Title

Individual differences in phonology: the realization of stem-final coronal obstruents in Korean

Author

Jo, Jinyoung

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Individual differences in phonology: the realization of stem-final coronal obstruents in Korean

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Linguistics

by

Jinyoung Jo

2024

ABSTRACT OF THE DISSERTATION

Individual differences in phonology: the realization of stem-final coronal obstruents in Korean

by

Jinyoung Jo

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2024

Professor Kie R. Zuraw, Chair

Individual speakers' speech patterns differ from one another, despite presumably similar language input. What are the sources of this individual variability? In this dissertation, I explore sources of individual differences in pronunciation of coronal obstruents (/s/, /t^h/, /c^h/, /c/) at the ends of nouns in Korean. Research in linguistics has traditionally focused on patterns that are shared across individuals of a speech community, and speakers who deviate from the group-level generalization were treated as outliers. Research on speaker variation has primarily been in sociolinguistics, which investigates how demographic and stylistic variables affect speech. However, it has recently been recognized that individuals with homogeneous sociolinguistic background also exhibit different linguistic behavior, and there is increasing interest in how and why people vary.

One view of language learning holds that individuals who received roughly the same linguistic input as a child converge on the same linguistic knowledge (grammar), and that inter-speaker variation arises only due to random noise or extra-grammatical factors such as differ-

ences in how speakers perform the experimental task (Chomsky, 1975). In a longitudinal study involving two experiments (Chapters 4-5), I show that speaker variation cannot be dismissed as random noise or as an artifact of the task. Speakers were self-consistent in their pronunciation for the target words in the two experiments, which employed different tasks and were separated by 1.5 years. Further, the results of *k*-means clustering analysis in each experiment indicate that distinct speaker types indeed exist, rejecting the hypothesis that the speakers are best understood as a single group.

That individual speakers should differ from one another, despite presumably similar language input, is an interesting puzzle. In Chapter 5, I explore potential sources of the individual variability, including differences in the lexicon, grammar, cognitive traits, and pressure toward normative pronunciation. Some of these factors reliably predicted individual differences, indicating that variation is systematic.

In Chapter 6, I model individual speakers' production patterns within the framework of Maximum Entropy Grammar (Smolensky, 1986; Goldwater & Johnson, 2003), employing Optimality-Theoretic constraints (Prince & Smolensky 1993, 2004). I provided the model with corpus frequencies as the learning data (obtained in Chapter 3) and varied the bias terms for the constraints across different learning simulations. This variability in bias terms results in different outputs from the model, replicating the individual speakers' production of obstruent variants observed in the experiment.

The dissertation of Jinyoung Jo is approved.

Bruce P. Hayes

Jongho Jun

Megha Sundara

Kie R. Zuraw, Committee Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation would not have been possible without the guidance and support I received from many people. Words cannot fully express my gratitude, but I will make my best attempt.

First, I am sincerely thankful to my advisor, Kie Zuraw. I feel incredibly fortunate to have had her as my dissertation advisor. She provided me with the most stimulating ideas and offered concrete, sensible feedback on many aspects of my projects. It was she who sparked my interest in individual differences in phonology. I enjoyed every single meeting with her, which made the dissertation journey all the more enjoyable. I also appreciated our conversations about life outside of linguistics; her advice on non-academic matters was just as thoughtful.

Bruce Hayes always encouraged me to think about the bigger picture in my projects. I was often amazed at how insightful his advice was. While working as his research assistant on our Balinese project, I found him to be an inspiring example of the researcher and teacher I want to become. I also enjoyed learning English expressions from him, which he refers to as "mid-century idioms".

I had the privilege of collaborating with Megha Sundara on several projects related to child language acquisition, and I am deeply grateful to her for supporting my continued interest in language acquisition at UCLA. She always helped me find direction when I felt lost in my projects. My projects benefited greatly from her concrete suggestions for research methods, analysis, and write-ups. I really appreciate her invaluable mentorship over the years.

A special thanks goes to Jongho Jun, who offered valuable comments and feedback on the many different Korean phonological processes I have studied. They were crucial in helping me get the details right about the processes I investigated in this dissertation. I am so glad I asked him to be on my dissertation committee. I am especially grateful to him, as I wouldn't have been able to join UCLA Linguistics without his support during my MA studies at Seoul National University (SNU). I truly enjoyed my time there, taking his classes, attending his office hours.

I also thank other professors I met at UCLA and at SNU. I am grateful to Sun-Ah Jun, with whom I enjoyed chatting in Korean. I often walked into her office with questions about K-ToBI, and we would end up discussing all sorts of things, both linguistic and non-linguistic. Her warmth and support provided me with comfort in the department. I am also thankful to Jae-Young Lee, my MA advisor at SNU, who always encouraged me to dream big. He believed in my work and consistently supported my learning. Eon-Suk Ko deserves a special mention for introducing me to the world of research. I remember walking into her office to ask what a t-test was, as she often recalls, and years later, finding myself equipped with a variety of research skills after working on her projects. I am continually impressed by her passion for research.

My friends and colleagues played a crucial role in helping me keep going. I am especially grateful for my cohort at UCLA: Jake Aziz, Phillip Barnett, Colin Brown, Noah Elkins, Jennifer Kuo, and Mia Teixeira, as well as DSPs Tyler Carson, Angelica Pan, and Shuangqi Joy Wu. They were the first friends I made in LA, my teachers for various problem sets (especially in syntax and semantics), and my comrades. I am particularly glad to have met Jennifer, with whom I felt comfortable sharing my research and personal experiences. She was always there to listen to me and offer help. I am also thankful for Hironori Katsuda and Jian-Leat Siah for our friendship.

I am also grateful for my Korean linguist friends. We studied linguistics together in Korea, and many of them continue to pursue careers in the field. I feel very fortunate to have had Juyeon Cho, Sanghee Kim, and Heemin Park in my Master's program cohort, as we all shared a passion for linguistic research. Juyeon and Sanghee are both very good friends and research collaborators, and I hope our friendship and collaboration will continue in the future. I took phonology courses with Sarang Jeong, Hyunjung Joo, and Ahram Kim, who are also good friends outside of linguistics. We struggled together while solving OT problem sets, often creating silly constraints without even realizing how silly they were. These are the kinds of memories I can look back on and smile. I am especially grateful for Sarang Jeong, who was a very close friend in Korea and continues to be one here in the States. She has provided endless support in both my academic and personal life, and I think that pursuing linguistics was a good decision because it led me to meet her. I would also like to thank Sujin Oh, Sarah Hye-yeon Lee, Jeonghwa Cho, and

Yeonseob Lee for sharing this experience with me during our graduate studies. Additionally, I must thank my non-linguist friends, who supported my goals, patiently listened to my research-related complaints, and always welcomed me when I visited Korea.

I am deeply indebted to my research participants, many of whom were also my friends, for their time and patience for my longitudinal study. This dissertation has benefited from audiences at UCLA Phonology Seminar, Annual Meeting on Phonology 2022, 2023 and Linguistic Society of America 2024, Stanford Phonetics and Phonology Workshop.

Finally, this dissertation would not have been possible without the support of my family. I thank my parents and my brother for their endless encouragement. My parents always believed in me and supported me in every way possible throughout my life. They were devoted to babysitting my son as the pace of this dissertation accelerated. Without their help, this dissertation would not have come into being. I also thank my in-laws for their support of my career goals and their commitment to babysitting as well. My son, Ian Choi, has brought me a type of joy I will never find elsewhere. Of course he didn't care about the ups and downs of my dissertation and kept me busy with his needs, reminding me that there's more to life than the dissertation. Lastly, my deepest gratitude goes to my husband, Sean Choi, for his support and advice, both practical and emotional. He has been supportive of my career goals. He helped me stay encouraged when journals rejected my work, solved many of the technical problems I encountered in my research (which I dared not tackle on my own), dragged me outside for walks to help me take breaks from my work, and made coffee every morning to start our day. He has also been a wonderful collaborator. I am grateful to have someone so close to me with whom I can share every aspect of my dissertation, research, and academic life, and he listened to me patiently. It must not have been easy to be a husband of someone doing a Ph.D. I thank him for being willing to take on that role.

This dissertation was supported by the UCLA Dissertation Year Fellowship, Overseas Doctoral Program Scholarship granted by Kwanjeong Educational Foundation, as well as Pony Chung Humanities Scholarship awarded by Pony Chung Foundation.

VITA

2011–2015 B.A. (English Language and Literature), Seoul National University

2015–2017 M.A. (English Language and Literature), Seoul National University

PUBLICATIONS

Jo, J., Sundara, M., & Breiss, C. (to appear). Predicting language outcomes at 3 years using individual differences in morphological segmentation in infancy. *Infant Behavior and Development.*

Jo, J., & Sundara, M. (to appear). Remote collection of language samples from 3-year-olds. *Journal of Child Language.*

Jo, J. (2024). Korean vowel harmony has weak phonotactic support and has limited productivity. *Phonology.*

Choi, S., & Jo, J. (2024). Leveraging linguistic signals to improve machine learning models for detecting ChatGPT usage on exams. *Artificial Intelligence, Machine Learning, Convolutional Neural Networks and Large Language Models.* De Gruyter.

Jo, J.*, Cho, J.*, Kim, S. J.*, & Jun, S.-A. (2024). What prosody does when morphosyntax is absent: the case of Korean relative clauses. *Proceedings of the Linguistic Society of America, 9*(1). (* = co-first authors)

Jo, J., & Choi, S. (2023). Analysis of plagiarism via ChatGPT on domain-specific exams. Proceedings of the 2023 Congress in Computer Science, Computer Engineering, & Applied Computing.

Breiss, C., & Jo, J. (2023). SIGMORPHON–UniMorph 2022 Shared Task 0: Cognitively Plausible Morphophonological Generalization in Korean. *Proceedings of the 20th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology.*

Ko, E.-S., Jo, J., On, K.-W., & Zhang, B.-T. (2020). Introducing the Ko Corpus of Korean mother-child interaction. *Frontiers in Psychology, 11.*

Jo, J., & Ko, E.-S. (2018). Korean mothers attune the frequency and acoustic saliency of sound symbolic words to the linguistic maturity of their children. *Frontiers in Psychology, 9.*

Jo, J. (2017). Learning bias of phonological alternation in children learning English. *Studies in Phonetics, Phonology and Morphology 23*(2), 261-292.

Jo, J. (2016). Native and non-native English speakers' production of voiceless stops: Effects of lexical stress and prosodic domain on VOT. *English Studies, 36*, 111-128.

Jo, J. (2015). Native and non-native speakers' perception of fluency in L2 speech. *SNU Working Papers in English Language and Linguistics, 13,* 40-62.

CHAPTER 1

Introduction

For the past few decades, phonological research has increasingly used quantitative methods (e.g. Sampson, 2005; Janda, 2013). Quantitative approaches have improved the empirical reliability and robustness of study results, since responses obtained from multiple speakers are resistant to noise that may arise when probing only one speaker's grammar. Because these studies focus on describing and explaining patterns that are shared across different individuals of a speech community, individual speakers who deviate from the phonological generalization at the group level are often treated as outliers.

While research in individual differences has been receiving more attention in the field of psychology, there has traditionally been a long division in psychology between experimental psychologists and correlational (differential) psychologists (Cronbach, 1957; Eysenck, 1966). Experimental psychologists manipulate conditions and observe their consequences while tightly controlling situational variables, environmental factors that could affect participants' performance. They focus on differences between treatment and non-treatment group, and tend to treat individual variation as error variance. In comparison, correlational psychologists investigate relationship among naturally occurring variables, and treat individual variation as important consequences of biological or social factors. In recent decades, however, the two approaches have increasingly informed each other, allowing researchers to address questions that would not be answerable with purely experimental or correlational approaches (Cronbach, 1975; Eysenck, 1997; Revelle et al., 2011).

Within differential psychology, a subfield dedicated to studying individual variation, research investigates how and why individuals differ from one another. This field encompasses a wide range of interests, including variations in affect, behavior, cognition, and motivation, and examines the contributions of genetic and environmental factors to these individual differences. Additionally, differential psychology has practical applications, such as predicting individuals' behavior in occupational settings based on personality and ability assessments (Barrick & Mount, 1991; Mount et al., 2005). As such, psychological research places a greater emphasis on understanding individual differences across various contexts.

One area of linguistic research that puts relatively greater emphasis on individual differences is the sociolinguistic literature, the variationist approach in particular. Research in this field focused on how sociolinguistic variables such as age, gender, and socioeconomic status affect linguistic behaviors of different speakers that belong to the same speech community (Fischer, 1958; Labov, 1963, 1966/2006, 1972). However, it has also been widely recognized that individuals that do not differ along those sociolinguistic variables also exhibit different production and perception behaviors, for variables that are not believed to carry social meaning. Research on individual variation has reported speaker variation in phonetic implementation of speech sounds (voice onset time in Newman, 1997 and Allen et al., 2003; consonant articulation in Mielke et al., 2016; vowel articulation in Johnson et al., 1993; vowel formants in Peterson & Barney, 1952; coarticulation in Zellou, 2017), and how often individuals apply a variable phonological rule (Zuraw, 2000; Bayles et al., 2016; Davidson et al., 2021). Yu & Zellou (2019) provides a review of studies on individual differences in phonological processing.

It has been also noted in the literature that individual variation is not just random noise or measurement errors that arise when the same individuals are tested multiple times; rather, it is highly structured and has meaningful implications for the grammar (Chodroff & Wilson, 2017; Wade et al., 2021). One systematic aspect of individual differences in phonological grammar is that a significant portion of the variance can be explained by factors such as the amount or type of language input that one receives, personality traits and cognitive processing styles. In addition, although individual speakers exhibit differences in phonetic implementation of speech sounds or in how frequently one applies a variable phonological rule, we still find some consistent pattern across individuals. For example, Chodroff & Wilson (2017) reported that although native English speakers' production of voice onset time (VOT) for aspirated stops are different from speaker to speaker, the pattern in which [k^h] is realized with a longer VOT than [p^h] was observed within individuals (Zlatin, 1974; Koenig, 2000; Newman, 2003; Theodore et al., 2009).

Another aspect of inter-speaker variability that attests to its systematicity is that speakers are self-consistent in their phonetic or phonological behavior when tested multiple times. If the differences among individuals were random, the across-speaker variability in one study would not be replicated in another study conducted with the same set of speakers. Based on my understanding, only a very small number of studies across all areas of linguistics directly examined whether speakers are consistent across different testing sessions (VOT imitation in Wade et al., 2021; cue weighting in Kong & Edwards, 2016; syntactic judgment in Han et al., 2016).[1] A similar point has been made by James et al. (2018) and Durvasula (2024) that studies have rarely tried to replicate the individual differences. Therefore, more research is needed to investigate whether speakers are self-consistent across different occasions to test whether individual variation is more than just chance fluctuations or random noise.

Given the prevalence of individual variation in phonological behaviors, it is important to understand the nature of such variation. First, as discussed above, we must ask: does it reflect mere random noise, or is it observed in a reliable and systematic manner? Second, what are the sources of across-speaker differences? Furthermore, if the differences are meaningful and systematic, we need to account for the variation in the formal apparatus of phonology. In the remainder of this chapter, I first lay out why it is important to study individual differences, and then review previous approaches to investigating sources of inter-speaker variation and modeling phonological variation.

[1] See also James et al. (2018) which investigated subjects' consistency across different test items (rather than different testing sessions) to investigate whether stable (and thus meaningful) individual differences exist. They report a low correlation in the participants' responses between one set of items and another, and attribute the lack of consensus among studies examining individual differences in syntactic processing to the low consistency.

1.1 Why should we investigate individual differences?

Investigating individual differences is important in the following respects. First, when researchers investigate a variable phonological process by calculating the rate of rule application, aggregated results might obscure the precise pattern of the variation. For instance, imagine a situation in which the rate of application of a phonological process is 70%. There are at least two different scenarios with respect to how the grammar should look like: (i) 30% of the speakers never apply the rule and the other 70% always apply the rule, in which case grammars do not need to produce variation; rather, one can simply posit two different categorical grammars for the speech community, and (ii) most speakers apply the rule about 70% of the time. In this respect, investigating across-speaker variation could help understand whether and how much within-speaker variation there is.[2] Under the scenario (i), each speaker shows invariant behavior, hence no intra-speaker variation. In (ii), we would expect to see both application and non-application of the rule within an individual. Previous studies have also noted that aggregating across many speakers' responses may obscure the fact that different individuals can possess phonological grammars that differ from each other to a substantial degree (Zsiga & Gouskova, 2006; de Lacy, 2009; Bennett & Braver, 2020). In cases where different speakers exhibit opposite behaviors concerning the effect of interest, the effect would not even be observable when analyzed in aggregate (Zee et al., in preparation). Thus, it is crucial to investigate a variable phonological process at the individual level to describe and explain the variation more accurately.

Second, investigating inter-speaker variation can have implications for various other linguistic phenomena such as phonologization (Yu, 2021) and origins of sound change (see Stevens & Harrington, 2014 for a review). For instance, it has been argued that sound change is initiated by individual differences in articulation (Baker et al., 2011), differences in how individuals link perception and production (Beddor, 2009, 2012), and variability in cognitive processing, which is in turn related to individuals' social traits (Yu, 2013).

[2] See, for instance, Davidson et al. (2021) for recent work that investigated whether it is across-speaker variation or within-speaker variation that underlies a variable phonological process.

Last, a systematic investigation of individual differences could challenge the assertion that all learners of a language converge on the same grammar, often put forth by proponents of Universal Grammar (UG). Chomsky (1975) claims that "individuals in a speech community have developed essentially the same language" and that "[t]his fact can be explained only on the assumption that these individuals employ highly restrictive principles that guide the construction of the grammar." However, the assumption that language acquisition is uniform across all individuals lacks empirical evidence. Instead, numerous quantitative studies have shown differences across individuals in various areas of linguistics, as mentioned earlier (Peterson & Barney, 1952; Johnson et al., 1993; Newman, 1997; Allen et al., 2003; Han et al., 2007; Bayles et al., 2016; Han et al., 2016; Mielke et al., 2016; Zellou, 2017; Yu & Zellou, 2019; Davidson et al., 2021; Zee et al., in preparation).

Given the assumption that individuals arrive at the same grammar, the strongest stance of UG would predict that the individual differences, if any, must stem from either (a) random errors or (b) differences in factors related to performance or language use, such as memory or attention limitations, access to lexical entries, and other physiological or cognitive constraints in actual production and comprehension of utterances, rather than from differences in grammar (also known as competence). As argued above, the kind of individual differences reported in previous studies is far from being random, and should not be dismissed as simple noise or uninteresting peculiarities of individual speakers. Furthermore, while we should acknowledge that factors in the purview of performance or language use contribute to across-speaker variation, it is worth investigating whether at least some part of individual variation is explained by grammatical components above and beyond factors related to performance. With this background, the present study explores whether we find non-random across-speaker variation in phonology and if so, to what extent performance-related factors can explain the variability. If the observed variation is not due to noise, but shows meaningful and systematic differences between speakers that might reflect differences in grammar, we should be able to capture and formalize it within phonological models.

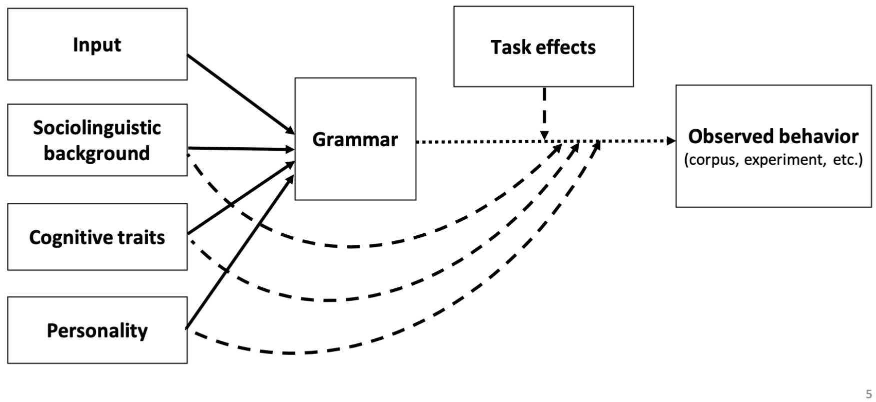

Figure 1.1: Factors influencing individual speakers' phonological behavior

1.2 Sources of individual differences

When individual speakers' speech patterns differ from one another, what are the sources of this individual variability? Previous research has identified various sources of across-speaker variation, including sociolinguistic background, language-related experience, cognitive traits (such as working memory capacity, attention, inhibitory skills and autistic traits) and personality, just to name a few. These factors affect individuals' speech perception and production in various ways, as schematically represented in Fig. 1.1. First, they can affect individuals' learning of the grammar, as represented by solid lines. They also influence how individuals use the grammatical knowledge in the actual tasks of speech comprehension and production (which can be observed in corpus studies, experiments, etc.), as represented by dashed lines. A component of Fig. 1.1 that has not been mentioned so far is task effects; individuals may also differ in how they perform a particular task in experiments, thus depicted as one of the factors influencing speakers' use of grammar in their perception and production. Below I outline previous studies investigating the effect of one or more of these factors on inter-speaker variation in phonetic or phonological behavior.

As mentioned above, one of the most extensively studied variables affecting speaker differences is sociolinguistic background. Numerous studies have reported that individuals differing in sociolinguistic variables such as age, gender, socioeconomic status, social group, and social attitudes exhibit different speech patterns (e.g. Labov, 1994, 2001; Eckert & McConnell-Ginet, 2003; Foulkes & Docherty, 2006; Sumner et al., 2014).

