moreover, the different results from the two methods of measuring proficiency also highlight the importance of using an objective measurement of proficiency in bilingual research. Furthermore, this result also emphasizes individual differences in lexical entrainment among bilinguals. These individual differences not only reflect the diversity within the bilingual population but also suggest possible differences in the strategies or mechanisms that bilinguals use in conversation. For instance, bilinguals who are less proficient in one language may use more entrainment to learn from their partner's language use. It is also possible that they strategically rely on their partner to ensure accurate word selection in conversation so that they reduce the effort of evaluating alternative lexical options. Although the result for Mandarin did not show a significant proficiency effect, it suggests a possible difference in lexical entrainment between a first language (L1) and a second language (L2). By comparing the results for English and Mandarin, I found that Mandarin-English bilinguals exhibited more entrainment in English (L2) than in Mandarin (L1), although this comparison was between groups and needs further evidence from future research.

In examining the effect of language switching on bilingual lexical entrainment, I observed a switching effect in English: Mandarin-English bilinguals entrained less in English after switching from Mandarin (L1). In contrast, the switching effect in Mandarin showed an opposite pattern, where greater Mandarin entrainment was observed after switching from English (L2). I hypothesized that the inhibitory theory may explain the increase in Mandarin entrainment after switching. Mandarin-English bilinguals must suppress the activation of their more dominant language (Mandarin) in order to produce the less dominant one (English), so the reactivation of Mandarin becomes more difficult and requires more cognitive effort. As a result, bilinguals employ increased lexical entrainment as a strategy to support accurate word production when Mandarin is not well-activated. On the other hand, this hypothesis is not adequate to explain the decreased English entrainment after switching, because the inhibition should have also affected English and resulted in more entrainment behaviors. I also discussed the other possible explanation that an individual's preference for entrainment may remain consistent even when language switching occurs. The preference for entrainment may be determined by the language used at first and then carried over to other languages. This could explain the interaction effect between switching and language observed in bilingual lexical entrainment. Future studies are necessary to test these explanations.

The results on proficiency in bilingual lexical entrainment appear to support the theory of rational processes in lexical entrainment, because speakers engage in the process of evaluating whether the word is better for the communicative purpose given the proficiency level of the bilingual, no matter whether the bilingual is the speaker or the partner in a conversation. However, the theory of automatic priming in lexical entrainment, although always being assumed in lexical entrainment, has less evidence to support its occurrence. To examine the automatic and rational processes underlying lexical entrainment, I developed a computational model to explore the mechanisms involved in this phenomenon. Taking the rate–distortion control (RDC) model as the basic framework, the model clearly separates the two mechanisms into two components and updates them respectively with policy gradient methods applied. The model successfully predicted the empirical results, demonstrating its reliability. The successful prediction also serves as evidence for the hypothesis that both automatic priming and the rational process of control (conceptual pact) engage in lexical entrainment. It demonstrates that the two mechanisms must work together, with manipulations of the learning rate parameters in each mechanism and the parameter that weighs the two mechanisms. These parameters can be defined differently under different circumstances. For instance, proficiency can be linked to the three parameters if we explain the effect of proficiency computationally. Based on the empirical finding that less proficient bilinguals tend to entrain more, proficiency may be related to both automatic and rational mechanisms of entrainment, and may even influence the parameter of weighing the two mechanisms. With the assumption that less proficient bilinguals may engage in learning language use through entrainment, this learning process may reflect a greater involvement of reward evaluation in the model, indicating a higher learning rate parameter α_θ for the reward component. In terms of the relationship between proficiency and the automatic process, low proficiency may indicate weaker semantic connections between semantic concepts and lexical items. As a result, the activation of words from the partner may be strong enough to override alternative lexical options. This process can be operationalized by a higher learning rate parameter α_P within the automatic mechanism. Furthermore, proficiency may also relate to how bilinguals weigh the two mechanisms. Less proficient bilinguals may rely more on rational processes than automatic priming for the purpose of language learning, which can be operationalized as a higher β parameter to increase the weight of the reward component in the model. Overall, bilinguals at different levels of proficiency may exhibit the three parameters in varying numbers, which may capture bilinguals' individual differences in applying these mechanisms. These differences can be explored in the next step of research.