However, even individuals from fairly homogeneous sociolinguistic backgrounds may differ in their comprehension and production patterns. Another source of inter-speaker variability is the language-related experience that one gains over their lifespan, that is, inter-speaker variation may arise because individuals are exposed to different portions of the target language. In the acquisition literature, it has been acknowledged that children's language development can be predicted by characteristics of their caregivers' speech (e.g. Bornstein et al., 1998; Cristia, 2001; Liu et al., 2003; Newman et al., 2016). Further, as argued in Lev-Ari (2016), language input and experience affect one's speech patterns beyond childhood. For instance, speech perception is affected by the number of speakers an individual is exposed to (Lev-Ari, 2018), and speech production is influenced by an individual's dialectal experience (Sanchez et al., 2015; Wade, 2022).

The effect of input differences on inter-speaker variation may be especially critical for a language learning environment in which the data are sparse, and the size of each set of input is small (Niyogi & Robert, 1997; Reilly, 2007). For instance, Reilly (2007) argued that the inter-speaker variation observed in Texistepec Popoluca stems from the sub-optimal learning environment in which the use of language is declining and the dominant language of the community is Spanish.

At first glance, it is not plausible to think that insufficient input is the source of individual variation in a language like Korean, the focus of this dissertation, as there are more than 50 million speakers in South Korea alone. It is not likely to suffer from declining language use, given that Korean is the only official language in the country. However, it is still not impossible that Korean speakers suffer from the problem of an inadequate amount of input for them to converge on the same grammar. For instance, Han et al. (2007, 2016) report that two distinct grammars are found among the Korean native speakers regarding V-movement in the syntactic structures of the language. They argue that the population is divided into two groups, one whose grammar has verb-raising and the other whose grammar does not, and that this is due to insufficient and ambiguous evidence for either grammar. Similarly, noun stems ending in a coronal obstruent are not very frequent in the Korean lexicon, and there is across-speaker variability in pronunciation of these consonants, as will be shown throughout this dissertation. This gives rise to the possibility that different speakers end up with different grammars due to insufficient input.

Recently, there has been growing interest in the relationship between cognition and linguistic behavior. Studies have shown that individual variation in phonetics and phonology is influenced by differences in cognition. Yu & Zellou (2019) reviewed studies investigating the link between individual differences in "cognitive processing styles" and their production or perception patterns. They outline various types of individual traits under the umbrella term cognitive processing styles. The first type is executive functions (a set of cognitive processes that are essential for controlling and managing behavior and thoughts), which in turn encompasses several different traits: attention-related factors (such as the ability to switch attention between different tasks or mental sets, and the ability to perform a task in the presence of a distractor), working memory capacity (concerning updating and monitoring information), and inhibition (the ability to suppress irrelevant and distracting responses). Previous studies have found a link between individuals' attention and cue weighting (Gordon et al., 1993; Kong & Lee, 2018), between working memory capacity and spoken word recognition (Conway et al., 2001; Nitsan et al., 2019) as well as prosodic phrasing (Bishop & Intlekofer, 2020; Bishop, 2021), and between inhibitory control and speech perception (Taler et al., 2010; Lev-Ari & Peperkamp, 2013, 2014) as well as production (Lev-Ari & Peperkamp, 2013).

The second type of cognitive processing styles discussed is autistic traits. Individuals with autism spectrum disorder are recognized for having distinct cognitive processing styles compared to neurotypical individuals. However, it is also acknowledged that individuals who are not clinically diagnosed with autism can exhibit autistic-like traits (Constantino & Todd, 2003; Lundström et al., 2012). Studies have found a relationship between autistic traits of individuals in the non-clinical population and their speech processing, specifically sensitivity to prosodic information (Bishop et al., 2015; Jun & Bishop, 2015), sensitivity to lexical knowledge in speech perception (Stewart & Ota, 2008), perceptual compensation for coarticulation (Yu, 2010), coarticulatory production (Yu et al., 2024), and phonetic imitation (Yu et al., 2013; Snyder et al., 2019).

Other types of cognitive processing styles discussed in Yu & Zellou (2019) include declarative memory (the memory of facts that can be consciously recalled) and procedural memory (the memory of how to perform tasks and skills that are usually unconscious). As noted in Yu & Zellou (2019), these memory types are most often studied in relation to lexical and morphological learning (Ullman, 1999, 2004), although some studies have investigated their effects on (morpho)phonological processing (Ettlinger et al., 2014; Arthur et al., 2021).

Despite these findings, a few studies suggest that little of the across-speaker variance in speech perception and production can be explained by these differences in domain-general traits (Kong & Edwards, 2016; Kapnoula et al., 2017; Hall-Lew et al., 2021; Wade, 2022). For instance, Wade (2022) investigated whether speakers' convergence behaviors are correlated with the Big Five inventory (John et al., 1991), Autism-Spectrum Quotient (Baron-Cohen et al., 2001), and Marlowe-Crowne social desirability scale (Crowne & Marlowe, 1960) and found little evidence that these cognitive and personality traits predict convergence. Hall-Lew et al. (2021) points out that only a subset of individual cognitive measures seem to be relevant to speech or language and it remains unclear whether individual differences are attributable to domain-general cognitive processing styles or more specific aptitudes for speech processing. Therefore, the extent to which individual differences in these cognitive processing styles and personality traits can explain the variance in speech patterns remains unanswered. Moreover, previous research investigating the relationship between cognition and speech primarily focuses on individual differences in phonetics, with relatively few studies examining whether inter-speaker variation at the segmental level can be explained by those cognitive characteristics. The present study aims to address this gap.

1.3 Computational modeling of individual differences

In the present study, I model the way speakers arrive at different probabilistic grammars as they learn a target language. I focus on the realization of stem-final coronal obstruents in Korean (see Chapter 2 for details), which shows a high degree of inter-speaker variation. I fit a Maximum Entropy (MaxEnt) Grammar to the corpus frequencies, which approximates the linguistic input that the speakers receive. Crucially, by modulating the bias terms of certain constraints, and thereby generating different outputs, each of which is expected to match some of the speakers' behaviors observed in the experiment, I model how individuals arrive at different end states of grammar with roughly the same linguistic data.

Many quantitative studies on phonological variation employ computational modeling as a tool to test and compare formal theories of learning (Jarosz, 2019). In studies comparing the pattern of variation observed in the corpus data against the speakers' behaviors in wug tests or judgment rating tasks, the corpus frequencies are fed to a model as learning data, and the model is expected to reproduce the speakers' behaviors in the experiments given an appropriate grammar. In such studies, it is typically the case that the speakers' behaviors are modeled only at the group level by averaging across all participants' responses. The assumption underlying the practice of aggregating data across individuals is that there are no between-speaker differences. Given that this is far from being true in many cases, modeling linguistic behavior must focus more on how people are different from each other, as has been argued in some previous works (Lee & Webb, 2005). Crucially, if individual variation is systematic, predictable by certain factors, and relevant to linguistic theory (as mentioned in Section 1.1), it needs to be modeled.

1.4 Overview of the dissertation

To delve into inter-speaker variation in phonology and the sources of this variability, this dissertation focuses on individual differences in the realization of stem-final coronal obstruents in Korean. I first conducted a corpus analysis to obtain a nuanced understanding of the distribution of the stem-final obstruent variants. Then I carried out a longitudinal study by conducting two experiments to closely examine the inter-speaker variation. The two experiments used the same target words but employed different tasks. I focused on whether the participants are consistent in their responses across different tasks and over time. Such self-consistency would suggest that inter-speaker variation is systematic, with individual differences largely unaffected by the type of task employed. I also investigate the sources of inter-speaker variation through specific tasks included in these experiments. Finally, I model different speakers' production patterns.

The rest of the dissertation is organized as follows. In Chapter 2, I provide background on the realization of stem-final obstruents in Korean. I present results of a corpus analysis on this variable process in Chapter 3, and those of the two experiments in Chapter 4 (Experiment 1) and Chapter 5 (Experiment 2). In Chapter 6, I model the experiment results. Chapter 7 provides a general discussion of the findings of this study.

CHAPTER 2

Stem-final obstruents in Korean

The present study investigates individual differences in the realization of stem-final coronal obstruents in Korean, which exhibit variation before vowel-initial suffixes. In this section, I first provide traditional analyses of how these obstruents alternate between their realization in the final position of an unsuffixed form and before a vowel-initial suffix. Then, I outline alternate explanations proposed in previous studies to account for the observed variation.

2.1 Basic patterns of alternation

Korean obstruents show a three-way laryngeal contrast, as shown in (1). The obstruents in the shaded cells are the focus of this study.

(1) *Obstruent inventory in Korean*

	labial	**coronal**			**velar**
	stop	stop	fricative	affricate	stop
lenis	p	t	s	c	k
aspirated	p^h	t^h		c^h	k^h
fortis	p*	t*	s*	c*	k*

These obstruents are involved in two phonological processes that create alternation within the inflectional paradigm of nouns ending in these obstruents. First, obstruents in coda position are neutralized to their (homorganic) unreleased lenis stop counterpart. For instance, as described in (2), coronal obstruents in /os/ 'clothes' and /pat^h/ 'field' are neutralized to [t] in their unsuffixed forms [ot] and [pat], respectively. Similarly, labial stops neutralize to [p], e.g.

/ip^h/ → [ip] 'leaf', and velar stops to [k], e.g. /puʌk^h/ → [puʌk] 'kitchen', /pak*/ → [pak] 'outside'.

(2) *Neutralization of coronal obstruents in coda*

	Underlying Form	**Unsuffixed**	ACCUSATIVE **/-ɨl/**	**Stem Meaning**
a.	/os/	[o**t**]	[o**s**-ɨl]	'clothes'
b.	/pat^h/	[pa**t**]	[pa**t^h**-ɨl]	'field'
c.	/k*oc^h/	[k*o**t**]	[k*o**c^h**-ɨl]	'flower'
d.	/nac/	[na**t**]	[na**c**-ɨl]	'daytime'

Among the coronal obstruents, /t*/, /s*/ and /c*/ do not appear in the final position of a noun stem. Additionally, there are only a few /t/-final nouns (e.g. /nat/ 'grain', /tikɨt/ 'the name for the Korean alphabet corresponding to *t*'), all of which are infrequent, such that Jun's (2010) corpus study reports zero occurrence of /t/-final stems. (2) therefore presents only /s/-, /t^h/-, /c^h/- and /c/-final stems.

Second, /t/ and /t^h/ are palatalized to [c] and [c^h], respectively, before high front vocoids /i/ or /j/, e.g. /pat^h-i/ → [pac^h-i] 'field-NOMINATIVE', across a morpheme boundary. These phonological rules then generate alternation within the inflectional paradigm, as presented in (3). In the classical analyses of the alternation, the variant that appears before vowel-initial suffixes is considered the underlying form of the obstruent (with an exception of palatalization of /t^h/ before /-i/). The surface forms are then derived through application of coda neutralization (/os/ → [ot]; /pat^h/ → [pat]) and the palatalization rule (/pat^h-i/ → [pac^h-i]).

(3) Basic patterns of alternation of stem-final coronal obstruents

	Unsuffixed	**/-i/ 'NOMINATIVE'**	**/-e/ 'LOCATIVE', /-ɨl/ 'ACCUSATIVE'**
UR	/os/ 'clothes'	/os-i/	/os-e/, /os-ɨl/
SR	[o**t**]	[o**s**-i]	[o**s**-e], [o**s**-ɨl]
UR	/pat^h/ 'field'	/pat^h-i/	/pat^h-e/, /pat^h-ɨl/
SR	[pa**t**]	[pa**c^h**-i]	[pa**t^h**-e], [pa**t^h**-ɨl]

Such analyses, however, fail to describe or explain the regularities that are observed in recent analogical changes in the realization of the obstruents. The following section describes these changes currently underway in Korean and outlines explanations for the emerging patterns based on previous studies.

2.2 Variation

Some obstruents in the final position of a noun stem are realized variably when it is followed by a vowel-initial suffix. When obstruent-final stems are immediately followed by a vowel-initial suffix, the obstruent is expected to be realized as such, as outlined above. However, many non-standard variants are observed. The variable realization of noun stems with final obstruents is illustrated in (4). As for labials, the aspirated stop /p^h/ is often realized as the lenis [p], in addition to its faithful, standard pronunciation [p^h], as illustrated by (a) and (b). There are no stems ending in /p*/. The velar aspirated stop /k^h/ is realized as either [k^h] or [k], as shown in (k) and (l). Among the vanishingly small number of stems ending in /k*/, only /pak*/ 'outside' appears to be familiar to Korean speakers. For this stem, the only acceptable pronunciation of the final /k*/ is [k*] (e.g. [pak*-e], *[pak-e]), according to my informal observation, and previous studies have similarly not reported variable realizations of stem-final /k*/.

The pattern of variation is more complicated for coronal stops. First, stem-final /t^h/ can be realized not only as the canonical [t^h] but also as [s], [c^h] or [t], as shown in (c)-(d). When the suffix is /-i/ 'NOMINATIVE', as in (e), the standard variant is [c^h] due to an independent process of palatalization mentioned above. In addition to the expected [c^h], however, [s] is also frequently chosen as the variant. Similarly, /c^h/ is realized as the canonical [c^h] or the non-canonical [s], [t^h], [c] or [t], as can be seen in (f), (g) and (h). Stem-final /c/, exemplified in (i) and (j), also exhibits variation between the standard form [c] and innovative forms like [s] and [c^h]. It is known that /s/-final stems show a very small degree of variation and is mostly realized with [s], although innovative variants are reported (e.g. /os-ɨl/ → [oc^h-ɨl] 'clothes-ACC'; Jun & Lee, 2007). Korean does not have stems with other coronal obstruents that are productively used.

(4) *Variation in the realization of stem-final obstruents (hankwuk cengsin mwunhwa yenkwuwen [The Academy of Korean Studies], 1995; Choi, 2004; Kang et al., 2004)*

	UR	Canonical	Innovative	Gloss
labial	a. /ip^h-ɨl/	[ip^h-ɨl]	[ip-ɨl]	'leaf-ACC'
	b. /mulɨp^h-i/	[mulɨp^h-i]	[mulɨp-i]	'knee-NOM'
coronal	c. /mit^h-ɨl/	[mit^h-ɨl]	[mis-ɨl], [mic^h-ɨl], [mit-ɨl]	'bottom-ACC'
	d. /k*ɨt^h-ɨl/	[k*ɨt^h-ɨl]	[k*ɨs-ɨl], [k*ɨc^h-ɨl]	'end-ACC'
	e. /pat^h-i/	[pac^h-i]	[pas-i]	'field-NOM'
	f. /pic^h-i/	[pic^h-i]	[pis-i]	'light-NOM'
	g. /k*oc^h-e/	[k*oc^h-e]	[k*os-e], [k*ot^h-e]	'flower-LOC/DAT'
	h. /toc^h-e/	[toc^h-e]	[tos-e], [tot^h-e], [toc-e], [tot-e]	'sail-LOC/DAT'
	i. /pic-ɨn/	[pic-ɨn]	[pis-ɨn], [pic^h-ɨn]	'debt-TOP'
	j. /nac-e/	[nac-e]	[nas-e]	'daytime-LOC/DAT'
velar	k. /puʌk^h-esʌ/	[puʌk^h-esʌ]	[puʌk-esʌ]	'kitchen-LOC'
	l. /sepjʌknjʌk^h-e/	[sepjʌknjʌk^h-e]	[sepjʌknjʌk-e]	'dawn-LOC/DAT'

Abbreviations: ACC = accusative, NOM = nominative, LOC = locative, DAT = dative, TOP = topic

The present study focuses on coronal obstruents, which exhibit a wider range of innovative variants than the labial or the velar stop. It is known that [s], [t^h], [c^h], [c], [t] are attested as variants of stem-final /t^h/, /c^h/ and /c/ (Choi, 2004; Kang et al., 2004; Jun, 2010). Among the non-standard variants, [s] has been reported to be the most frequent pronunciation, [t^h] and [c^h] are moderately attested, while [c] and [t] are infrequent. It has also been found that the relative frequency between [t^h] and [c^h], the two variants with an intermediate frequency, is conditioned by the type of suffix that follows. Specifically, [t^h] is more frequent before [e]-initial suffixes than before [ɨ]-initial ones, and [c^h] is more frequent before [ɨ]-initial suffixes than before [e]-initial ones.

While there are many explanations in the literature for the predominance of [s] (Ko, 1989; Kim, 2001), I adopt the explanation that locate the source of the attested pattern in the lexical distribution of final obstruents (Hayes, 1998; Kang, 2003, 2005; Albright, 2008; Jun, 2010), as will be summarized below. For other accounts that explain the variation, I refer readers to Jun (2010).

In order to understand why [s] is the most frequent variant, we need to consider a number of facts about Korean. First, coronal obstruents are neutralized to [t] in coda position, as mentioned above, repeated as part of (5). The neutralization creates various types of alternation, e.g. [t]~[s] alternation in (5a), [t]~[t^h] in (5b), etc. Therefore, the prevalence of [s] as a variant of non-/s/ obstruents, e.g. [pas-ɨl] for /pat^h-ɨl/, [k*os-ɨl] for /k^*oc^h-ɨl/ and [nas-ɨl] for /nac-ɨl/, can be seen as an extension of [t]~[s] alternation to other environments (b-d). Then the question is, why is it [s] that is extended to non-/s/ stems, rather than some other way around? For instance, why is the logically possible pattern in which [t^h] is extended to others (e.g. *[ot^h-ɨl] for /os-ɨl/) rarely attested? This is because /s/-final stems are the most frequent among the stems that end in coronal obstruents (approximately 50%, with some variability depending on which corpus or dictionary is consulted, e.g. Kang & Kim, 2004; Albright, 2008). Whenever one sees the [t] in the unsuffixed form, chances are high that it will map to [s] in the suffixed form (when the suffix is vowel-initial). In other words, the fact that Korean is rich in /s/-final stems makes [s] a safe "guess" as a variant to be used in suffixed forms.

(5) *Alternation caused by coda neutralization*

	Underlying Form	**Unsuffixed**	ACCUSATIVE **/-ɨl/**	**Alternation**	**Stem Meaning**
a.	/os/	[o**t**]	[o**s**-ɨl]	[t]~[s]	'clothes'
b.	/pat^h/	[pa**t**]	[pa**t^h**-ɨl]	[t]~[t^h]	'field'
c.	/k^*oc^h/	[k*o**t**]	[k*o**c^h**-ɨl]	[t]~[c^h]	'flower'
d.	/nac/	[na**t**]	[na**c**-ɨl]	[t]~[c]	'daytime'

This explanation is convincing, given that some suffixes (especially the nominative and the accusative suffix) are often omitted in Korean, which leads the native speakers to be exposed to unsuffixed forms frequently. For instance, two utterances [o**s-ɨl** ip-ʌ] and [o**t** ip-ʌ] 'clothes(-ACC) wear-DECLARATIVE' can be used interchangeably to mean 'wear clothes'. As such, the native speakers have only a small amount of exposure to the suffixed forms, especially for infrequent stems whose final obstruent is not /s/. As a result, alternations involving these infrequent stems are replaced by the dominant [t]~[s] alternation.

In fact, the observed relative preference among variants is matched by the distribution of lexical final obstruents in noun stems. As mentioned earlier, [s] is the most preferred, followed by [t^h] and [c^h], while [c] and [t] are least preferred. The distribution of stem-final coronal obstruents in the lexicon is such that /s/-final nouns are the most common, while /t^h/-final and /c^h/-final stems are moderately frequent, and /c/-final and /t/-final stems are least common. (/t/-final stems are close to non-existent, as noted earlier.)

Similarly, the suffix type effect for [t^h] and [c^h] can be understood with some generalizations about the lexical distribution of stem-final obstruents and suffix-initial vowels. To recap how the relative preference is conditioned by the suffix, [t^h] is relatively favored before [e]-initial suffixes and [c^h] is preferred before [ɨ]-initial suffixes. This pattern is most noticeable for /t^h/-final nouns. Jun (2010) shows that this reflects the fact that [e]-initial suffixes are more likely to combine with /t^h/-final stems than /c^h/-final stems in the Korean lexicon, while the opposite is true for [ɨ]-final stems. It seems that many of the /t^h/-final stems happen to denote location, e.g. /mit^h/ 'bottom, under', /pak*at^h/ 'outdoor', /kʌt^h/ 'outside', thereby frequently combining with suffixes that denote location, e.g. /-e/ 'LOC/DAT' and /-esʌ/ 'LOC', which happen to be [e]-initial. Given that the aforementioned stems are high in frequency, and that these locative suffixes cannot be omitted, we can imagine that speakers will hear [mit^h-e], [kʌt^h-e], etc. very frequently.[1]

Although the variable realization of stem-final obstruents has been investigated in previous studies (e.g. Choi, 2004; Albright, 2008; Jun, 2010), the variation has only been reported at the group level by aggregating all the speakers' responses. Some of these studies do mention, however, that speaker-by-speaker variation might exist. The present study examines between-speaker variation in detail to see how much speakers differ from each other and in which environments, focusing on coronal obstruents.

[1]One exception to this pattern is /-ɨlo/ 'DIRECTIVE (DIR)', which is [ɨ]-initial but frequently combines with location-denoting stems. It will be shown in Chapter 3 that /t^h/ is mostly produced as [t^h] before /-ɨlo/.

CHAPTER 3

A corpus study

3.1 Introduction

There are few corpus studies that investigated the realization of noun-final obstruents in Korean. This is presumably because there are only a small number of obstruent-final nouns in the language, which makes it hard to find their occurrences even in a fairly large corpus. For instance, according to Albright (2008), only 18% of Korean nouns are obstruent-final. Using a large-sized corpus was especially important for the present work, because a large number of tokens must be provided by each speaker in order to make generalizations at the individual level.

3.2 Methods

3.2.1 Data

I used a speech corpus consisting of recordings of conversations between pairs of speakers discussing everyday topics (National Institute of Korean Language, 2022). The total duration of recordings reaches approximately 500 hours, consisting of around 900,000 intonational phrases. Each phrase was tagged with speaker ID, so it was in principle possible to investigate inter-speaker variation.

3.2.2 Target words

I created a list of target words based on Choi's (2004) survey items and the frequency list of Korean (Kang & Kim, 2004). The target words consisted of noun stems ending in a coronal obstruent followed by a V-initial suffix.

As for the target stems, I used Choi's /t^h/-final and /c^h/-final nouns except for /sat^h/ 'crotch', /nat^h/ 'piece' and /oc^h/ 'a tree species', because I judged them to be infrequent or unfamiliar to Korean speakers. Choi's survey did not include /s/-final and /c/-final nouns, presumably because they show less variation. Target stems with final /s/ or /c/ consisted of those listed in Kang & Kim (2004) that were judged to be fairly familiar to native Korean speakers and were either native-Korean or Sino-Korean (i.e. loanwords were excluded.) The target stems included both monomorphemic and polymorphemic ones. For instance, not only did I include /pic^h-i/ 'light-NOM', but /$hespic^h$-i/ 'sunlight-NOM' was also extracted. Although it is possible that morphologically simplex and complex words behave differently in terms of how the stem-final obstruent is realized, a visual inspection of the data suggested no crucial difference. I aggregated the two groups when reporting the results. The list of target stems is provided in (6).

(6) Target stems in the corpus analysis

Stem	Gloss
kʌs	'thing'
tɨs	'as if, like'
kos	'place'
muʌs	'what'
t*ɨs	'meaning'
os	'clothes'
mas	'taste'
t^has	'blame'
ius	'neighbor'
cis	'act'
calmos	'fault'
nolɨs	'circumstance'
pʌlɨs	'habit'
kʌcis	'lie'
mʌs	'stylishness, smartness'
amk^{h}ʌs	'female (animal)'
mos	'nail' / 'pond'
suk^{h}ʌs	'male (animal)'
s*ias	'seed'
jʌlʌs	'several'
kɨlɨs	'plate'
pus	'brush'
jʌs	'taffy'
pʌsʌs	'mushroom'
kis	'feather' / 'collar'
cas	'pine nut'
pis	'brush (for hair)'
iptʌs	'morning sickness'

Stem	Gloss
kʌt^h	'outside'
kjʌt^h	'side'
k*ɨt^h	'end'
mʌlimat^h	'bedside'
mit^h	'bottom'
pak*at^h	'outside'
pat^h	'field, farm'
pjʌt^h	'sunlight'
sot^h	'cauldron'
sut^h	'thickness (of hair)'
p^hat^h	'red bean'
k*oc^h	'flower'
nac^h	'face'
mjʌc^h	'some, how many'
pic^h	'light'
salkac^h	'skin'
suc^h	'charcoal'
nac	'day'
cʌc	'milk, breast'
pic	'debt'

I then extracted all the words that consist of one of the target stems and an [i]-, [e]- or [ɨ]-initial suffix (N=9444).[1] A total of 16 suffixes were initially found. I only included six suffixes that

[1]Among the seven vowels in Korean, [i], [e] and [ɨ] are the ones most frequently used in the initial position of a

are fairly frequent and have been often discussed in previous studies, presented in (7), because the other suffixes were not frequent enough to form a generalization with regard to the effect of suffix type. These six suffixes made up 91.5% of the total data (N=8638). The cases in which suffixes were stacked, e.g. /mit^{h}-e-man/ 'bottom-LOC-only', were also included in the analyses and treated as the same category as words with only the first suffix, e.g. /mit^{h}-e/ 'bottom-LOC'.

(7) Suffixes included in the corpus analysis

Suffix	Gloss
-e	LOCATIVE/DATIVE
-esʌ	LOCATIVE
-i	NOMINATIVE
-ɨl	ACCUSATIVE
-ɨlo	DIRECTIVE
-ɨn	TOPIC

3.2.3 Analysis

The target words were coded for their surface form of the final obstruent by perceptual judgment, and based on acoustic cues when necessary. In the analysis, I only included those whose final obstruent was realized as one of the five variants that are often considered in the literature, [s], [t^{h}], [c^{h}], [c] and [t].[2] This excluded 0.7% of the total data. The final data set for analysis included 8579 tokens (7049 /s/-final, 870 /t^{h}/-final, 512 /c^{h}/-final and 148 /c/-final), with 48 stem types and 167 word types.

suffix. There are no suffixes that begin with an [o], [u] or [ʌ]. There is only one suffix that begins with [a], i.e. the VOCATIVE suffix /-a/. This suffix has limited usage, as it typically combines with a noun that indicates a person's name or an animate noun, and is exclusively used when addressing the noun. Choi (2004) included this suffix in her survey, but many of the test items employing it sound unnatural to me, as most of the target stems are inanimate. Therefore, it is not included as one of the target suffixes in this study.

[2]The excluded ones are: (i) cases where the consonant was deleted due to casual and fast speech, (ii) cases in which the recording quality was not good enough and (iii) a handful of cases in which it was realized as a consonant other than the five considered above.

3.3 Results

The frequency of each variant for the target words are calculated based on token-weighted type frequency: each word contributes 1 unit, divided into fractions based on the number of tokens in which individual variants occur for that word. For example, the word /pat^{h}-e/ 'field-LOC' occurred 19 times in total (token frequency), among which 13 were [pat^{h}-e], 4 were [pas-e], and 2 were [pac^{h}-e]. Each of these thus contributed 0.68 (13/19), 0.21 (4/19), and 0.11 (2/19).

Fig. 3.1 shows how stem-final /s/, /t^{h}/, /c^{h}/, and /c/ are realized in the corpus, which largely replicates findings of the previous studies. As mentioned earlier, because each word contributes 1 unit, the bar height represents the number of words in each condition (final consonant + suffix). For instance, there were a total of 10 words whose stem-final consonant was /s/ and the suffix was /e/, all of which were realized as [s].

As we can see in the leftmost panel, stem-final /s/ is virtually always realized as [s] regardless of the suffix type (99.95%), as expected. A few cases did appear in which /s/ was realized as something else ([t] or [c]), but this constituted only 0.05% of the total number of /s/ tokens.

The realization of /t^{h}/ is conditioned by the suffix type. When followed by /-e/ or /-esʌ/, the majority of /t^{h}/ tokens are realized as [t^{h}] (80.2% before /-e/, 88.6% before /-esʌ/). In contrast, when followed by /-ɨl/, the proportion of the canonical pronunciation [t^{h}] (35.5%) is comparable to that of [s] (36.3%) and only slightly higher than that of [c^{h}] (28.4%). Similarly, the proportion of [t^{h}] before /-ɨlo/ (69.0%) and that before /-ɨn/ (33.3%) was lower than that before [e]-initial suffixes. This confirms findings of previous studies that [t^{h}] is preferred before [e]-initial suffixes more than before [ɨ]-initial ones. In the case of /-i/, due to the palatalization rule mentioned earlier, /t^{h}/ is expected to be realized as [c^{h}]. While the most frequent variant for /t^{h}/ before /-i/ was indeed [c^{h}] (49.2%), the proportion of [s] was similarly high (44.5%). Surprisingly, an illegal form [t^{h}] was found in this context.[3]

[3]There was only one token of /t^{h}-i/ that was realized as [t^{h}-i], namely /mit^{h}-i/ → [mit^{h}-i]. Since /mit^{h}-i/ had two tokens and one of them was produced as [mit^{h}-i], this contributed 0.5 (1/2) out of 9 units (as represented by the bar height in Fig. 3.1) to the /t^{h}-i/ category. The suffix was phonetically realized as [i] (hence the transcription

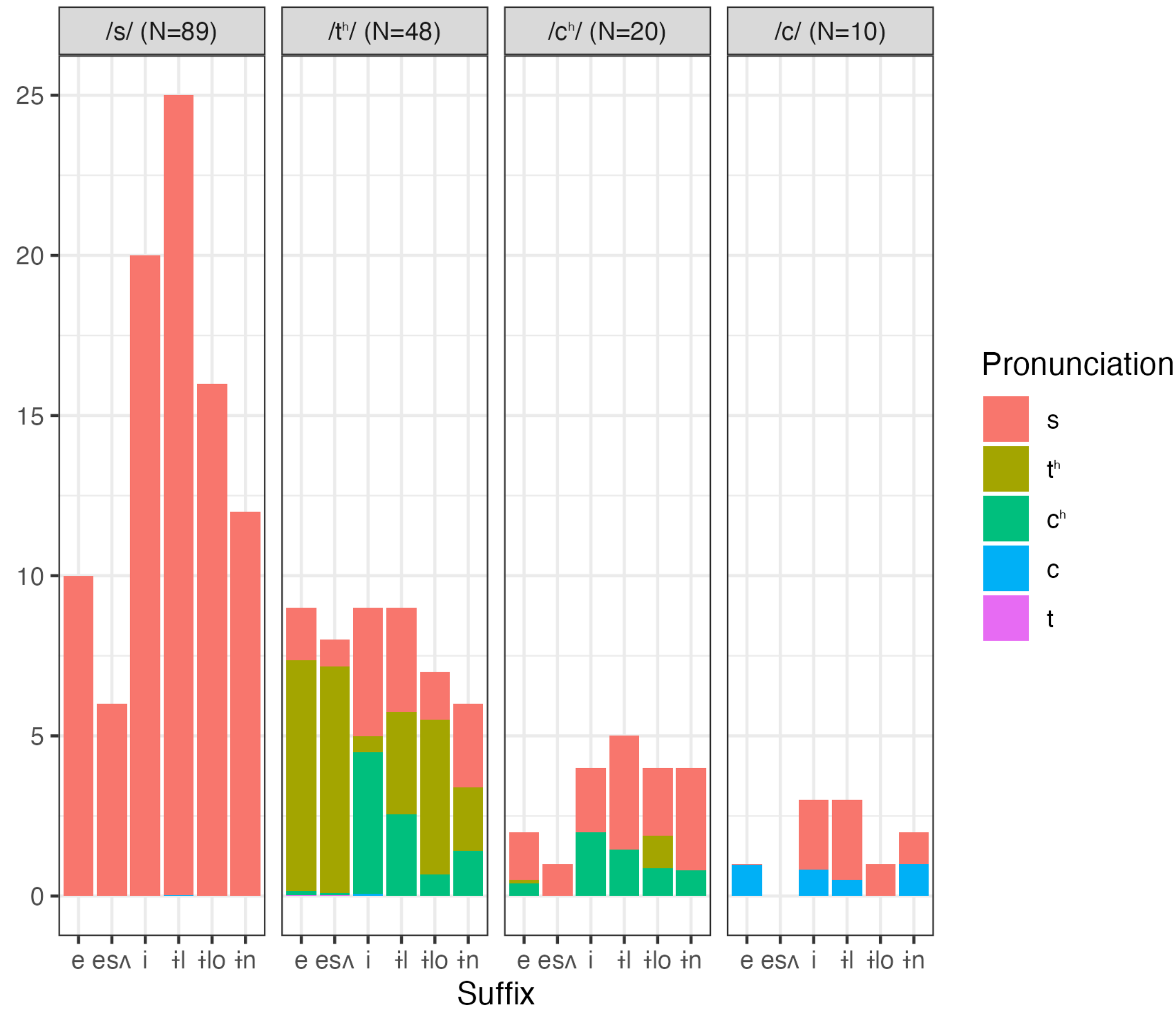

Figure 3.1: Aggregated responses of all speakers in the corpus, calculated in token-weighted type frequency. The numbers at the top of each panel indicates the number of word types for each of the final obstruents.

For /cʰ/-final stems, the competition arises mainly between [s] and [cʰ]. It is noticeable that the proportion of [s] was 50% or higher in all suffixes, outnumbering the canonical variant [cʰ]. It is also notable that there are two suffixes for which [tʰ] is observed (namely /-e/ and /-ɨlo/) and these are precisely the ones which exhibited relatively high proportions of the canonical [tʰ]

by the coder), but semantically, /-e/ would have been more appropriate in the utterance. I thus suspect that the speaker intended to produce [e] but ended up producing [i], which can be considered a speech error.

responses in /tʰ/. (Another suffix which also had a high proportion of [tʰ] in /tʰ/ is /-esʌ/, but this suffix combined with /cʰ/-final stems infrequently.) Therefore, we can say that the frequencies of variants for /cʰ/ reflect the lexical distribution that /tʰ/-final words frequently combine with /-e/ and /-ɨlo/.

Finally, the probability of /c/ being realized as the standard [c] is also conditioned by the suffix type: it is highest before /-e/ (98.2%), intermediate before /-i/ (27.8%), before /-ɨl/ (16.7%) and before /-ɨn/ (50%), and 0% before /-ɨlo/. The high proportion of [c] before /-e/ can be explained with a similar account to that of /tʰ/-final stems. The most frequent one among a small number of /c/-final stems is /nac/ 'daytime', which is frequently suffixed with /-e/. Note that [c] is chosen as a variant only for /c/; it is not extended to other consonant contexts. This is presumably because there are only a few /c/-final noun stems in the lexicon.

3.4 Discussion

The present corpus study using a large-sized speech data sample provided a comprehensive picture of how coronal obstruents in stem-final position are realized, replicating major findings of previous studies such as the preference for [s] and the effect of suffix type. Assuming that the corpus data approximate the language input that Korean native speakers receive while learning the grammar, the corpus frequencies obtained will serve as the learning data in the future modeling work (Chapter 6) that formalizes the speakers' grammars.

However, there are a few limitations of the corpus study. First, although the speakers were asked to have a "free conversation", the recording took place in the lab rather than in a naturalistic setting. Moreover, in many cases, the two interlocutors did not know each other before coming to the lab. Therefore, the speech analyzed here might not accurately reflect the native Korean speakers' production patterns of spontaneous speech. Second, the data includes not only the Seoul Korean (which is often assumed to be the "standard" dialect) but also other dialects, most notably Gyeongsang dialect. Thus, the results might not represent a single speech community's production patterns.

Importantly, it was not possible to analyze the pronunciation at the individual level. Although the size of the corpus is quite large (see Section 3.2.1), each pair of speakers contributed only about 15 minutes of recording, and only an average of 3.6 target tokens were found per speaker (recall that nouns with a final coronal obstruent are not very frequent in the Korean lexicon.) Given that such a small number of tokens were obtained from each speaker from 15 minutes of their speech, it might not be feasible to collect a large number of target tokens in any kind of corpus that would be sufficient to investigate individual differences. In the next section, I present a production study conducted to explore the speaker variation in the realization of stem-final coronal obstruents by eliciting relatively large number of tokens from each speaker.

CHAPTER 4

Experiment 1

4.1 Introduction

I conducted an online production experiment to investigate inter-speaker variation in their pronunciation of stem-final coronal obstruents and the factors contributing to this variation. To minimize potential influence of individual differences in their execution of the experimental task, such as their understanding of the task and familiarity with the experimental setup, I opted for a straightforward task. Participants were given the flexibility to complete it at their convenience and in a location of their choice.

4.2 Methods

4.2.1 Participants

35 native speakers of Korean participated in Experiment 1 (15 males and 20 females, age mean: 30.1, range: 21-48). Participants were recruited online (Prolific) and by word-of-mouth. They were paid for their participation.

4.2.2 Stimuli

A total of 30 stems with a final coronal obstruent were used (four /s/-final, thirteen /t^h/-final, ten /c^h/-final, and three /c/-final stems). I followed the same procedures for choosing the target words as I did for the corpus study (see Section 3.2.2), except that the number of /s/-final target stems in the experiment was much smaller. It was expected that /s/-final stems would exhibit only a small degree of variation, as outlined in Chapter 3, so only a small number of /s/-final stems were included to shorten the duration of the experiment session. Thus, the distribution of stem-final obstruents among the experimental stimuli was not the same as that in the corpus, as /s/-final stems were underrepresented in the experiment. The relative number of /t^h/-final, /c^h/-final, and /c/-final stems remains consistent with the corpus data, where the order is /t^h/-final stems > /c^h/-final stems > /c/-final stems. I also included 12 filler stems, whose final segment was one of /m, n, ŋ, l/ or a vowel. The full list of the target stems is provided in (8).

(8) Target stems used in the experiments

Stem	Gloss
os	'clothes'
calmos	'fault'
kɨlɨs	'plate'
pus	'brush'
kʌt^h	'outside'
kjʌt^h	'side'
k*ɨt^h	'end'
twik*jʌt^h	'backyard'
mʌlimat^h	'bedside'
mit^h	'bottom'
pak*at^h	'outside'
pat^h	'field, farm'
pjʌt^h	'sunlight'
sot^h	'cauldron'
sut^h	'thickness (of hair)'
mut^h	'land'

p^hat^h	'red bean'
k^*oc^h	'flower'
nac^h	'face'
tac^h	'anchor'
toc^h	'sail'
$mj\Lambda c^h$	'some, how many'
$t\Lambda c^h$	'trap'
pic^h	'light'
$salkac^h$	'skin'
suc^h	'charcoal'
juc^h	'a type of traditional game'
nac	'day'
cʌc	'milk, breast'
pic	'debt'

A pilot study that included all six suffixes from the corpus analysis extended the duration of the experiment session, which was likely to cause participant fatigue. Therefore, only the three most common suffixes, namely /-e/ 'LOC/DAT', /-i/ 'NOM', and /-ɨl/ 'ACC', were included as the target suffixes in the experiment. (9) provides the target suffixes.

(9) Target suffixes used in the experiments

Suffix	Gloss
-e	LOC/DAT
-i	NOM
-ɨl	ACC

Not every possible combination of the target stems and target suffixes was included as a test word. This is because some combinations resulted in a word that is unlikely to be used by the native speakers due to their semantics. The final set of test items consisted of 104 words, each of which was produced twice by the same speaker (block 1 and block 2).

For each target word, a unique sentence was presented that provided a context in which the word can be used. These sentences indicated which suffix should be attached to the stem. For instance, for /kʌt^h-e/ 'outside-LOC', the associated sentence was "poŋt^{h}u (kʌt^h-e) cuso-lɨl s*ʌ cu-se-jo" 'Please write the address on the outside of the envelope', where (kʌt^h-e) was replaced with a blank (see below for the procedure of the experiment). I intentionally rendered the sentences in a colloquial style to encourage more natural speech.

4.2.3 Design

When a trial began, a sentence with a blank was presented that the participants were expected to complete with the target word. The stem was played through the audio (in the unsuffixed form), and the participants were required to come up with a suffix that would best align the word's meaning with the provided sentence. In the aforementioned example, the participants saw the sentence "poŋt^{h}u _________ cuso-lɨl s*ʌ cu-se-jo" 'Please write the address _________ of the envelope', heard the stem [kʌt] (from /kʌt^h/), and were expected to read the sentence aloud by filling in the blank with a variant of /kʌt^h-e/ (e.g. [kʌt^h-e], [kʌs-e], etc.) They pressed the *Start recording* button, read the entire sentence aloud, and then hit the *Stop recording* button to end the trial. They were asked to read the whole, completed sentence, rather than just say the target word, to make them less conscious about their pronunciation of the target word and to elicit more natural speech. The speakers were allowed to listen to the stem as many times as they wanted, but were given only one chance to record.

Because noun suffixes in Korean can be omitted in many cases, participants were given an explicit instruction that they must attach a suffix to the stem when filling in the blank. They were told that one of the purposes of the experiment was to see whether native speakers can use suffixes correctly, in order to make sure that they always attach a suffix and to hide the real purpose of the experiment.

4.2.4 Procedure

The entire experimental session was held online. Once participants accessed the link, they were presented with a consent form. After a microphone and speaker test, participants read the instructions and went through three practice trials with stems ending in a sonorant (e.g. /tut^{h}oŋ/ 'headache'). Each of the three suffixes was the correct answer in each of the three trials. Participants were provided with the correct inflected form after they finished recording (e.g. /tut^{h}oŋ-i/). The main session consisted of two blocks, and they were told to take a break as long as they liked between the blocks.

After the main session, participants filled out the free-form question "What do you think this experiment is investigating?". The participants' answers were coded as one of "correct guess" "partially correct guess" and "incorrect guess". If their answer indicated that they knew that the pronunciation of the stem-final obstruents was the focus of the study, it was coded as "correct". If the answer alluded to "pronunciation" but did not point out that it was about the obstruents, it was coded as "partially correct". Answers that did not mention pronunciation and instead suggested other purposes, such as testing knowledge about the suffixes, were coded as "incorrect". As will be discussed in detail in Section 5.3.5, 21 participants (60%) correctly guessed the purpose, 8 participants (22.9%) were partially correct, and 6 participants (17.1%) made an incorrect guess.

The next set of questions tested the participants' knowledge of spelling. They heard each target stem played through the audio in their unsuffixed form (with the neutralized form [t] in the final position), and typed the spelling in Hangul (the Korean alphabet system). For stem-final obstruents, the orthography indicates the lexical form of these consonants. Thus, these questions were included to tap into participants' knowledge about the underlying forms of the stems as represented in their mental lexicon. It turns out that the participants provided the correct spelling in most (97.9%), if not all, responses, and the results will not be discussed further in Section 4.3. There was minimal individual variability in spelling knowledge, as most participants performed exceptionally well on the spelling test (see Section 5.7 for details).

The participants also answered the question "How important do you think it is to speak with a correct pronunciation?" on a Likert scale of 1 (Not important at all) to 7 (Very important). They were also asked to rate each stem for familiarity, i.e. how often they hear and use each stem, on a Likert scale of 1 (rarely) to 7 (very often). A demographic survey was provided at the end, which consisted of questions about age, gender and language background.