Lastly, I investigated the effect of production modality on lexical entrainment by comparing the entrainment behavior of Mandarin-English bilinguals when typing versus speaking. Because speaking is argued to be more difficult and more anxiety-inducing than typing (and writing) for L2 speakers, it places greater demands on cognitive resources. Then, fewer resources are available for bilinguals to engage in the rational process of evaluating word options in lexical selection. Therefore, I hypothesized that modality brings about differences in entrainment, with bilinguals showing less entrainment in speaking due to the reduced involvement of rational processes under limited cognitive resources. According to the results, Mandarin-English bilinguals entrained more in the typing task than in the speaking task, while no significant main effect of proficiency or interaction effect of modality and proficiency was observed. This result highlights the importance of modality in language production, particularly in interactive contexts. I explained the main effect of production modality with the argument that the rational process of control engages more in typing, leading to more entrainment, while the automatic process plays a larger role in speaking. Because speaking is assumed to be more difficult for bilinguals, I argued that bilinguals would spend fewer cognitive resources on evaluation and selection, allowing more resources to be allocated to other steps that require more effort. For an easier modality of producing words, such as typing, more cognitive resources may be available to evaluate which word is better to use in conversation. Besides that, although I failed to find a significant effect of proficiency on entrainment as I had previously, a similar trend was still observed to support the conclusion of the previous study. This result also suggests that further research is necessary to investigate the impact of proficiency and individual differences on bilingual lexical entrainment.

In future work, the effect of proficiency still needs more evidence to support its role in determining entrainment behaviors. Combining the computational modeling method, proficiency can be operationalized as the three parameters in the model, and the behavioral results of bilingual lexical entrainment can be modeled through the manipulation of these parameters. In this way, the effect of proficiency can be mapped onto the two mechanisms respectively so that the model can explain why proficiency brings about such an effect and if any other components or parameters are needed to account for the gap in why its effect does not always occur in entrainment. The model can also be refined and developed to account for the effects observed in research, such as proficiency and production modality.

Moreover, the investigation of bilingual lexical entrainment has mainly relied on experimental tasks such as picture matching and naming, which are not as natural as real conversations. Although entrainment behavior is robustly observed across most studies, entrainment in real conversations can be far more complex. In addition to proficiency and production modality, many other factors also contribute to entrainment, such as social dynamics, contextual factors, and different communicative goals. Although these factors were not discussed in this dissertation, they are worth exploring further by combining our current knowledge of the mechanisms and factors in bilingual lexical entrainment with both computational and empirical approaches.

The results of this dissertation contribute to a clearer understanding of how mechanisms operate in lexical entrainment using a computational modeling method. This is a first exploration of a new method in studying lexical entrainment. It also provides evidence for the involvement of both automatic and rational processes from an aspect other than empirical results. The investigation on proficiency in bilingual lexical entrainment highlights the individual differences in bilinguals' entrainment behavior. Finally, the comparison of entrainment between typing and speaking provides insights into how word selection changes in different modalities.

Bibliography

Abel, J. and M. Babel (2017, September). Cognitive Load Reduces Perceived Linguistic Convergence Between Dyads. *Language and Speech 60*(3), 479–502.

Baus, C., K. Strijkers, and A. Costa (2013). When does word frequency influence written production? *Frontiers in Psychology 4*.

Beatty-Martínez, A. L. and P. E. Dussias (2017, August). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language 95*, 173–189.

Beatty-Martínez, A. L., C. A. Navarro-Torres, P. E. Dussias, M. T. Bajo, R. E. Guzzardo Tamargo, and J. F. Kroll (2020, June). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition 46*(6), 1022–1047. Num Pages: 1022-1047 Publisher: American Psychological Association (US).

Biber, D. (2009). Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. *Cambridge handbook of literacy*, 75–91.

Bonin, P. and M. Fayol (2000, July). Writing words from pictures: What representations are activated, and when? *Memory & Cognition 28*(4), 677–689.

Branigan, H. P., M. J. Pickering, J. Pearson, J. F. McLean, and A. Brown (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition 121*(1), 41–57.