4.2.5 Analysis

I listened to each token and judged how the stem-final obstruents were realized. Only the five variants [s], [t^h], [c^h], [c] and [t] were included in the analysis. Responses were excluded when the speaker gave two different pronunciations within a single trial. In some of these cases, it sounded like the speaker had tried to correct their pronunciation in the second attempt[1] (although it was not always the case that the second attempt was the canonical pronunciation). Occasionally, the speakers produced a suffix that was not the intended suffix.[2] If the alternative suffix was consonant-initial, it must be excluded, simply not being the target of this study[3]; if it was vowel-initial, it could in principle be included in the analysis, but was excluded for ease

[1]This suggests that they were conscious about their pronunciation of stem-final obstruents and might have realized the purpose of the study. This issue is discussed in detail in Section 4.5.

[2]The sentences were still grammatical with such alternative suffixes. For instance, the nominative /-i/ and the topicalizing /-ɨn/ can be interchangeably used in many sentences in Korean with minimal change in meaning. Although I tried to come up with test sentences that sound much more natural with /-i/ than with /-ɨn/, on some occasions participants chose /-ɨn/.

[3]Responses with consonant-initial suffixes may be interpreted as the participants' strategy to avoid guessing the alternant of the obstruent before vowel-initial suffixes. A similar pattern was observed in Do (2013), where children used consonant-initial suffixes to avoid phonological alternation. Only 0.2% of the total responses employed consonant-initial suffixes in the current study.

of analysis. Responses were also deleted when the speaker failed to record, when it was hard to determine how the consonant was realized due to poor recording quality, or when the consonant was not among the aforementioned five variants. Approximately 4% of the total number of tokens were excluded, leaving 1382 tokens for the final analysis.

4.3 Results

I first present the results for all the speakers aggregated. Fig. 4.1 shows the distribution of [s], [t^h], [c^h], [c] and [t], grouped by suffix and underlying consonant. Unlike Fig. 3.1, in which token-weighted type frequencies were used, Fig. 4.1 presents results based on token frequency. This choice is due to all target words being presented an equal number of times (N=2; once in each block), so we do not need to worry about a few frequent stems dominating the pattern when using token frequency. Additionally, the number of /s/-final stems was intentionally reduced in the experiment to decrease the duration of the session, so using the raw number of stems would not be particularly informative.

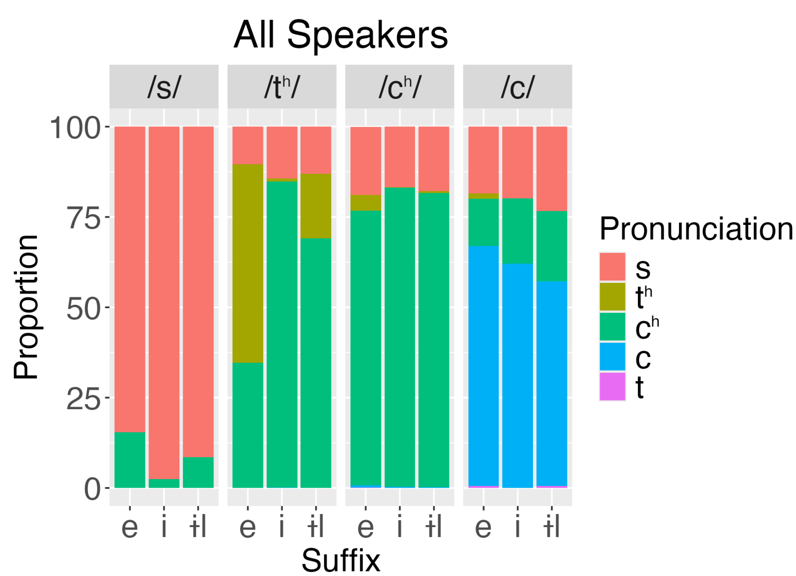

Figure 4.1: Aggregated responses of all speakers in Experiment 1

The majority of /s/-final tokens were produced as [s], as expected, but the non-canonical [c^{h}] was also observed in all suffixes, as high as 15.4% before /-e/. This is in contrast to the corpus data, in which [c^{h}] was absent for stem-final /s/. Among the four /s/-final stems included in the experiment, it was the responses for the least frequent stem /pus/ 'brush' that accounted for most of these non-canonical tokens; the other three stems were realized as [s] more than 99% of the time.

The realization of stem-final /t^{h}/ was influenced by the suffix, replicating findings of previous studies (Jun, 2010). The proportion of [t^{h}] was higher before /-e/ (54.9%) than before /-ɨl/ (17.9%), as was the case in the corpus data (80.2% before /-e/, 35.5% before /-ɨl/). For these two suffixes, [s] and [c^{h}] were observed as innovative forms. In the palatalization context (before /-i/), 84.7% of the responses were the canonical [c^{h}], and 14.3% were the innovative [s].

The proportion of the canonical [c^{h}] for /c^{h}/-final stems was overall higher than that in the corpus study (27.6%), ranging from 76.0% before /-e/ to 82.8% before /-i/. It is notable that there were some [t^{h}] responses for /-e/ (4.3%), which was rarely observed before /-i/ (0.1%) or before /-ɨl/ (0.6%). This again suggests that [t^{h}] is preferred before /-e/ than before /-ɨl/, replicating findings of previous studies (Jun, 2010).

As for /c/, the proportion of the canonical variant [c] did not greatly differ across the suffix types, ranging from 56.7% before /-ɨl/ and 66.5% before /-e/. This is in contrast with the results of the corpus study, in which the rate of [c] responses was substantially higher before /-e/ (98.2%) than before the other two suffixes (27.8% before /-i/, 16.7% before /-ɨl/). Another difference is that while [c^{h}] tokens were not observed for /c/-stems in the corpus, [c^{h}] responses were present in /c/-final stems in the experiment – in fact, they were found as frequently as [s] responses. Similar to the case in /c^{h}/-final stems, there was a small number of [t^{h}] responses for /-e/ (1.5%), but not for other suffixes. Finally, [t] was occasionally found.

The occurrence of innovative [c^{h}] has been reported in previous research (Jun, 2010; Albright, 2005, 2008), which was attributed to the high frequency of /c^{h}/-final stems in the lexicon (although they are still less frequent than /s/-final stems). Notably, the proportion of [c^{h}] is higher than that of [s] in the current experiment, whereas previous studies reported that [s] is the most frequent innovative variant (Jun, 2010; Albright, 2005, 2008; Kang, 2003; Hayes, 1998; Ko, 1989). I speculate that the higher frequency of [c^h] is an experimental artifact: there were only four /s/-final stems but ten /c^h/-final stems in the stimuli, so encountering /c^h/-final stems more frequently than [s]-final stems during the experimental session has overridden the relative frequency difference between [s] and [c^h] in the lexicon (namely [s] being more frequent than [c^h]).

If there was no variation across speakers, all speakers would have the distribution of the variants as shown in Fig. 4.1. It was found, however, that speakers differ substantially from each other in several aspects, as will be outlined in the following sections.

4.3.1 Proportion of canonical variants

I first examined the proportion of canonical variants in individual speakers' speech (Fig. 4.2). A moderate degree of inter-speaker variation was observed, with the proportion ranging from 25.7% to 84.7%. Every participant produced at least some non-canonical responses.

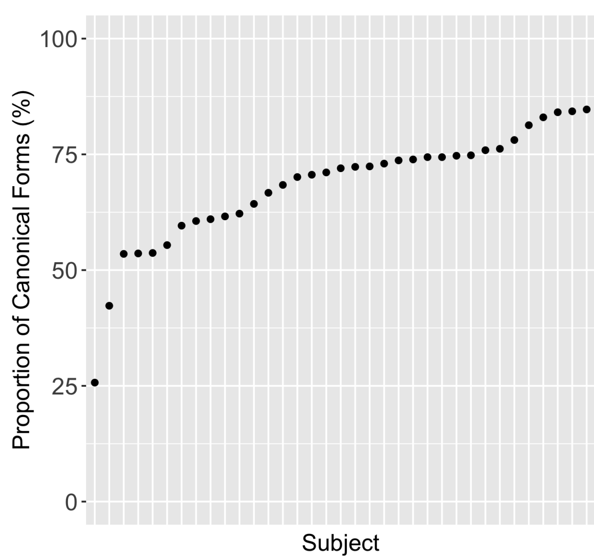

Figure 4.2: Individual speakers' proportion of canonical forms in Experiment 1

I then investigated individual differences in more specific phonological environments that might have led to differences in the frequency of canonical forms. I first examined whether speakers differed from each other concerning some of the trends discussed in the literature. One of them was occurrence of [s] as an innovative variation; do individuals differ in their propensity to produce non-canonical [s] responses? Another observation is the suffix effect in which [t^h] tends to be preferred before /-e/ but [c^h] is more frequent before /-ɨl/. Since this effect of suffix was most prominent in /t^h/ (see Fig. 4.1), I investigated differences in the proportions of [t^h] responses between /t^h-e/ and /t^h-ɨl/ as an index of the suffix effect. To be consistent with other variables, which focus on innovative responses, I examined differences in the proportions of *innovative* responses between /t^h-e/ and /t^h-ɨl/; this is essentially a mirror image of the differences in the proportions of [t^h] responses. Next, I examined individual variability in the rather unexpected tendency documented in previous studies (Jun, 2010; Jun & Lee, 2007), in which /s/-final words are produced with a non-[s], specifically [c^h].

I focused on two additional phonological environments that have not been extensively discussed in previous studies. The first of these two variables is the proportion of (non-)canonical responses for /c/. As demonstrated below, individuals exhibited a substantial degree of variation in this regard. The second is the occurrences of the innovative [c^h], a trend that is more frequently observed in the experiment compared to the corpus data, especially in /s/ and /c/, as noted above.

In sum, the following subsections outline five variables for which the speakers substantially differ from one another: (1) innovative [s] responses (the proportion of [s] responses observed for non-/s/ words) (2) the difference in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/ (3) innovative responses for /c/ (4) innovative responses for /s/, and (5) innovative [c^h] responses (the proportion of [c^h] responses in non-/c^h/ words).

4.3.2 Speaker variation in more specific environments

For each environment, I first present the proportions of all participants' production of the variant of interest (e.g. innovative [s] responses). Additionally, I compare the response patterns of speakers who substantially differ in their use of that variant, with four speakers chosen for illustration in Fig. 4.3. The full set of participants' responses across all words is presented in Section 4.3.3.

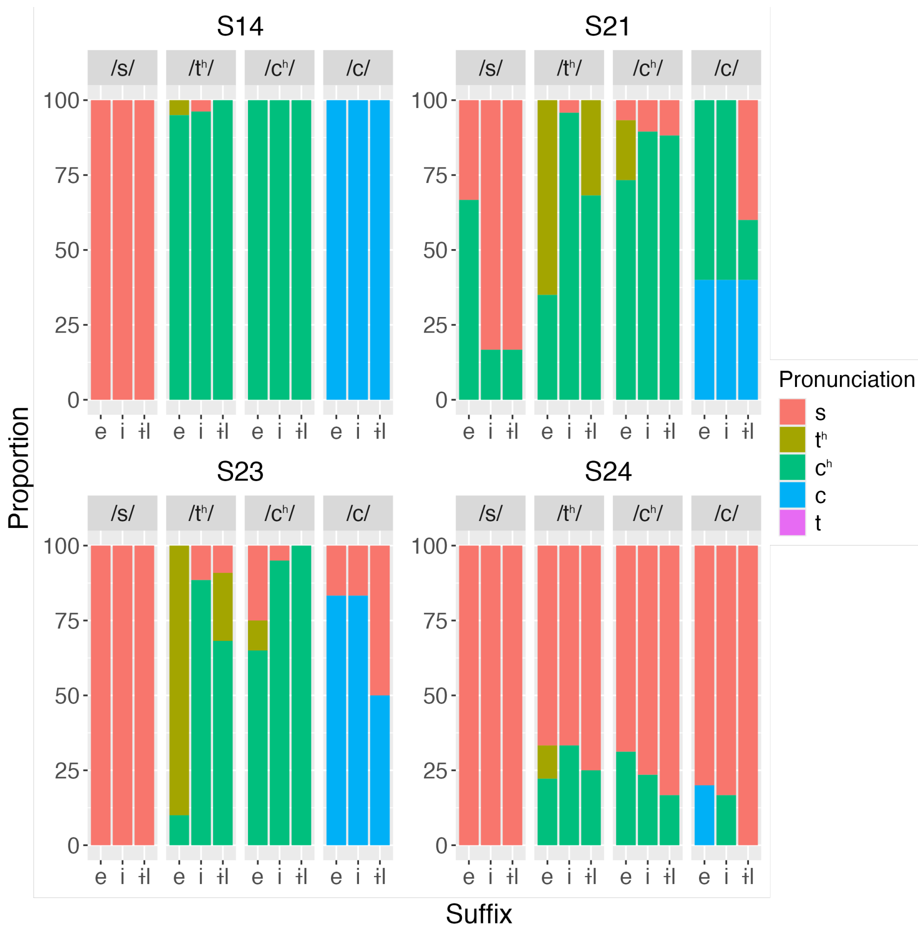

Figure 4.3: Four participants' responses in Experiment 1

Proportion of innovative [s] responses

One phonological environment in which individual variation was observed is the degree to which [s] is extended to /t^h/, /c^h/ and /c/. In Fig. 4.4, we observe that the proportion of [s] responses for non-/s/ stems ranges from 0% to to 77.8%. In Fig. 4.3, we see a stark contrast between S24, who frequently used [s] for non-/s/, and S14, who rarely provided [s] responses in the same contexts. We can speculate that speakers like S24 are more likely to utilize the fact that /s/-final stems account for the majority of the stems ending in a coronal obstruent in the language. One might think that S14 is simply being more prescriptive in general, not willing to use the innovative [s]. This possibility is ruled out, however, because this speaker frequently gave the non-canonical [c^h] responses for /t^h/.

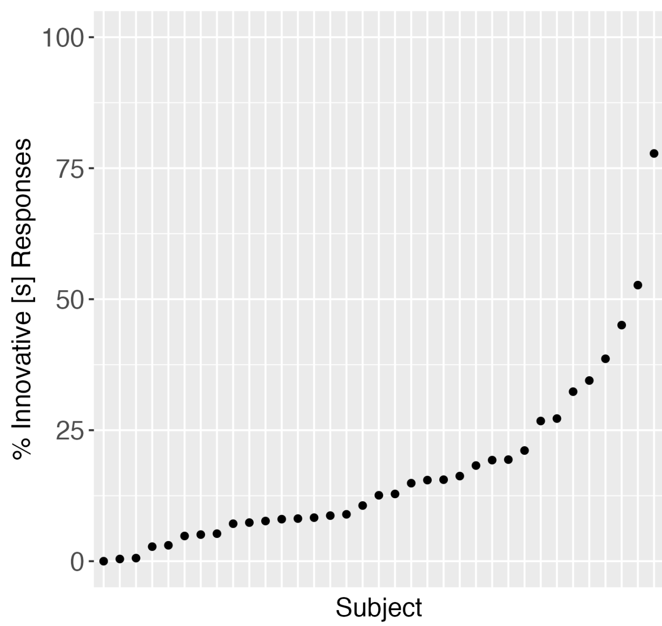

Figure 4.4: The proportion of innovative [s] responses in Experiment 1

Difference in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/

Another environment for which speakers showed a great amount of variation is how sensitive they are to the suffix effect in /t^h/. Individual speakers' proportion of innovative responses for /t^h/-final stems before /-e/ and before /-ɨl/ is plotted in Fig. 4.5. A higher proportion of innovative responses before /-ɨl/ would indicate the suffix effect. First, we can observe that all speakers but one produced a higher proportion of innovative responses before /-ɨl/ than before /-e/ (the data points above the y=x line), suggesting that most speakers are sensitive to the lexical statistics in which /t^h/-final stems are relatively frequent among those suffixed with /-e/, but they are less so among those suffixed with /-ɨl/. However, the speakers differed from each other in their sensitivity to the suffix effect. That is, the farther a data point in Fig. 4.5 deviates from the y=x line, the greater the difference between /-e/ and /-ɨl/ it represents. The proportion of innovative responses ranged from 10.0% to 100% for /-e/, and from 50.0% to 100% for /-ɨl/.

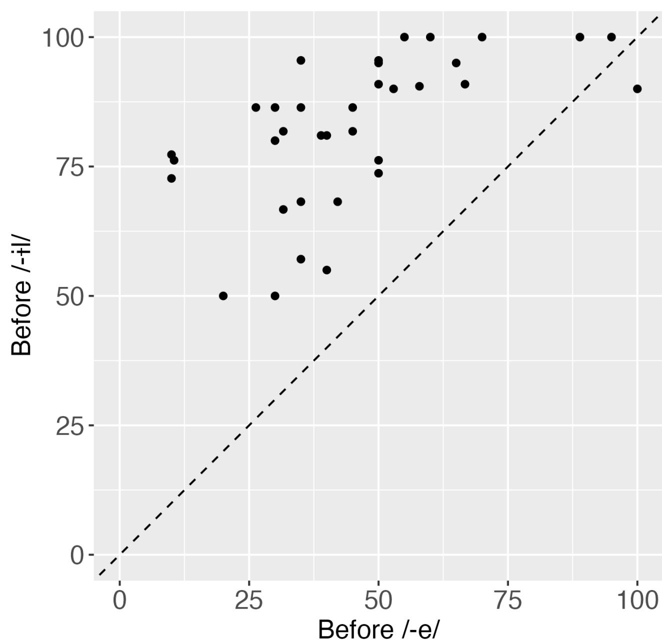

Figure 4.5: The proportion of innovative responses for /t^h/ before /-e/ and before /-ɨl/ in Experiment 1

Proportion of innovative responses in /c/

Speakers also differed substantially from each other regarding whether and how frequently they realized /c/ with a non-canonical variant. Fig. 4.6 shows that the proportion of innovative responses for /c/ ranges from 0% to 100% depending on the speaker, with a wide dispersion. Two contrasting speakers can be found in Fig. 4.3: while S14 always produced [c] for /c/, S24 did so infrequently. Again, this difference cannot be fully explained by differences in general preference towards prescriptive pronunciation; if S14 was trying to be more prescriptive, we would have observed more frequent faithful realizations of /t^h/.

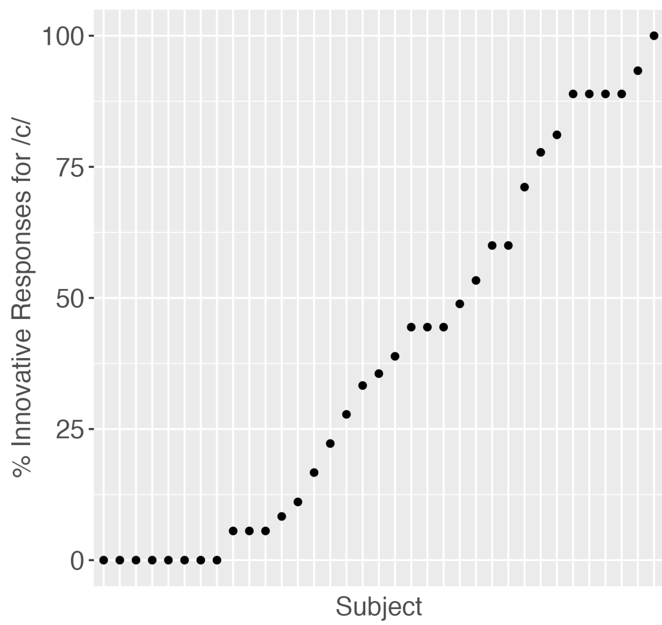

Figure 4.6: The proportion of innovative responses for /c/ in Experiment 1

Proportion of innovative responses in /s/

Speakers were split into two groups with regard to whether they always pronounced /s/ as the canonical [s]. As can be seen in Fig. 4.7, while 22 speakers never produced an innovative variant for /s/, 13 speakers produced a few non-[s] tokens ([c^h]) for /s/. In Fig. 4.3, we see that S21 frequently produced non-[s] variants (specifically, [c^h]) for /s/, whereas S14, S23 and S24 produced [s] for /s/ 100% of the time.

Proportion of innovative [c^h]

It was observed that the speakers frequently produced [c^h] in /s/, /c/, or /t^h/. Fig. 4.8 shows the proportion of [c^h] tokens in these non-/c^h/ contexts, which ranges from 10.8% to 79.8%. S21

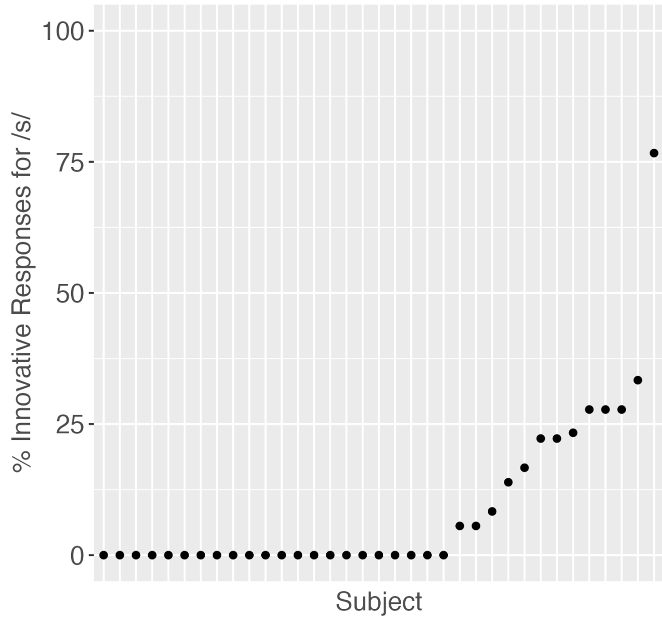

Figure 4.7: The proportion of innovative responses for /s/ in Experiment 1

and S24 in Fig. 4.3 show a contrasting behavior; while S21 extended [c^h] to non-/c^h/ contexts extensively, S24 did so only in a limited way. In fact, S24 did not frequently produce [c^h] even for /c^h/ or for the palatalization environment /t^h-i/. S14 and S23 did not produce [c^h] in /s/ or /c/, but frequently produced [c^h] in /t^h/.

In sum, a substantial degree of individual differences are observed in the realization of stem-final obstruents. Specifically, speakers diverged from the group-level generalizations (i.e. aggregated responses) and from each other not only in the proportion of canonical forms, but also in five distinct ways: (1) the proportion of innovative [s] responses, (2) the difference in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/ (3) the proportion of innovative responses for /c/ (4) the proportion of innovative responses for /s/, and (5) the proportion of innovative [c^h] responses.

I have thus far argued that participants differ across these five variables. In the next section, I conduct a clustering analysis to objectively verify whether distinct speaker types exist, or if the speakers are best understood as a single group.

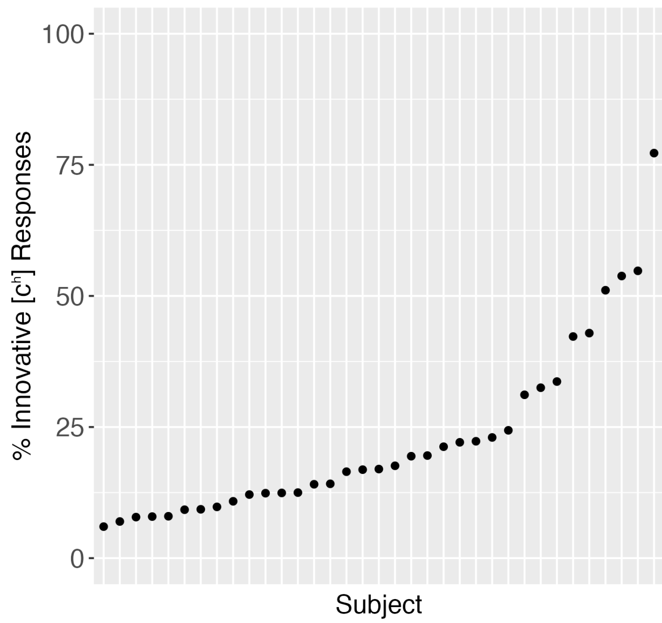

Figure 4.8: The proportion of innovative [c^{h}] responses in Experiment 1

4.3.3 K-means clustering

In this section, I use k-means clustering (Macqueen, 1967) to categorize individual participants into distinct sets according to their response patterns. Specifically, I group together participants who behave similarly with regard to the five variables examined above, namely, (i) the proportion of [s] for non-/s/, (ii) the difference in the proportion of innovative variants between /t^{h}-e/ and /t^{h}-ɨl/, (iii) the proportion of innovative variants in /c/, (iv) the proportion of innovative responses in /s/, and (v) the proportion of [c^{h}] for non-/c^{h}/. By performing a clustering analysis, we can assess similarities and differences among individual participants in an objective way using a machine learning technique. An additional goal of this clustering is to choose a speaker in each cluster that is most representative of that group (the one who is nearest to the centroid), and model these representative speakers' responses.

Cluster analysis is a machine learning technique that categorizes unlabeled data points into groups. The goal of a cluster analysis is to find distinct groups within a data set such that data points in a group have similar characteristics to each other. One of the most widely used clustering techniques is k-means clustering (Macqueen, 1967), which is a method used to partition a set of observations into k clusters, where k is a natural number. The algorithm starts by randomly selecting k data points from the dataset as the initial centroids of the clusters, representing the centers of the clusters. Each data point is then assigned to the nearest centroid based on Euclidean distance. Thus, a data point belongs to the cluster whose centroid is closest to it. After all data points have been assigned to clusters, the centroids are recalculated for each cluster. This process moves the centroid to the center of its cluster. These steps (assigning data points and updating the centroids) are repeated until the centroids no longer change more than some pre-defined threshold or when the pre-specified maximum number of iterations is reached. In this way, k-means clustering can discover underlying patterns in a data set and assign labels to data points in an objective way. In the study of speech sounds, k-means clustering is primarily used to classify vowels based on their acoustic properties (Grabowski & Kuo, 2023; Bissell, 2021; Renwick & Ladd, 2016; Mayer, 2020). Similar to the present study, Zee et al. (in preparation) used k-means clustering to group speakers based on their phonological behavior.

K-means clustering analysis was performed using the *kmeans* function of the *stats* R package (R Core Team, 2018). In selecting the optimal number of clusters, I used two methods: calculating the Average Silhouette Width (Rousseeuw, 1987) and the within-cluster sum of squares (also called inertia; Thorndike, 1953). In the first method, the Silhouette Width measures how similar a data point is to its own cluster compared to other clusters. The Silhouette Width is defined as:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

where *a(i)* is the average distance between sample i and other samples in its cluster, and *b(i)* is the smallest average distance between sample i and all samples in any other cluster (in other words, the average distance between sample i and samples in the nearest cluster.) The value of *s(i)* ranges between -1 and 1. Values near 1 indicate that data point *i* is much closer to the other points in the same cluster than to points of the nearest other cluster. Less optimal cases include values of *s(i)* near 0, which indicate that the data point is located in an intermediate position between the two clusters, as well as values near -1, which indicate that the it would be more appropriate for the data point to be clustered in the nearest other cluster. For a given *k*, the Silhouette Width is calculated for each sample and then averaged over the entire dataset to assess how closely all the samples in the cluster are grouped. According to this method, the optimal number of clusters should be chosen to maximize the Average Silhouette Width. It does not inherently increase (improve) as the number of clusters increase, contrasting with the within-cluster sum of square explained below; instead, it provides a trade-off between having a large number of clusters to ensure a data point is close to other points in its own cluster (cohesion) and having few clusters so that a data point is far from points in other clusters (separation).[4]

In the second method, the within-cluster sum of squared Euclidean distances is calculated, which measures the variance from each data point in the cluster from its centroid, and plotted against the number of clusters. As the number of clusters increases, the within-cluster sum of squares decreases. The optimal value *k* is determined to be the point after which the change in the sum of squared distances begins to level off (forming an "elbow" shape), since a good model is one that has a low within-cluster sum of squares *and* a small number of clusters (which are in a trade-off relationship). Note that because this method requires visual inspection, the researcher's subjective judgment is involved. It is also worth mentioning that different metrics used to determine the optimal number of clusters might yield diverging results, requiring the researcher's judgment call.

In the current analysis, the Average Silhouette Width suggested that the optimal number of clusters is 2 or 3 (2 is slightly better than 3), as presented in Fig. 4.9. Comparing the within-

[4]In an extreme case of good cohesion, every point is its own cluster. Although, in principle, the *a(i)* of the above equation should be 0, and the *s(i)* should become 1, the *s(i)* in such a case is not clearly defined. The *fviz_nbclust* function of the *factoextra* R package (Kassambara & Mundt, 2020) used to visualize the change of Average Silhouette Width as a function of number of clusters in the present study sets *s(i)* = 0 when there is only one data point in a cluster.

cluster sum of squares between 2 and 3 clusters, the decrease was considerable when moving from 2 to 3 clusters (see Fig. 4.10), suggesting that three clusters might be more appropriate. Therefore, I decided to cluster participants into three groups.

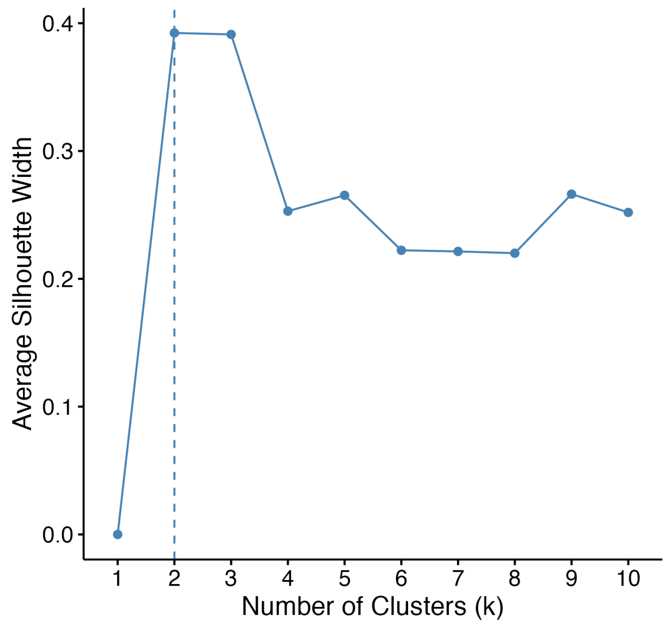

Figure 4.9: Average Silhouette Width for each number of clusters in Experiment 1

To visualize the clusters, I performed principal component analysis using the *fviz_cluster()* function of the *factoextra* package (Kassambara & Mundt, 2020) in R (R Core Team, 2018) to reduce 5 dimensions, the original variables used to cluster the participants, to 2 dimensions, principal component 1 (PC1) and principal component 2 (PC2). In Fig. 4.11, the clusters are plotted against PC1 and PC2. Table 4.3 presents the loadings that indicate which of the original variables or dimensions have the largest effect on each principal component. Loadings can range from -1 to 1, in which a loading close to either end of this continuum indicates that the variable in question has a strong effect on that principal component. It was found that PC1, which separates Cluster 3 from the other two clusters (see Fig. 4.11), is most affected by the innovative [c^h] responses (0.58), the proportion of innovative responses for /s/ (0.57), and the proportion of innovative responses for /c/ (0.55). PC2, which separates Cluster 2 from all other

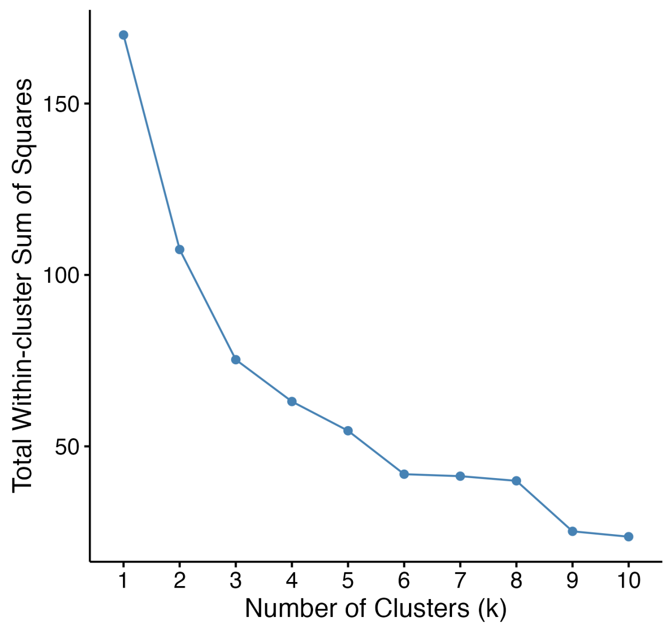

Figure 4.10: Total within sum of squares for each number of clusters in Experiment 1

clusters, is most affected by the proportion of innovative [s] responses (0.81). Differences in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/ did not substantially affect either principal component.

	PC1	PC2
% Innovative [s] responses	0.05	0.81
Differences in the % innovative responses between /t^h-e/ and /t^h-ɨl/	-0.18	-0.30
% Innovative responses in /c/	0.55	0.35
% Innovative responses in /s/	0.57	-0.22
% Innovative [c^h] responses	0.58	-0.28

Table 4.3: Loadings for PC1 and PC2 in Experiment 1

Fig. 4.12 shows all participants' production variants, sorted by their cluster membership. The first 22 participants, from S1 to S9, belong to Cluster 1. Taking together our understanding of the two principal components (see Fig. 4.11 and Table 4.3) and visual inspection of the individual data in Fig. 4.12, these participants can be characterized by low proportions of in-

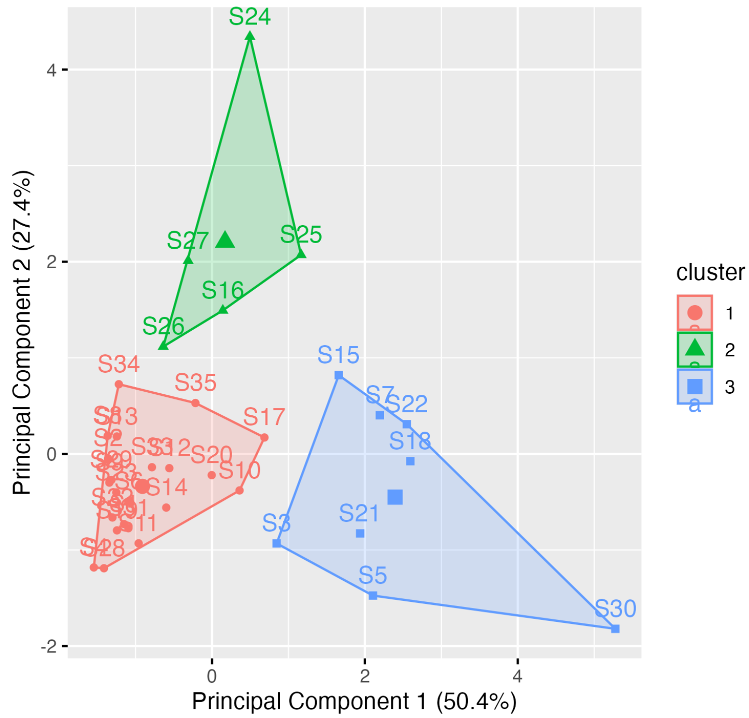

Figure 4.11: Three clusters of participants plotted against principal component 1 and principal component 2.

novative [s] responses for non-/s/, as well as low proportions of innovative [c^h] responses for non-/c^h/. Fig. 4.13A shows the participant of Cluster 1 who is closest to the centroid. Compared with response patterns of the other two participants in Fig. 4.13, each being the individual who is closest to the centroid of the cluster that they belong to, this participant indeed exhibits a lower proportion of [s] for non-/s/ and a lower proportion of [c^h] for non-/c^h/.

Cluster 2 has only five participants, from S16 to S27 in Fig. 4.12. Cluster 2 is distinguished from other clusters in that its participants produced a high proportion of non-canonical [s] responses. The participant who is closest to the centroid of Cluster 2 is presented in Fig. 4.13B.

Finally, the bottom eight participants in Fig. 4.12, from S15 to S7, belong to Cluster 3. These participants can be characterized by relatively high proportions of [c^h] for non-/c^h/. They produced [c^h] even for /s/-final stems, whereas speakers of other clusters rarely produced non-

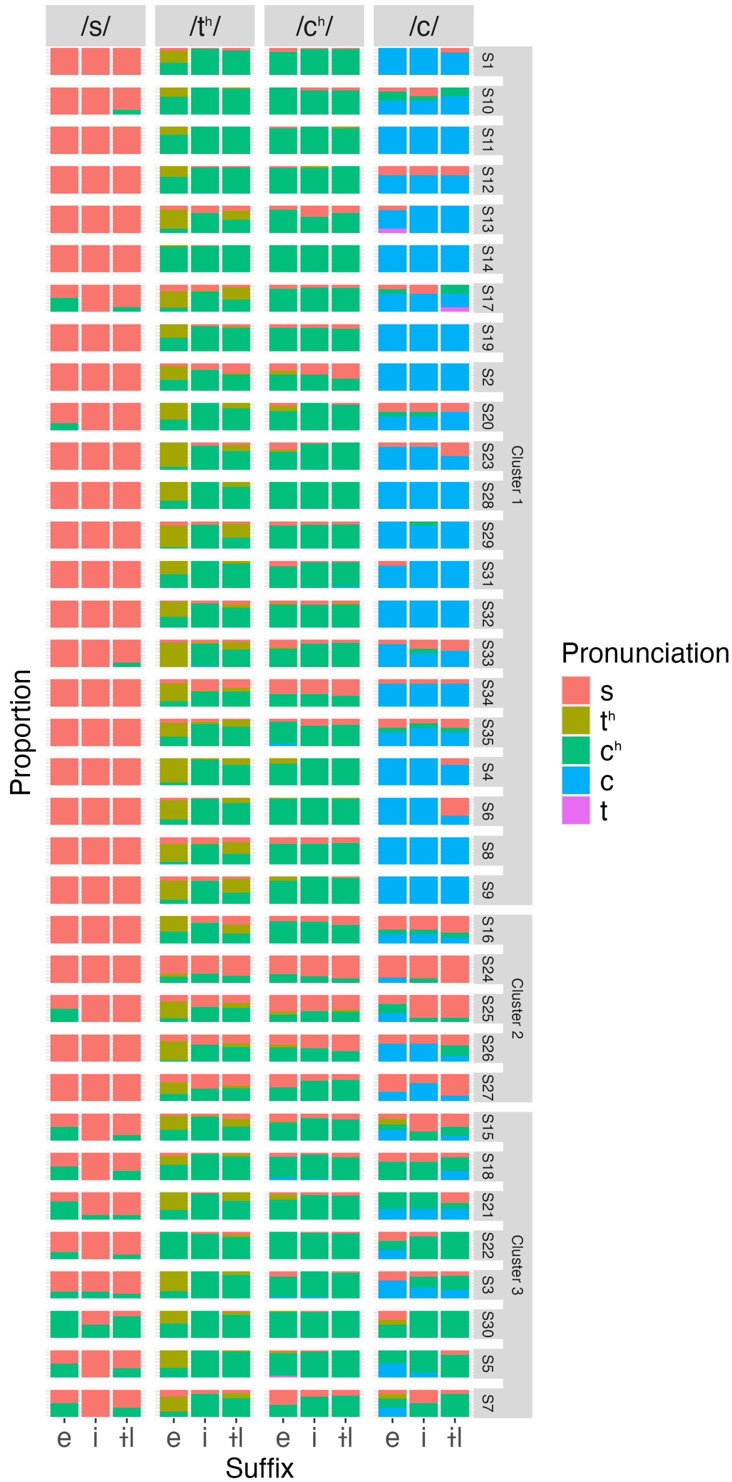

Figure 4.12: Response patterns of individual speakers in each cluster in Experiment 1

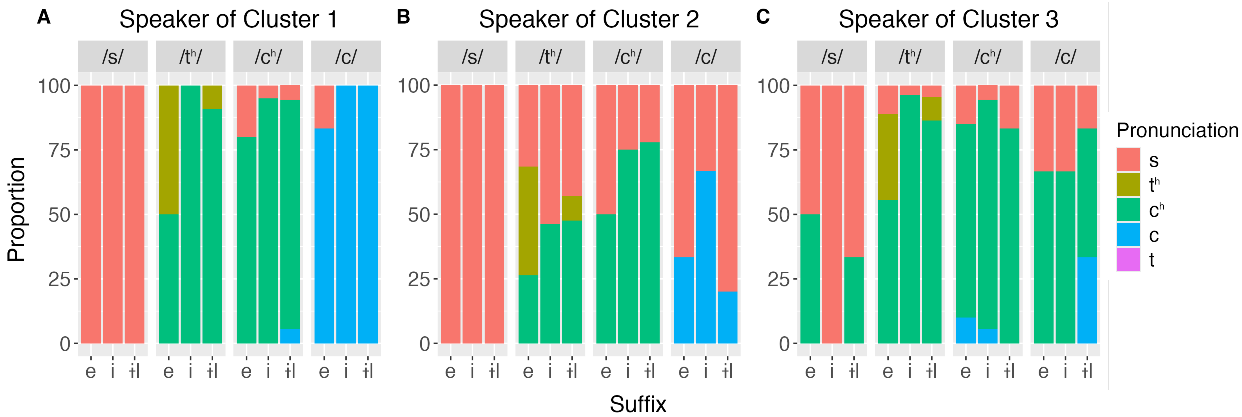

Figure 4.13: Speakers nearest to the centroid of each cluster in Experiment 1

canonical variants for /s/. It is also shown that speakers of Cluster 3 produced [$\mathrm{c^h}$] for stem-final /c/ more frequently than speakers of other clusters did. This pattern is evident in the speaker presented in Fig. 4.13C.

To summarize, k-means clustering analysis grouped participants into three clusters, each group differing from others in a subset of the five dimensions examined in Section 4.3.2, specifically, (i) the proportion of [s] responses for non-/s/, (ii) differences in the proportion of innovative responses between /$\mathrm{t^h}$-e/ and /$\mathrm{t^h}$-ɨl/, (iii) the proportion of innovative responses for /c/, (iv) the proportion of innovative responses for /s/, and (v) the proportion of [$\mathrm{c^h}$] for non-/$\mathrm{c^h}$/. The results suggest that individuals indeed differ from one another in their phonological behavior, with each individual being more similar to members of their own cluster than to members of other clusters. If the clustering analysis had identified 1 as the optimal number of clusters, it would have suggested that speakers are best understood as a single group rather than as different types. In Chapter 6, I model the production patterns of one speaker from each cluster presented in Fig. 4.13.

4.4 Statistical testing

I ran a Bayesian version of logistic regression model to statistically examine the factors that affect the realization of canonical pronunciations. The dependent variable was the binary outcome of canonical vs. non-canonical (reference level). Although I considered five variants as possible pronunciations in previous sections (namely [s], [t^{h}], [c^{h}], [c] and [t]), I categorized the responses into canonical pronunciation and innovative ones in order to render statistical modeling and interpretation of the results easier. While this binary coding may obscure specifics of the variation, it could still provide important insights to the variation; as we will see below, the probability of producing a canonical form can be predicted with various factors. Individual differences in their tendency to be faithful to the lexical form of the obstruent may be explained by their different sensitivity to these factors.