Brennan, S. E. and H. H. Clark (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition 22*(6), 1482–1493.

Cai, Z. G., Z. Sun, and N. Zhao (2021). Interlocutor modelling in lexical alignment: The role of linguistic competence. *Journal of Memory and Language 121*, 104278.

Chaouch-Orozco, A., J. González Alonso, and J. Rothman (2021, March). Individual differences in bilingual word recognition: The role of experiential factors and word frequency in cross-language lexical priming. *Applied Psycholinguistics 42*(2), 447–474.

Cheng, Y., E. K. Horwitz, and D. L. Schallert (1999, September). Language Anxiety: Differentiating Writing and Speaking Components. *Language Learning 49*(3), 417–446.

Cho, M. (2018, February). Task complexity, modality, and working memory in L2 task performance. *System 72*, 85–98.

Collins, A. G. E. and M. J. Frank (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*.

Collins, A. M. and E. F. Loftus (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review 82*(6), 407–428.

Costa, A., M. Miozzo, and A. Caramazza (1999, October). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection? *Journal of Memory and Language 41*(3), 365–397.

Costa, A., M. J. Pickering, and A. Sorace (2008, June). Alignment in second language dialogue. *Language and Cognitive Processes 23*(4), 528–556.

Costa, A. and M. Santesteban (2004, May). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language 50*(4), 491–511.

De Leeuw, J. R., R. A. Gilbert, and B. Luchterhandt (2023, May). jsPsych: Enabling an Open-Source CollaborativeEcosystem of Behavioral Experiments. *Journal of Open Source Software 8*(85), 5351.

Declerck, M. and A. M. Philipp (2015, December). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review 22*(6), 1630–1645.

Dell, G. (2014). Word production: behavioral and computational. In *The Oxford Handbook of Language Production*, pp. 49.

Dell, G., M. Schwartz, N. Martin, E. Saffran, and D. Gagnon (1997, October). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review 104*, 801–838.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review 93*(3), 283–321. Publisher: American Psychological Association.

Futrell, R. (2023, September). Information-theoretic principles in incremental language production. *Proceedings of the National Academy of Sciences 120*(39), e2220593120.

Garrod, S. and A. Anderson (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition 27*(2), 181–218.

Green, D. W. (1998, August). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition 1*(2), 67–81.

Gullifer, J. W., X. J. Chai, V. Whitford, I. Pivneva, S. Baum, D. Klein, and D. Titone (2018, August). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia 117*, 123–134.

Hoshino, N. and G. Thierry (2011, January). Language selection in bilingual word production: Electrophysiological evidence for cross-language competition. *Brain Research 1371*, 100–109.

Ivanova, I., H. Branigan, J. McLean, A. Costa, and M. Pickering (2021, October). Lexical Alignment to Non-native Speakers. *Dialogue & Discourse 12*(2), 145–173.

Ivanova, I., D. C. Hernandez, and A. Atiya (2025, March). Automatic and strategic components of bilingual lexical alignment. *Cognition 256*, 106046.

Jacobs, C. L. (2015). Predictions for self-priming from incremental updating models unifying comprehension and production. In *Proceedings of the 6th Workshop on Cognitive Modeling and Computational Linguistics*, Denver, Colorado, pp. 1–8. Association for Computational Linguistics.

Jasmin, K. and D. Casasanto (2012). The QWERTY Effect: How typing shapes the meanings of words. *Psychon Bull Rev*.

Kaschak, M. P., T. J. Kutta, and J. L. Jones (2011, December). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review 18*(6), 1133–1139.

Koranda, M. J., M. Zettersten, and M. C. MacDonald (2022). Good-Enough Production: Selecting Easier Words Instead of More Accurate Ones. *Psychological Science*.

Kroll, J. F., S. C. Bobb, M. Misra, and T. Guo (2008, July). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica 128*(3), 416–430.

Kroll, J. F. and A. M. B. d. Groot (Eds.) (2005). *Handbook of bilingualism: psycholinguistic approaches.* Oxford ; New York: Oxford University Press.