The independent variables are presented in (10), grouped into either categorical or continuous variables. All categorical variables were dummy-coded, and the underlined level is the baseline. The identity of the stem-final obstruent (FINAL C) and the suffix (SUFFIX) were included, as well as their interactions. Although it is impossible to test whether [s] is the most preferred pronunciation among the innovative variants in the current regression model, as the dependent variable is the binary outcome of canonical vs. non-canonical, we can at least partially test the preference for [s] by examining whether /s/-final stems have a higher likelihood of being realized with a canonical form than stems with other final obstruents.

(10) *Independent variables in the logistic regression analysis (Experiment 1)*

- Categorical variables (underlined = baseline level)
  - FINAL C: /s/, /t^{h}/, /c^{h}/, /c/
  - SUFFIX: /-e/, /-i/, /-ɨl/
  - FINAL C × SUFFIX: interaction between stem-final obstruent and suffix
  - RELATIVE FREQ: stem is more frequent (stem > word), stem and word have the

same frequency (stem = word), word is more frequent (stem < word)

- BLOCK: block 1, block 2
- GENDER: female, male
- PURPOSE KNOWN: no, partially, yes
- SPELLING: incorrect, correct

• Continuous variables

- LOG STEM FREQ: the log of the stem's frequency (in its standalone form)
- LOG WORD FREQ: the log of the word's frequency
- LOG STEM FREQ × LOG WORD FREQ: interaction between the log of the stem's frequency and the log of the word's frequency
- STEM FAMILIARITY: response for "How often do you hear and use this word?" on a Likert scale of 1 (Rarely) to 7 (Very often)
- AGE: participant's age
- NORM IS IMPORTANT: response for "How important do you think it is to speak with correct pronunciation?" on a Likert scale of 1 (Not important at all) to 7 (Very important)

Both the frequency of the target word (suffixed form; e.g. /os-i/ → [os-i] 'clothes-NOM') and that of the stem (unsuffixed form; e.g. /os/ → [ot]) were included as predictors. While previous studies have explored the influence of frequency on the realization of stem-final obstruents (e.g. Kang, 2003, 2007; Ko & Jun, 2024), relatively few have undertaken a detailed examination of the effects of word frequency, stem frequency, and their interaction. I expected that words whose *stem* is relatively frequent are less likely to be produced with a canonical form in the experiment, because speakers frequently hear these stems with a neutralized final [t] and thus exposed to its canonical, underlying form relatively infrequently. In comparison, it is likely that higher *word* frequency is associated with higher probability of being realized with a canonical form, because the likelihood of hearing the underlying, lexical form of the stem-final obstruents is higher in suffixed forms compared with unsuffixed forms.[5] Frequency-related factors included the log of the word's frequency (LOG WORD FREQ; the frequency of the inflected form), the log of the stem's frequency (LOG STEM FREQ; the frequency of stems in the unsuffixed form) and an interaction between LOG WORD FREQ and LOG STEM FREQ. A ternary variable of relative frequency between word and stem was also included (RELATIVE FREQ with "stem is more frequent", "word is more frequent", "word and stem have the same frequency"), as the relative frequency between stem and word may affect the word's pronunciation (Hay, 2003). Frequencies were obtained from the corpus used in Chapter 3. When the frequency of a word or stem was 0, it was adjusted to 0.01 (representing a small number that is close to zero), since taking log of 0 results in negative infinity. Finally, participants' responses for "How often do you hear and use this word?", rated on a Likert scale of 1 (Rarely) to 7 (Very often), were entered into the model (STEM FAMILIARITY).

The model also included several variables that were not related to the characteristics of the words themselves. BLOCK was included to test whether the participants produced canonical variants more frequently in the second block, possibly because they realized the purpose of the experiment as they progress and tended to produce the "correct" form of the word. As demographic factors, AGE and GENDER were included. AGE was entered as a continuous variable; the model results did not change considerably when it was coded as a binary variable, dividing the participants into older group (those who are older than the median) and younger group. Whether the participant correctly guessed what the experiment was testing was also included (PURPOSE KNOWN), with three levels "no" (wrong guess), "partially" (partially correct guess) and "yes" (correct guess); see Section 4.2.4 for the coding criterion. SPELLING was a binary predictor indicating whether the participant spelled the stem correctly or not. NORM IS IMPORTANT was entered as a continuous variable. The model results did not change considerably when it was coded binarily, dividing the participants into those who gave a rating of 6 or 7 for the normative question and those who gave 5 or lower.

[5] See Baayen et al. (1997) which shows that morphologically complex words with a high frequency are stored in the mental lexicon in its full form.

In addition to these predictors, random intercepts for participant and stem were also included. When by-participant random slopes for FINAL C, SUFFIX, and frequency measures were added, the model became excessively complex, leading to impractical run times. Consequently, these random slopes were excluded from the final model.

The model was fit using the *brms* package (Bürkner, 2017, 2018, 2021) in R (R Core Team, 2018). 10,000 samples were drawn in each of four chains from the posterior distribution over parameter values. The first 1,000 samples were discarded for warm-up. Following recommendations of Ghosh et al. (2018) on priors for Bayesian logistic regression, a weakly-informative Student-*t* prior with 5 degrees of freedom was used for the intercept and all coefficients. There were no divergent transitions after warm-up. All parameters had an R-hat value of 1, indicating that the model has converged.

Table 4.4 provides coefficient estimates along with their 95% Credible Interval (CrI) for each parameter. In Bayesian analyses, 95% CrI is the range within which one can be 95% certain that the true value of a parameter lies. When the CrI excludes zero, it suggests that the factor of interest credibly affects the outcome. When the CrI includes zero, the probability of an effect in the direction of the coefficient's sign (p-direction) is also reported, representing the proportion of the posterior distribution that shares the sign of the median. This value typically ranges between 50%, i.e. the posterior distribution is centered around zero, and 100%, i.e. the whole posterior distribution lies to either positive or negative side. The closer the p-direction is to 50%, the weaker the evidence for an effect of either direction; conversely, the closer it is to 100%, the stronger the evidence for an effect in that direction. In Table 4.4, parameters whose CrI does not include zero are boldfaced.

I first discuss the effect of FINAL C, SUFFIX and their interactions. First, the coefficient for FINAL C (s) was positive and the CrI did not include zero, which suggests that /s/-final stems were more likely to be produced with a canonical form than the baseline /t^h/-final stems, when the suffix is /-e/. The coefficient for FINAL C (c^h) was also positive and the CrI did not include zero, indicating that /c^h/-final stems were more likely to be produced with a canonical form

Parameter	Estimate	Est. Error	95% CrI lower limit	95% CrI upper limit	p-direction
Intercept	-1.06	0.53	-2.10	0.00	
FINAL C (s)	3.13	0.76	1.66	4.65	
FINAL C (c^h)	1.84	0.49	0.86	2.80	
FINAL C (c)	0.93	0.70	-0.47	2.32	90.96%
SUFFIX (i)	2.49	0.17	2.17	2.82	
SUFFIX (ɨl)	-1.62	0.17	-1.96	-1.28	
FINAL C (s) : SUFFIX (i)	-1.83	0.75	-3.27	-0.32	
FINAL C (c^h) : SUFFIX (i)	-1.99	0.23	-2.44	-1.54	
FINAL C (c) : SUFFIX (i)	-2.53	0.32	-3.17	-1.90	
FINAL C (s) : SUFFIX (ɨl)	1.89	0.49	0.94	2.84	
FINAL C (c^h) :SUFFIX (ɨl)	1.79	0.27	1.26	2.32	
FINAL C (c) : SUFFIX (ɨl)	1.33	0.33	0.68	1.97	
LOG STEM FREQ	0.32	0.26	-0.19	0.82	89.35%
LOG WORD FREQ	-0.10	0.17	-0.45	0.24	72.62%
LOG STEM FREQ : LOG WORD FREQ	0.37	0.24	-0.10	0.85	94.02%
RELATIVE FREQ (stem = word)	-0.21	0.67	-1.52	1.09	62.75%
RELATIVE FREQ (stem < word)	0.97	0.23	0.50	1.43	
STEM FAMILIARITY	0.05	0.07	-0.09	0.18	74.00%
BLOCK (block 2)	0.00	0.08	-0.15	0.15	50.01%
AGE	-0.07	0.21	-0.48	0.34	63.93%
GENDER (male)	-0.25	0.40	-1.04	0.54	73.86%
NORM IS IMPORTANT	0.11	0.20	-0.28	0.50	70.69%
PURPOSE KNOWN (partially)	0.77	0.55	-0.32	1.83	92.22%
PURPOSE KNOWN (yes)	0.74	0.43	-0.10	1.57	
SPELLING (correct)	-0.03	0.28	-0.56	0.52	54.61%

Table 4.4: Results of the Bayesian logistic regression analysis of Experiment 1

than /t^h/-final stems, when the suffix is /-e/. As for FINAL C (c), the coefficient was positive and the CrI included zero with the p-direction of 90.96%, providing moderate to strong evidence that /c/-final stems are more likely to be produced with a canonical form than /t^h/-final stems before /-e/.

As for the effect of SUFFIX, SUFFIX (i) had a positive coefficient, with its CrI not including zero, which indicates that the baseline /t^h/-final stems are more likely to be produced with a canonical form when the suffix is /-i/ than when the suffix is /-e/. This is expected, since palatalization is an automatic process in Korean (the palatalized form was coded as canonical). SUFFIX (ɨl) had a negative coefficient with its CrI not including zero, indicating that stem-final /t^h/ are less likely to be produced with a canonical form when the suffix is /-ɨl/ than when the suffix is /-e/. This confirms the suffix effect, in which [t^h] is preferred before /-e/ than before /-ɨl/.

All of the FINAL C × SUFFIX interaction terms credibly predicted the probability of producing a canonical form. As can be seen in Fig. 4.14, the negative coefficients for FINAL C (s) : SUFFIX (i), FINAL C (c^h) : SUFFIX (i) and FINAL C (c) : SUFFIX (i) indicate that the effect of FINAL C (s), (c^h) and (c) on the outcome is reduced when the suffix is /-i/, compared to the baseline suffix /-e/. This is likely due to the high proportion of canonical forms (palatalized form) in /t^h/ when the suffix is /-i/. The positive coefficients for FINAL C (s) : SUFFIX (ɨl), FINAL C (c^h) : SUFFIX (ɨl) and FINAL C (c) : SUFFIX (ɨl) indicate that the effect of FINAL C (s), (c^h) and (c) increases when the suffix is /-ɨl/ compared to when the suffix is /-e/. This is likely due to the low proportion of canonical forms in /t^h/ when the suffix is /-ɨl/, as shown in Fig. 4.14.

I next investigated the effect of factors related to word or stem frequency. The coefficient for LOG STEM FREQ was positive and its CrI included zero; the p-direction of 89.35% suggests weak to moderate evidence that more frequent stems are more likely to be produced with a canonical variant, contrary to the prediction that a higher stem frequency would be associated with a lower proportion of canonical forms. LOG WORD FREQ did not reliably predict the likelihood of producing a canonical form. Then the prediction that a higher word frequency would be

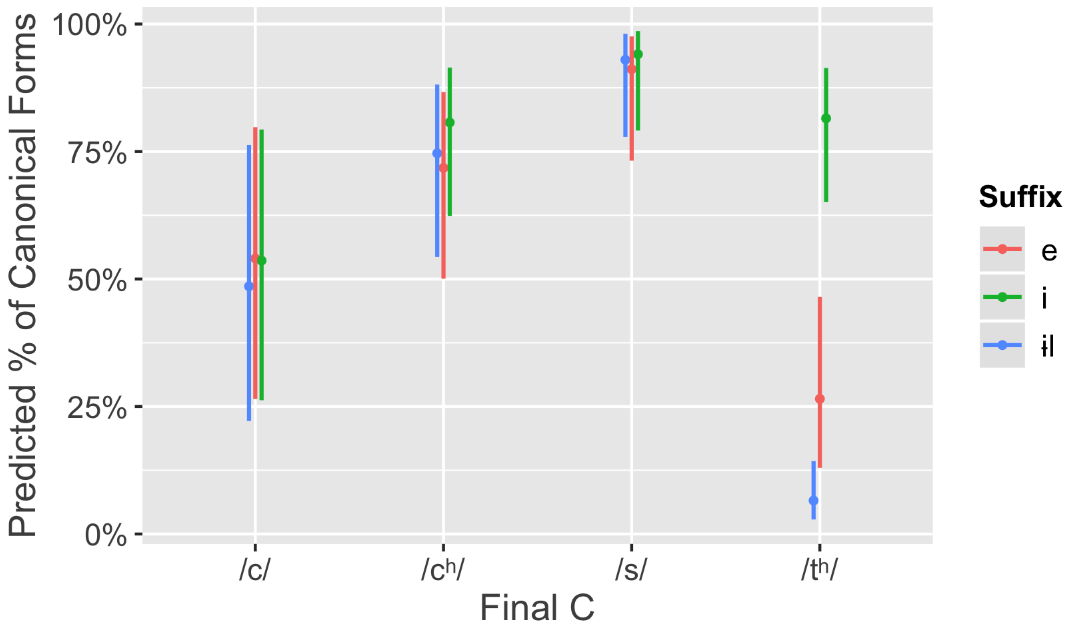

Figure 4.14: Interaction of stem-final obstruent and suffix in Experiment 1

associated with a higher proportion of canonical forms is not borne out. The positive coefficient for LOG STEM FREQ × LOG WORD FREQ, whose CrI included zero and had a p-direction of 94.02%, suggests that the effect of stem frequency (in which a higher stem frequency is associated with a higher proportion of canonical forms) is stronger when the word frequency is also relatively high, as shown in Fig. 4.15. Put differently, the probability of producing a canonical form was high when both word and stem frequencies were high. The categorical variable RELATIVE FREQ credibly predicted the likelihood of producing a canonical form; specifically, the probability of being realized with a canonical form was higher when the word frequency was higher than stem frequency, compared to the other way around. Finally, STEM FAMILIARITY did not credibly predict the likelihood of producing a canonical form.

Turning to the predictors that were unrelated to the characteristics of the words themselves, it was found that BLOCK did not credibly predict the outcome. None of the two demographic factors, AGE and GENDER, credibly predicted the probability of canonical forms. The coefficient

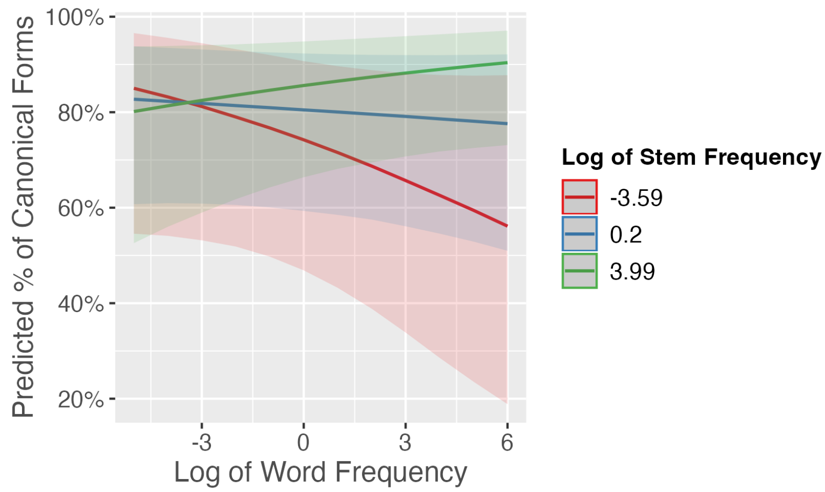

Figure 4.15: Interaction of log of word frequency and log of stem frequency in Experiment 1

for NORM IS IMPORTANT was not credible, either. The coefficient for PURPOSE KNOWN (partially) was positive, and the CrI included zero; the p-direction of 92.22% suggests that participants who figured out the purpose of the experiment partially are more likely to produce canonical forms than those who did not realize the purpose even partially. The coefficient for PURPOSE KNOWN (yes) was also positive with the CrI not including zero, indicating that those who correctly guessed what the experiment was testing are more likely to produce canonical variants than those who did not. Finally, the effect of SPELLING was not credible.

In sum, /s/- and /c^h/ stems are more likely to be realized as a canonical form than /t^h/-final stems when the suffix is /-e/. The effect of final obstruent was conditioned by the suffix, as shown by credible interactions between the identity of the final obstruent and that of the suffix. Further, when the word's frequency is higher than its stem's frequency, the probability of producing a canonical form was higher. Participants were more likely to produce a canonical form when they knew what the experiment was testing. Overall, the variation is predicted by both linguistic factors and individuals' metalinguistic awareness.

4.5 Discussion

Experiment 1 investigated variable realization of stem-final coronal obstruents in Korean, focusing on individual differences. The results confirmed findings of previous studies that [s] is a frequent, preferred variant (Jun, 2010; Albright, 2005, 2008; Ko, 1989; Hayes, 1998; Kang, 2003). We thus observed a high proportion of faithful, canonical responses for /s/-final stems. The effect of suffix, specifically that [t^h] is more frequently observed before /-e/ than before /-ɨl/, was also confirmed, leading to a relatively high proportion of faithful, canonical realization of stem-final /t^h/ before /-e/. Both of these generalizations were found to be statistically reliable in the regression model. Note that the regression model included predictors of word and stem frequency to test the possibility that frequent lexical items (those with stem-final /s/ or /t^h-e/ sequences) are more likely to be realized with its canonical variant. It is interesting to note that even when those frequency-related variables were included in the model, the effect of the final obstruent (that suggests a higher probability of canonical forms for /s/-final stems) was credible, and so was effect of suffix (that suggests a higher probability of canonical forms for /t^h-e/ compared to /t^h-ɨl/). The higher probability of /s/ being realized with a canonical form has been attributed to its high frequency, but the statistical analyses revealed that the frequency measures included in the current model could not explain this effect entirely.

The primary purpose of the experiment was to investigate inter-speaker variation in their pronunciation of stem-final obstruents. Speakers diverged from the group-level generalizations (i.e. aggregated responses) and from each other in the proportion of canonical responses, as well as in five distinct phonological environments: (1) the proportion of [s] observed for non-/s/ (2) the difference in the proportion of innovative pronunciation between /t^h-e/ and /t^h-ɨl/ (3) the proportion of innovative responses for /c/ (4) the proportion of innovative responses for /s/ (5) the proportion of [c^h] in non-/c^h/.

What could be the sources of this inter-speaker variation? In the regression model, several factors were found to reliably predict the variation. Then one can reasonably think that individuals vary in their pronunciation because they differ along these variables that modulate the variation. For instance, the prevalence of [s] responses can be attributed to the fact that Korean is rich in /s/-final stems, and across-speaker differences in their preference for [s] may be explained by differences in sensitivity to such lexical statistics. It was also found that the probability of being realized with a canonical variant was higher when the word's frequency was higher than the stem's frequency, compared to the other way around. Individual speakers may be sensitive to this frequency effect to different degrees, which may contribute to different propensity to produce canonical forms. Finally, the participants who realized the purpose of the experiment produced canonical forms more frequently than those who did not realize it, likely because they were prone to produce what they think is the correct pronunciation for each target once they were aware of what they were tested for.

I found no evidence that the demographic factors investigated in this study, age and gender, reliably predict the proportion of canonical responses. Lack of an age effect may be attributed to the small range of variation in participants' age; most were in their 20s or 30s. I leave it to future studies to investigate whether and how older speakers and younger speakers differ in their pronunciation of stem-final obstruents. The tendency for female speech to be more innovative than male speech has been found in sociolinguistic research, but there was no difference between female and male participants in their likelihood of producing a canonical form in the present study.

Participants were asked to indicate how strongly they believe that it is important to pronounce words correctly. It is plausible to think that those who believe correct pronunciation is important produce canonical forms more frequently. However, individual differences in the proportion of canonical variants was not predicted by their responses to the question that directly asks their tendency to follow the normative rule. This finding suggests that individual differences in phonological behavior cannot be reduced to differences in an extralinguisic factor that is apparently closely related to the variable of interest (in this case, the likelihood of producing a canonial variant).

One factor that might have affected the individual variation but was not included in the regression model is individual speakers' performance in the experimental task. For instance, some speakers might have found it more difficult to understand the instructions about the task than others; some might have been distracted in the middle of the experiment. Although it was hard to measure individual participants' reactions to the experimental task, there are reasons to believe that the individual variation is little affected by factors related to how individuals perform the task. First, because the experiment was held online, the participants performed the tasks in an ideal environment; they were given as much time as they want in producing the target words; they were told to take as long break as they want between the two blocks of recording[6]; they were allowed to participate in the experiment at their most convenient time, and in the most comfortable place. In addition, the task was easy enough that all participants finished the experiment within a reasonable length of time, and their performance was at ceiling for choosing the correct suffix.[7] A similar argument is made in Bennett & Braver (2020); they rule out the possibility that the inter-speaker variation observed in the Xhosa labial palatalization is due to differences in how well individuals understood the wug test, and argue instead that the variation arises due to differences in individuals' grammar. Moreover, the speaker variation is not random, but rather can be accounted for by the same factors that are well established in the literature to explain the variation (Albright, 2005, 2008; Jun, 2010; Hayes, 1998). Therefore, the inter-speaker variation observed in the experiment is unlikely to be best explained by individual differences in how (successfully) individuals perform the experimental task.

In subsequent chapters (Chapter 5 and Chapter 6), I investigate the potential influence of other sources of individual variability, including cognitive traits and linguistic input that one receives in the course of language learning.

[6]It turns out that the duration of the break time does not differ much from speaker to speaker.

[7]Even in cases in which their response deviated from the expected answer, the alternative suffix provided by the participants still made sense in my judgment.

CHAPTER 5

Experiment 2

5.1 Introduction

The primary purpose of Experiment 2, the second half of the longitudinal study reported in this dissertation, was (i) to investigate whether individuals' responses are consistent across different tasks, (ii) to probe individual differences in production of obstruent variants for nonce words, and (iii) to conduct a preliminary analysis on the relationship between the speakers' choice of variants and their cognitive traits, which could be a possible source of individual differences.

Before outlining the tasks included in Experiment 2, I would first like to discuss why it is important to investigate consistency of individual variability across two different experiments. As was mentioned earlier in Chapter 1, some scholars has claimed that all learners of a speech community develop the same language (Chomsky, 1975), which predicts that individual differences are merely random errors. One way to test whether inter-speaker variation is systematic or random is to examine whether individuals' behaviors are consistent across time, and robust to minor variations in task. If individuals are consistent within themselves, the finding would suggest that the variability is systematic and meaningful. A similar argument was made in James et al. (2018), which investigated individual differences in the reading difficulty of object- vs. subject-extracted relative clauses. The authors point out that studies need to directly test whether there are genuine and stable individual differences in syntactic processing to begin with. To this end, they investigated whether consistent individual differences exist across different test items, such that some individuals consistently find object-extracted relative clauses easier to read than others. Studies that conducted two testing sessions to investigate the consistency and systematicity of individual differences include Han et al. (2016)[1], Kong & Edwards (2016), and Wade et al. (2021). In these studies, the same tasks were employed in both testing sessions, and the interval between the two sessions was relatively short.

In Experiment 2, a set of multiple choice questions were employed to address the goal (i) mentioned above, namely to investigate whether the individual differences are stable and reliable. There were several considerations in designing the second task for investigating whether speakers' behavior in this task is consistent with the first one (i.e. production test in Chapter 4). First, the nature of the task should be moderately similar to the production task. That is, the two tasks have to be different from each other in order to test whether participants show consistent responses even across different tasks, but the task effect should not be too large to severely perturb individual behavior and remove most of the consistent patterns that would have otherwise been observed. Second, the results of the two studies must be comparable. Since participants provided a single response in each trial in the production study, they were allowed to choose only one pronunciation for a target word among a number of possible pronunciations in the multiple choice questions as well. An alternative format would be to give a numerical rating for each of the variants for a target word; however, then the outcome of the study would be hard to directly compare with that of the production study.

Unlike in the production study, multiple choice questions presented possible pronunciations in the orthography (as the choice of alternant can be expressed orthographically in the Korean alphabet system.) This could have allowed participants to access a more diverse array of pronunciations that they might not have come up with in the production test. As will be demonstrated later, despite this difference, individual participants' preference for certain variants were observed consistently across the two tasks.

[1] The focus of Han et al. (2016) was to argue for between-speaker variation, or grammar split among the population, and against within-speaker variation. The present study also focuses on across-speaker variability but does not make strong claims about whether within-speaker variation exists. It is anticipated that speakers' responses would still be consistent across different testing sessions even if within-speaker variation exists.

Along with the multiple choice questions, a nonce word production test was conducted to address (ii). As Experiment 1 only employed existing words in Korean, we cannot determine whether the individual differences stem from variability in lexicon or grammar. Because a nonce word test precludes the effect of lexicon, we can test whether the speaker variation can be traced to differences in grammar. If individuals behave similarly across real words and novel words, indicating that the pattern is productive, the results would suggest that the speaker variation does not only stem from differences in the lexicon.

Finally, three cognitive tasks were also included in the experimental session to address (iii), namely, a nonword repetition test (Gathercole & Baddeley, 1989) for assessing phonological working memory, Stroop task (Stroop, 1935) for assessing inhibitory skills (the ability to suppress irrelevant responses), and Autism-Spectrum Quotient questionnaire (Baron-Cohen et al., 2001) to measure autistic traits. More specifically, investigating the relationship between language and cognitive traits will help us understand the extent to which language relies on domain-general cognition and to what extent it is modular (Fodor, 1983). Moreover, if individuals' response patterns in the experiments can be reliably predicted by differences in their cognitive traits, the findings will further document the systematicity of individual variation.

The first cognitive trait of interest was phonological working memory. Phonological working memory is essential for storing and processing verbal and acoustic information. Phonological working memory is known to support a variety of skills including vocabulary acquisition, sentence processing, and reading skills (Adams & Gathercole, 1996; Baddeley et al., 1998; Dufva et al., 2001; van der Schuit et al., 2011; Kormos & Sáfár, 2008). However, its role has been studied mostly within the L2 learning literature, where the ability to accurately repeat nonwords is found to predict better L2 vocabulary acquisition (Baddeley et al., 1998; French, 2006; Speciale et al., 2004), while its influence on first language acquisition has received less attention. Nevertheless, despite the limited research on the relationship between phonological working memory and L1 phonology, it is reasonable to believe that phonological working memory affects an individual's learning of variants for stem-final obstruents.

The task employed to assess the participants' phonological working memory was the nonword repetition test. Nonword repetition test has been widely used to measure individuals' phonological working memory in previous research; it was first used to investigate the relationship between phonological working memory and children's language learning (Gathercole & Baddeley, 1989, 1990; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000), yet has been used for adults as well (Gupta, 2003; Sasisekaran, 2013).

Another aspect of individuals' cognitive traits examined in the present study is inhibitory skills, the ability to inhibit an irrelevant dimension of a stimulus. For instance, when asked to identify the color of the word *red* presented in green ink, the subject is supposed to say "green", but it is likely that they are tempted to say "red". Individuals with better inhibitory control would correctly say "green", successfully suppressing "red".

As with phonological working memory outlined above, more research has been done on the effect of inhibitory skill in the L2 acquisition or bilingualism literature (Lev-Ari & Peperkamp, 2013; Gollan et al., 2011; Festman et al., 2010) compared with L1 acquisition. This is understandable, as bilinguals must switch between languages, suppressing interference from the non-target language in comprehension and production. The findings suggest that individuals with better domain-general inhibition exhibit stronger language control ability. Yet, a number of studies suggest that inhibition is an essential part of language processing even among monolinguals (Lev-Ari & Peperkamp, 2014; Taler et al., 2010; Gernsbacher et al., 1990; El Bouzaïdi Tiali et al., 2021). It has been reported that inhibitory skill is associated with the magnitude of neighborhood density effect on speech perception (Taler et al., 2010), with individual variations in phonological representations (Lev-Ari & Peperkamp, 2014), and with homophone processing El Bouzaïdi Tiali et al. (2021) in monolinguals. In the present case, Korean speakers can potentially encounter phonological neighborhoods or even homophony when they produce a stem with a final coronal obstruent, due to the coda neutralization outlined in Chapter 2. For instance, /pis/ 'comb', /pic^{h}/ 'light', and /pic/ 'debt' can be considered close phonological neighbors, and are homophonous when unsuffixed as they are all realized as [pit] with the coda neutralized; the same holds for /nas/ 'sickle', /nac^h/ 'face' and /nac/ 'day'. As a result, when speakers are producing /pic-ɨl/ 'debt-ACC', it is possible that variants of /pic^h-ɨl/ are activated. We can therefore reasonably expect that individuals' inhibitory control is related to the production of noun stems with final obstruents in Korean.

I employed the Stroop task (Stroop, 1935), a widely-used psychological test to assess inhibitory skills. In the original study, participants were asked to name the ink color of color words, in which the ink color and the meaning of word did not match. For instance, *red* was printed in blue and participants were asked to name the color blue instead of reading the word "red". As controls, they were also asked to name the color of solid squares. It was found that the participants took a significantly longer time to name the color in which the words were printed than to name the color of squares.

I used a manual version of the Stroop task as a proxy for inhibitory skills, following Lev-Ari & Peperkamp (2014). Participants were instructed to press a key to indicate the color of the printed words. There were two critical conditions: (i) congruent, in which the ink color matched the meaning of the color word, (ii) incongruent, in which the ink color did not match the meaning of the color word. In addition, there were trials in which color words were presented in black ink and the participants were asked to identify the meaning of the color word rather than naming the ink color. This condition was added to increase interference.

Lastly, participants' autistic-like traits were assessed. Autistic traits are among the most widely studied cognitive characteristics in relation to phonetic and phonological processing. A number of studies found a link between autistic traits and speech comprehension and production, including sensitivity to prosodic information (Jun & Bishop, 2015; Bishop et al., 2015), sensitivity to lexical knowledge in speech perception (Stewart & Ota, 2008), perceptual compensation for coarticulation (Yu, 2010), coarticulatory production (Yu et al., 2024), and phonetic imitation (Yu et al., 2013; Snyder et al., 2019). However, a few recent studies either did not find a link between individuals' autistic traits and patterns of perceptual compensation (Lai et al., 2022), or found results that contradict previously reported effects of autistic-like traits on phonetic convergence (Wade, 2022). Based on my understanding, the majority of studies have explored the relationship between individuals' autistic traits and their phonetic perception or production, while research on the impact of these traits on phonological patterns is rare.

I assessed the participants' autistic traits using the well-known Autism-Spectrum Quotient (AQ) questionnaire (Baron-Cohen et al., 2001). It comprises 50 questions, and assesses 5 different areas, i.e. social skill, attention switching, attention to detail, communication and imagination, each domain made up of 10 questions. More information on the questionnaire is provided below in Section 5.2.6.

Assessing individuals' cognitive traits poses a challenge due to the various types of executive functions and the multiple tests available for each. Given practical constraints, only three types of executive functions were tested. Additionally, while it is ideal to use multiple tasks to assess a single construct (see Section 5.7), I employed only one task per executive function. Therefore, the findings of this study should be interpreted with caution.

The participants first performed the nonce word test, then answered multiple choice questions, which were followed by a nonword repetition test, Stroop task, and Autism-Spectrum Quotient. The nonce word test preceded the multiple-choice questions to prevent the participants' responses to the nonce words from being influenced by exposure to real words in orthography in the multiple choice questions. As with Experiment 1, all experiment sessions were held online.

5.2 Methods

5.2.1 Participants

27 native speakers of Korean participated in the experiment (10 males, 16 females, and gender of one speaker unreported; age mean: 30.0, range: 21-48) during June and July 2023. One speaker's data were excluded from analysis due to background noise in the recording. Among

the remaining 26 participants, 17 had participated in Experiment 1 between October 2021 and January 2022 (Long Delay Group); for these group of participants, then, there was a 1.5-year period between the two experiments. The other 9 participated in both experiments in June or July 2023, with a 10-day separation between Experiment 1 and Experiment 2 (Short Delay Group). It is plausible to think that participants of Short Delay Group would be more consistent across Experiment 1 and Experiment 2 than those of Long Delay Group are, due to the shorter interval between the two experiments. This issue is addressed in Section 5.3.5. Participants were not informed that the two experiments were related.

Participants were recruited online (Prolific) and by word-of-mouth. They were paid for their participation.

5.2.2 Multiple choice questions

Stimuli

The target words used in the multiple choice questions were the same as those of the production task (N=80; see Section 4.2.2). Unlike the production test, no filler items were included; the participants were explicitly asked to choose their pronunciation for the target words, which made the purpose of the experiment already obvious.

Procedure

In each trial, a sentence containing a target word was presented. For example, for the target word /mit^{h}-e/ ‘under’, the sentence “kojaŋi-nɨn c^{h}imte **mit^{h}-e** sum-ʌ is*-ʌ” ‘The cat is hiding under the bed’ was shown on the screen in Korean orthography. The target word was boldfaced and printed in a bigger font compared to other words in the sentence. The set of sentences used in the multiple choice questions were different from those of the production task. For each target word, five options were provided in Korean orthography, each representing pronunciation of the word with one of the five variants considered, namely [s], [t^{h}], [c^{h}], [c] and [t]. For instance, for the target word /mit^{h}-e/, the options provided were *mise*, *mit^{h}e*, *mic^{h}e*, *mice* and *mite*. Participants were instructed to choose one option that represents their pronunciation for the target word, where the prompt was "How would you pronounce this word?" The order of options were randomized.

5.2.3 Production test with nonce stems

Stimuli

Twenty-five target nonce stems were created, presented in (11). All nonce stems were made monosyllabic, since the majority (24 out of 30) of real stems used in Experiment 1 and Experiment 2 were monosyllabic (see Section 4.2.2). Further, there was no difference in the likelihood of producing a canonical form between monosyllabic and polysyllabic real stems. In creating the target stems, all of which took the form of CV[t], I first made a list of every possible combination of CV followed by [t]. Then I excluded not only the items that have the same pronunciation with existing stems, but also those that have similar pronunciation with existing ones. For instance, the nonce item [p*at] was removed from the list because it was deemed similar enough to the real stem [pat] (/paT/ 'field') to cause participants to base their response on the real stem. Finally, I rated the remaining items, i.e. those determined to be phonologically distant to existing stems, for their well-formedness as a Korean stem on a Likert scale from 1 (impossible to be a Korean stem) to 7 (very good as a Korean stem). The items rated as 6 or 7 were included in the final list of target nonce stems.

Fifteen filler stems were also included in the list of nonce stems. They were monosyllabic C_1VC_2 in which C_2 is a sonorant ([n], [m], [ŋ] or [l]).

(11) Nonce stems

net	nʌt	nut	tut	t*at
pet	p*et	p*ʌt	p*ot	sʌt
sit	s*at	s*ʌt	s*it	cet

cut	cɨt	c*at	c^hat	c^hot
c^hut	t^hʌt	t^hut	t^hɨt	p^hʌt

As was the case in the production test and the multiple choice questions, the target suffixes were /-e/, /-i/ and /-ɨl/, as shown in (12). For each suffix, two carrier sentences were invented in which the target suffix is required in order for the sentence to make sense. For instance, one of the two carrier sentences for the ACCUSATIVE /-ɨl/ was "c^heksaŋ-e ____ noh-ɨsejo" 'Please put down ____ on the desk', where a word inflected with a suffix other than /-ɨl/ would be unacceptable.

(12) Suffixes employed in the nonce word test

Suffix	Gloss
-e	LOC/DAT
-i	NOM
-ɨl	ACC

Procedure

The procedure was the same as the production test of Experiment 1. On trial start, a carrier sentence with a blank was presented, which the participants were expected to complete with the target word. They were instructed to play the audio of the nonce stem (recorded in its unsuffixed form, ending in unreleased [t]), and come up with the inflected form of the stem that would make the sentence semantically most natural. They were asked to read the entire sentence (not just the target word) so that they were less conscious about their pronunciation of the nonce word and produce more natural speech. The participants were allowed to listen to the stem as many times as they want, but were given only one chance to record the sentence. Since the noun suffixes in Korean can be left out in many cases, participants were explicitly told that they *must* attach the suffix when they fill in the blank.

5.2.4 Nonword repetition test

Stimuli

The stimuli set consisted of 30 nonwords, five per each syllable length from 5 syllables to 10 syllables. All syllables in the nonwords were CV.

In creating the nonwords, I first prepared a list of all consonants in Korean that can appear in onsets, [p], [p*], [p^h], [t], [t*], [t^h], [k], [k*], [k^h], [c], [c*], [c^h], [s], [s*], [n], [m], [l], and [h] (N=18), and a list of all monophthongs in Korean, [a], [e], [i], [o], [u], [ɨ] and [ʌ] (N=7). The consonants were ordered in a way that avoided adjacent homorganic consonants or adjacent segments that are both laryngeally marked (that is, tense and aspirated stops) as much as possible, as these sequences may cause articulatory difficulties. The vowels were ordered as [o], [a], [i], [ʌ], [u], [e], [ɨ] (without a specific rationale). I then created a series of CVs, with Cs iterating through the list of consonants and Vs iterating through the list of vowels. After exhausting all 18×7=126 combinations of consonants and vowels, I repeated the procedure with the vowels in reverse order. From this list of syllables, I used the first five syllables to create one 5-syllable nonword, the next five syllables to create another 5-syllable nonword, and so on, until I created all 30 nonwords. None of the substrings of these nonwords formed existing words in Korean.

(13) Stimuli for the nonword repetition task

Syllable number	Nonword
	pos*ahic^{h}ʌ lu
	p*ekɨt*osap^{h}i
5 syllable	nek*ucʌt^{h}ɨmo
	c*atik^{h}ʌpus*e
	hɨc^{h}okʌp*ila

	t*usep^{h}ɨnik*aco
	t^{h}ʌmuc*etɨk^{h}opa
6 syllable	s*ihʌc^{h}ulep*ɨko
	t*asip^{h}ʌcunɨk*e
	t^{h}omac*itʌpek^{h}u
	s*ɨhoc^{h}alip*ʌkut*e
	k*icaponʌt^{h}umesɨ
7 syllable	k^{h}apis*ʌhuc^{h}etoc*ɨ
	t*isʌp^{h}ucelɨp*oka
	c*ʌtuk^{h}enok*ɨmita
	pɨs*ohac^{h}ilʌp*uket*ɨ
	sok*ʌcip^{h}anut^{h}emɨc*o
8 syllable	tak^{h}ipʌs*uhec^{h}ɨlop*a
	kat*osɨp^{h}ecuk*ʌnit^{h}a
	moc*ɨtek^{h}upʌs*ic^{h}oha
	lɨp*ekut*ʌsip^{h}acok*ɨne
	c*imʌt^{h}utak^{h}opɨs*ehuc^{h}ʌ
9 syllable	nak*icʌp^{h}uset*ɨkop*ali
	homɨc*etuc^{h}ɨpik^{h}ʌt^{h}os*a
	p*ulep^{h}okʌt*isak*ecɨnu
	t^{h}ʌmic*atok^{h}ɨpes*uhʌc^{h}ila
	p*okɨp^{h}ʌsut*ecit^{h}ɨnomek*a
10 syllable	k^{h}itʌc*upas*ohɨc^{h}elup*ʌki
	t*asop^{h}ɨcek*unʌt^{h}imac*otɨ
	k^{h}epus*ʌhic^{h}ap*ɨloket*usʌ

Most studies of nonword repetition tests in Korean were conducted with children (e.g. Lee & Sim, 2003; Hwang, 2015; Ha, 2020; Yang et al., 2013), and so the stimuli used in the present study differed from those of these previous studies in two ways. First, while their stimuli did not involve sounds that children cannot articulate reliably, such as /c/, /c^{h}/, /c*/, /s/, /s*/ and [l], the nonwords used in this study did contain these sounds. Second, the stimuli of the current study had a greater number of syllables than those created for children in previous studies, in order to make the task challenging enough to observe individual variation among adult speakers.

Procedure

In each trial, participants first listened to a nonword by clicking on a button. They then clicked on another button to start recording and spoke aloud the nonword they had just heard. After speaking, they clicked the same button to finish recording. Each nonword was played only once, and participants had only one chance to record their response.

Before beginning the task, participants were informed that they would be listening to nonwords that do not exist in Korean. They were instructed to repeat as much of the nonword as they could remember, even if they could not recall the entire form.

Analysis

There are two different ways of scoring participants' performance in a nonword repetition task. The first method is to regard a response as either correct (1) or incorrect (0). The second is to grade each segment (consonant or vowel) produced by the participant as correct (1) or incorrect (0) in relation to its target segment in the stimulus (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000). I adopted the second approach to allow for a more nuanced, gradient scoring across individuals. More specifically, if a segment produced by the participant is present in the target form, 1 was added to the score; if a segment in the participant's production was not present in the target, it did not add anything to the score.

For each target nonword, I divided the number of segments scored as correct by the total number of segments contained in the target. Then I calculated the average of these scores for each participant and multiplied it by 100. The nonword repetition test score, therefore, ranged from 0 to 100 (%).

5.2.5 Stroop task

Procedure

Four colors were used, namely, red, blue, green and yellow. Each session consisted of training, practice and test phases. The training phase was designed to let participants familiarize themselves with the key-color mapping, i.e. D-red, F-blue, J-green, K-yellow. The symbol "@@@@" was printed in one of the four colors, and the participants pressed the key that corresponded to the ink color. The training consisted of 20 trials. In the practice phase, the color words were printed in one of red, blue, green, yellow, and black. As mentioned earlier, if the ink color was one of red, blue, green and yellow, the participants had to press the key that corresponded to the ink color. If, on the other hand, the ink color was black, they had to press the key that corresponded to the meaning of the word. Feedback was provided to the participants in the practice phase (correct, wrong) and there were 40 trials. The test phase consisted of two blocks, each comprising 100 trials. No feedback was provided in the test phase.

Analysis

I obtained two measures of individuals' performance from their responses. First is the difference in the accuracy between congruent and incongruent stimuli, calculated by subtracting the accuracy in the incongruent trials from that of congruent trials. Accuracy is the number of trials in which the participant correctly identified the color of the text divided by the total number of trials. Second is the different in the reaction time (RT) between congruent and incongruent stimuli, calculated by subtracting the RT in the incongruent trials from that of congruent trials. I present the results based on the differences in RT, rather than the differences in accuracy (see Section 5.5 for more information).

5.2.6 Autism-Spectrum Quotient

Stimuli and procedure

As mentioned earlier, Baron-Cohen et al.'s (2001) Autism-Spectrum Quotient questionnaire was used, consisting of 50 items. I translated the original items into Korean. Items were presented one at a time on the screen.

Analysis

In each question, respondents are asked to indicate the degree to which they (dis)agree with a given statement by choosing one of the four options "definitely agree", "slightly agree", "slightly disagree" and "definitely disagree". Approximately half the questions were worded to yield a "disagree" response by a highly autistic individual, and half an "agree" response. The AQ can range from 0 to 50; the higher the score, the more autistic traits an individual possesses. In Baron-Cohen et al.'s (2001) scoring, each item scores 1 point if the respondent indicates the presence of an autistic-like behavior, whether mild or strong; "definitely agree" and "slightly agree" were treated the same, as were "definitely disagree" and "slightly disagree". However, some studies, such as Hoekstra et al. (2008), assign higher weights to responses indicating strong autistic behavior. In this approach, each item can score between 1 and 4 points, which makes the minimum AQ score 50 and the maximum 200. This continuous scoring method allows for a greater degree of individual differences. I present the results of both scoring methods in Section 5.5.