Lai, L. and S. J. Gershman (2021). Policy compression: An information bottleneck in action selection. In *Psychology of Learning and Motivation*, Volume 74, pp. 195–232. Elsevier.

Levelt, W. J. M. (1993). *Speaking: From Intention to Articulation.* The MIT Press.

Levelt, W. J. M., A. Roelofs, and A. S. Meyer (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences 22*(01).

Lewis, R. L., A. Howes, and S. Singh (2014, April). Computational Rationality: Linking Mechanism and Behavior Through Bounded Utility Maximization. *Topics in Cognitive Science 6*(2), 279–311.

Lupker, S. J. (1979, November). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition 7*(6), 485–495.

MacIntyre, P. D. and R. C. Gardner (1994, June). The Subtle Effects of Language Anxiety on Cognitive Processing in the Second Language. *Language Learning 44*(2), 283–305.

Marian, V., H. K. Blumenfeld, and M. Kaushanskaya (2007, August). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research 50*(4), 940–967.

Meuter, R. F. and A. Allport (1999, January). Bilingual Language Switching in Naming: Asymmetrical Costs of Language Selection. *Journal of Memory and Language 40*(1), 25–40.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language 30*(1), 69–89.

Meyer, A. S. (1996, August). Lexical Access in Phrase and Sentence Production: Results from Picture–Word Interference Experiments. *Journal of Memory and Language 35*(4), 477–496.

Meyer, D. E. and R. W. Schvaneveldt (1971, October). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology 90*(2), 227–234.

Misra, M., T. Guo, S. C. Bobb, and J. F. Kroll (2012, July). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language 67*(1), 224–237.

Neely, J. H. (1976, September). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition 4*(5), 648–654.

Perret, C. and M. Laganaro (2012, January). Comparison of Electrophysiological Correlates of Writing and Speaking: A Topographic ERP Analysis. *Brain Topography 25*(1), 64–72.

Pickering, M. J. and S. Garrod (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences 27*(02).

Pickering, M. J. and S. Garrod (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation 4*(2-3), 203–228.

Polinsky, M. and G. Scontras (2020, January). Understanding heritage languages. *Bilingualism: Language and Cognition 23*(1), 4–20.

Qu, Q., Q. Zhang, and M. F. Damian (2016, August). Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain and Language 159*, 118–126.

Rosselli, M., A. Ardila, M. B. Jurado, and J. L. Salvatierra (2014, December). Cognate facilitation effect in balanced and non-balanced Spanish–English bilinguals using the Boston Naming Test. *International Journal of Bilingualism 18*(6), 649–662.

Scheepers, C. (2003, October). Syntactic priming of relative clause attachments: persistence of structural configuration in sentence production. *Cognition 89*(3), 179–205.

Schriefers, H., A. Meyer, and W. Levelt (1990, February). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language 29*(1), 86–102.

Simon, H. A. (1955, February). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics 69*(1), 99.

Son, M. (2022, November). The role of modality and working memory capacity in L2 production. *Language Teaching Research*, 136216882211352.

Speer, R. (2022). rspeer/wordfreq: v3.0.

Strijkers, K., A. Costa, and G. Thierry (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral cortex 20*(4), 912–928.

Suffill, E., T. Kutasi, M. J. Pickering, and H. P. Branigan (2021). Lexical alignment is affected by addressee but not speaker nativeness. *Bilingualism: Language and Cognition 24*(4), 746–757.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: an introduction.* Adaptive computation and machine learning. Cambridge, Mass: MIT Press.

Tobar-Henríquez, A., H. Rabagliati, and H. P. Branigan (2020). Lexical entrainment reflects a stable individual trait: Implications for individual differences in language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition 46*(6), 1091–1105. Publisher: American Psychological Association.

Trebits, A. (2016, April). Sources of Individual Differences in L2 Narrative Production: The Contribution of Input, Processing, and Output Anxiety. *Applied Linguistics 37*(2), 155–174.

Zhang, D. and J. Nicol (2022, July). Lexical alignment in second language communication: evidence from a picture-naming task. *Language, Cognition and Neuroscience 37*(6), 732–749.