5.3 Results: multiple choice questions

I first present all the speakers' aggregated responses in the multiple choice questions. Figure 5.1 shows the distribution of [s], [tʰ], [cʰ], [c] and [t], grouped by suffix and underlying consonant. Comparing this result against Fig. 4.1, aggregated responses in the production experiment (Experiment 1) and those in the multiple choice questions shared important features in common. First, [s] was generalized to /tʰ/-, /cʰ/- and /c/-final stems to a moderate degree. In addition, [tʰ] was more frequent before /-e/ than before /-ɨl/, in both /tʰ/-final and in /cʰ/-final stems. Such similarities in the response patterns between the two experiments suggest that participants at the group level were consistent in their responses across different task types and different time points.

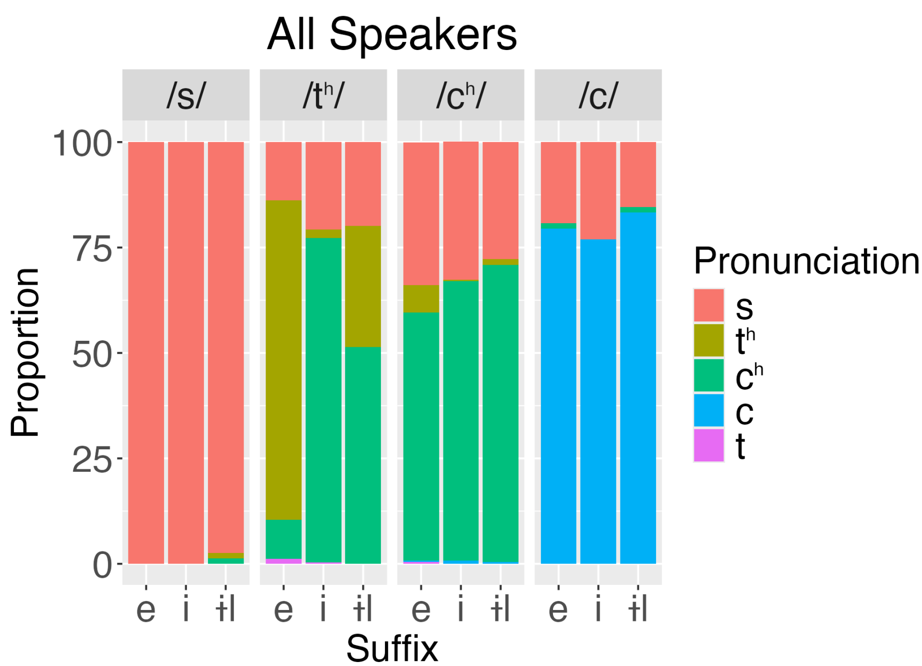

Figure 5.1: Aggregated responses of all speakers in Experiment 2

One qualitative difference was also observed between the two experiments, however. In Experiment 1, participants produced [cʰ] in /s/-final and /c/-final stems. Such generalization was rarely observed in Experiment 2. Given that [cʰ] was not observed in these contexts in the corpus data (see Fig. 3.1), the responses in the multiple choice questions match the corpus frequencies more closely than those in the production test do in this respect.

Interestingly, there was a small proportion of [t^h] responses in /t^h/ before /-i/, despite this being illegal in Korean due to the obligatory palatalization. This likely occurred because the target words were presented in written form, reflecting the underlying representation. For instance, when participants saw the target word *pat^h-e*, they were inclined to choose [pat^h-e] as their pronunciation.

5.3.1 Proportion of canonical variants

The proportion of canonical variants in individual speakers' responses is presented in Fig. 5.2. A moderate degree of inter-speaker variation was observed, with the proportion ranging from 37.5% to 76.2%. Participants were evenly distributed within this range. This is slightly lower than the individual participants' proportion of canonical variants in Experiment 1, which ranged from 42.3% to 84.7% with an outlier of 25.7% (see Fig. 4.2). In both experiments, every participant produced at least some non-canonical responses.

The following section presents participants' responses in the same phonological environments in Experiment 2 as was examined in Experiment 1 (Section 4.3.2): (1) innovative [s] responses (the proportion of [s] responses observed for non-/s/ words) (2) the difference in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/ (3) innovative responses for /c/ (4) innovative responses for /s/, and (5) innovative [c^h] responses (the proportion of [c^h] responses in non-/c^h/ words).

5.3.2 Speaker variation in more specific environments

Proportion of innovative [s] responses

As was the case in Experiment 1, one phonological environment in which individual variation was observed is the degree to which [s] is extended to /t^h/, /c^h/ and /c/. As shown in Fig. 5.3,

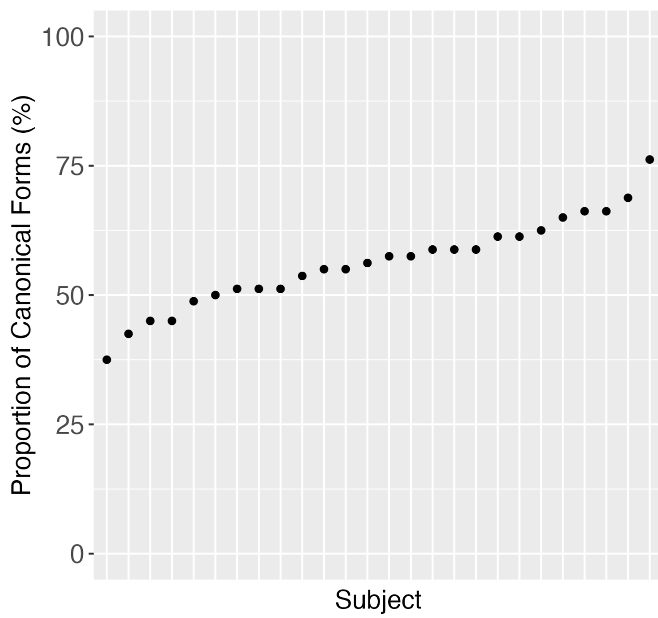

Figure 5.2: Individual speakers' proportion of canonical forms in Experiment 2

the proportion of [s] responses for non-/s/ stems ranges from 0% to 51.9%. This range is similar to that of Experiment 1, shown in Fig. 4.4 (0% - 52.7%, with an outlier of 77.8%), even though Experiment 1 has 9 more participants than Experiment 2. In both experiments, the speakers were evenly distributed within each proportion range.

Difference in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/

Fig. 5.4 shows the proportion of innovative responses for /t^h/-final stems before /-e/ and before /-ɨl/ suffixes. As was the case in the production of /t^h/-final stems (see Fig. 4.5) all but one participants chose innovative responses more frequently before /-ɨl/ than before /-e/ (above the y=x line). The speaker who provided innovative responses more frequently before /-e/ than before /-ɨl/ in Experiment 1 was not the same one as the one who did so in Experiment 2. In Experiment 2, all speakers selected innovative responses for /t^h-e/ 50% or less of the time.

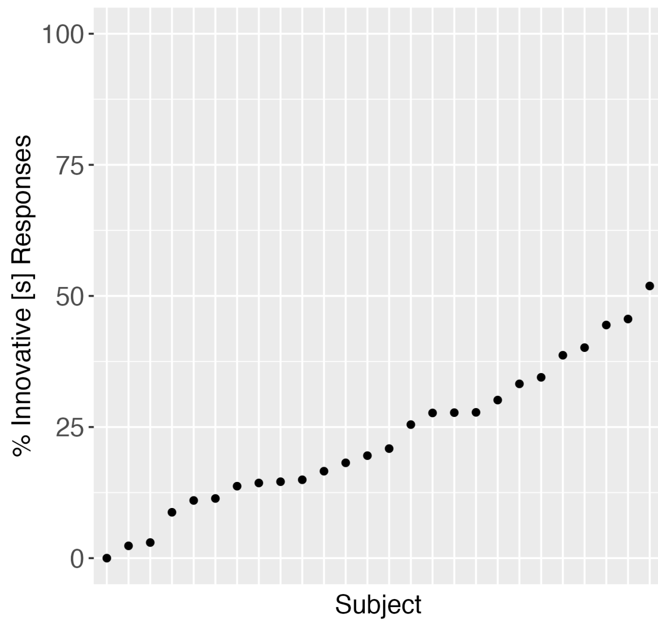

Figure 5.3: The proportion of innovative [s] responses in Experiment 2

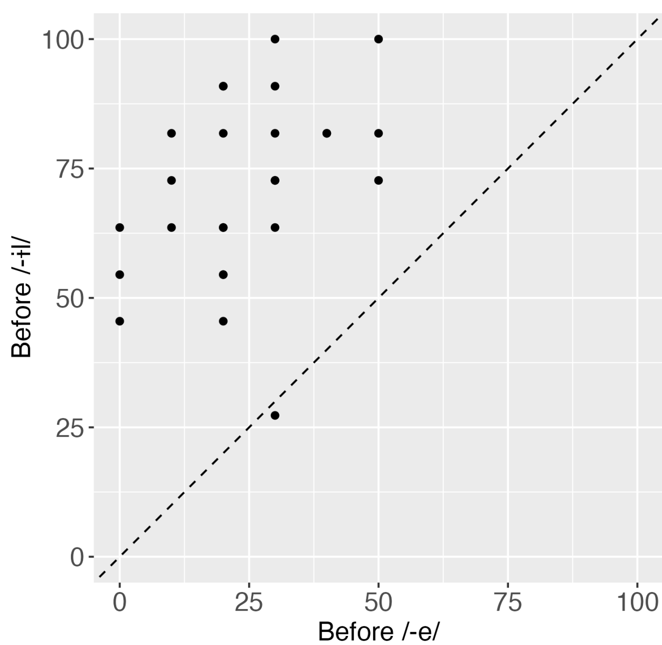

Figure 5.4: The proportion of innovative responses for /t^h/ before /-e/ and before /-ɨl/ in Experiment 2

Proportion of innovative responses in /c/

Similar to the results of Experiment 1, speakers differed substantially from each other regarding how frequently they realized /c/ as a (non-)canonical response in Experiment 2. As presented in Fig. 5.5, the proportion of innovative responses for /c/ ranges from 0% to 100% depending on the speaker, although there was only one speaker who produced 100% [c].

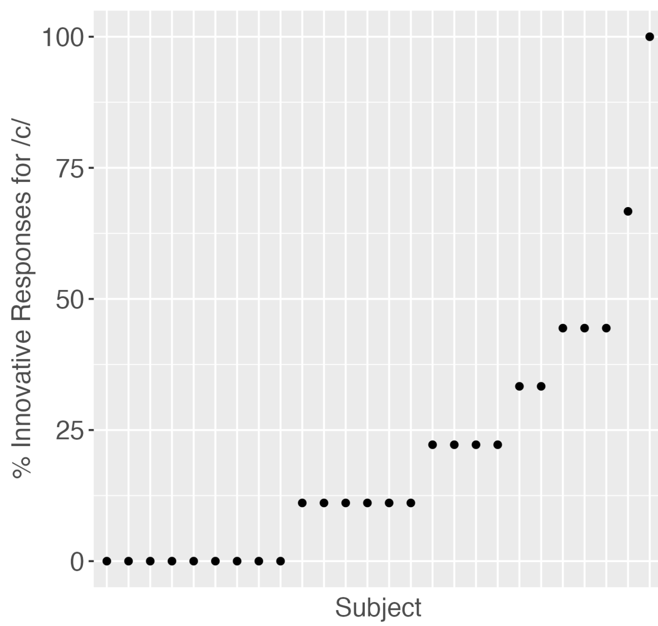

Figure 5.5: The proportion of innovative responses for /c/ in Experiment 2

Proportion of innovative responses in /s/

As Fig. 5.6 shows, only two speakers selected innovative responses for /s/, accounting for 7.7% of the participants. Both of them provided innovative responses 10% of the time. In comparison, a higher proportion of participants (37.1%) produced innovative variants in /s/ in Experiment 1.

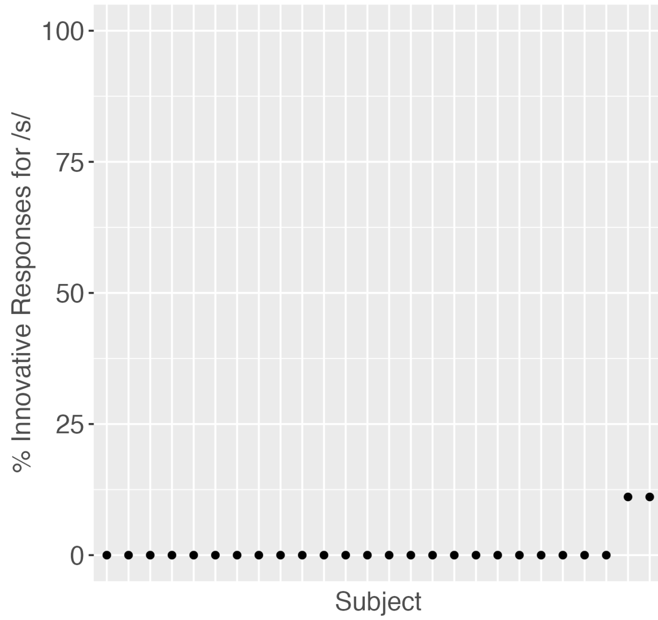

Figure 5.6: The proportion of innovative responses for /s/ in Experiment 2

Proportion of innovative [c^h]

Fig. 5.7 shows the proportion of non-canonical [c^h] responses in individual speakers' responses. Speakers produced [c^h] for non-/c^h/ as low as 6.0% and as high as 27.8% of the time. This range is both narrower and lower than that found in Experiment 1 (10.8% to 79.8%; see Fig. 4.8). The difference between the two experiments seems to arise from [c^h] responses in /s/ and /c/ in Experiment 1; as mentioned earlier, [c^h] was rarely chosen as a response for /s/ and /c/ in Experiment 2.

5.3.3 K-means clustering

As was done in Experiment 1 (see Section 4.3.3), individual participants in Experiment 2 were categorized into distinct groups according to their response patterns using k-means clustering (Macqueen, 1967). I used the same set of five variables, namely (i) the proportion of [s] for non-/s/, (ii) the difference in the proportion of innovative variants between /t^h-e/ and /t^h-ɨl/, (iii) the

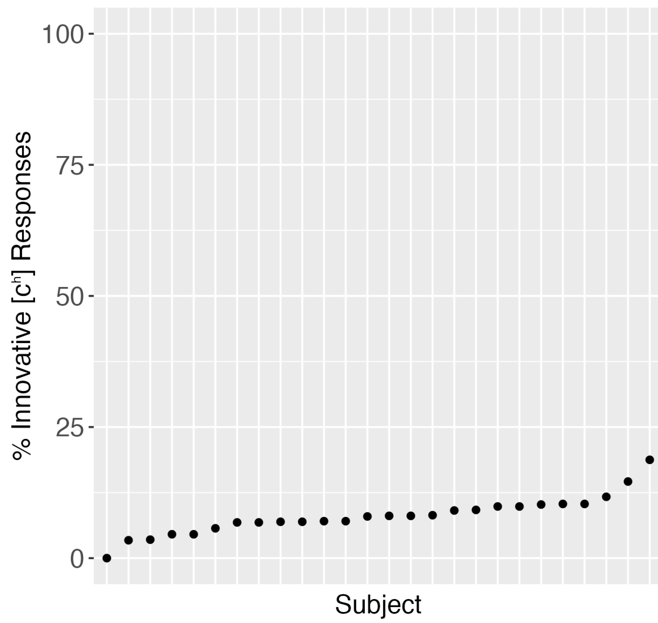

Figure 5.7: The proportion of innovative [c^h] responses in Experiment 2

proportion of innovative variants in /c/, (iv) the proportion of innovative responses in /s/, and (v) the proportion of [c^h] for non-/c^h/, as the dimensions for clustering. Performing a clustering analysis with data in Experiment 2 had two goals. First, clustering can reveal similarities and differences among individual participants that may not be easily identified with simple visual inspection of the data. In addition, by comparing results of the cluster analyses in Experiment 1 and 2, we can assess whether the same individuals tend to cluster together in both experiments – did participants who showed similar behavior in Experiment 1 pattern similarly in Experiment 2 as well?

The k-means clustering analysis was performed in the same way as done in Experiment 1 (see Section 4.3.3). The *kmeans* function of the *stats* R package (R Core Team, 2018) was used. I consulted both the Average Silhouette Width (Rousseeuw, 1987) and the within-cluster sum of squares (also known as inertia or elbow method; Thorndike, 1953) in choosing the optimal number of clusters. The Average Silhouette Width was highest when *k*=5, as shown in Fig. 5.8, suggesting that five clusters provide the most appropriate partitioning of the data. The within-cluster sum of squares for different values of *k*, shown in Fig. 5.9, further supports this, as the decrease from 4 to 5 clusters is considerable, while the decrease from 5 to 6 is relatively small. Therefore, the optimal number of clusters is considered to be 5.

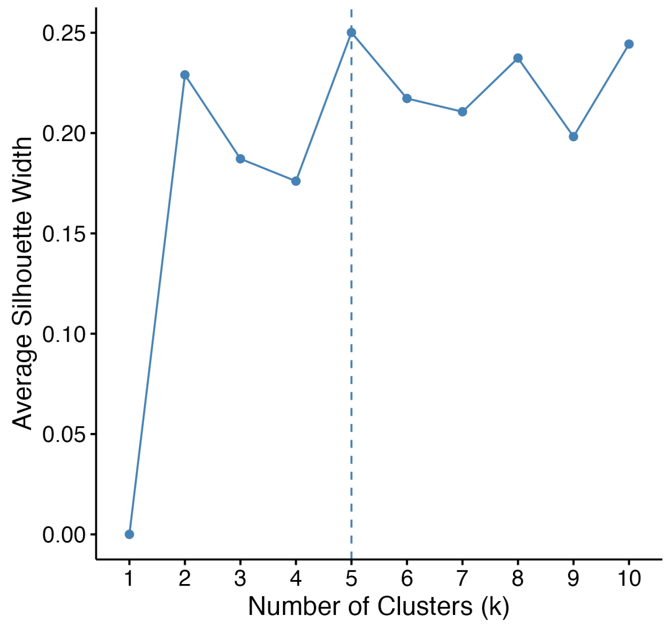

Figure 5.8: Average Silhouette Width for each number of clusters in Experiment 2

Fig. 5.10 plots the five clusters against PC1 and PC2, and Table 5.4 provides the loadings for each principal component, a number between -1 and 1 that indicates the degree to which each original variable contributes to the principal component. A larger absolute value indicates that the variable in question has a stronger effect on the principal component. PC1, which explains 41.0% of the variance of the data, is most affected by the proportion of innovative [s] responses (0.63), followed by the proportion of innovative responses for /c/ (0.59). The positive loadings of innovative [s] responses and innovative responses for /c/ suggest that PC1 can roughly be considered a tendency toward innovative pronunciations, or non-adherence to orthography. PC2, which explains 25.6% of the variance, is influenced by the proportion of innovative responses for /s/ and innovative [c^h] responses most strongly (-0.56 for both), followed by the differences in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/ (-0.53). The negative load-

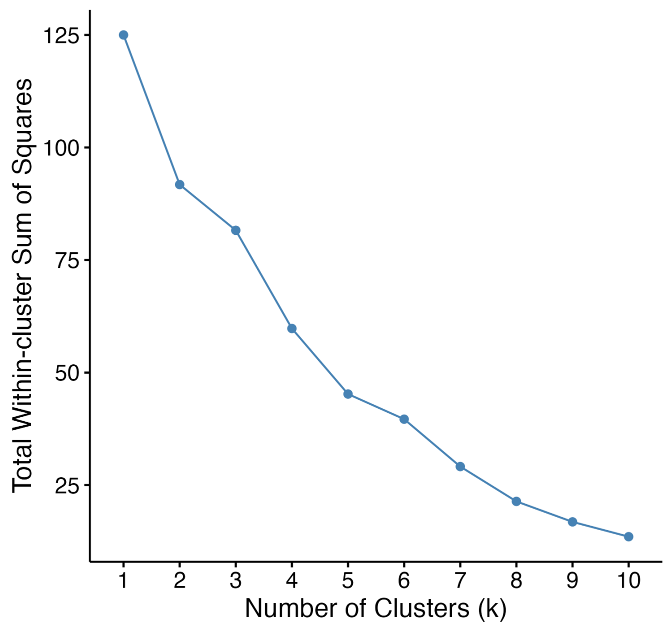

Figure 5.9: Total within sum of squares for each number of clusters in Experiment 2

ings of innovative responses for /s/ and innovative [c^h] responses suggest that PC2 indicates an unwillingness to choose innovative forms for /s/ and innovative [c^h] responses. The negative loading of the differences in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/ suggests a smaller difference in the proportions, or insensitivity to the suffix effect in /t^h/.

	PC1	PC2
% Innovative [s] responses	0.63	0.02
Differences in the % innovative responses between /t^h-e/ and /t^h-ɨl/	-0.19	-0.53
% Innovative responses in /c/	0.59	-0.29
% Innovative responses in /s/	0.28	-0.56
% Innovative [c^h] responses	-0.37	-0.56

Table 5.4: Loadings for PC1 and PC2 in Experiment 2

Characteristics of each cluster can be better understood once we look at all participants' production variants, presented in Fig. 5.11, as well as the response patterns of the participant in each cluster who is closest to its centroid, presented in Fig. 5.12. The first nine participants

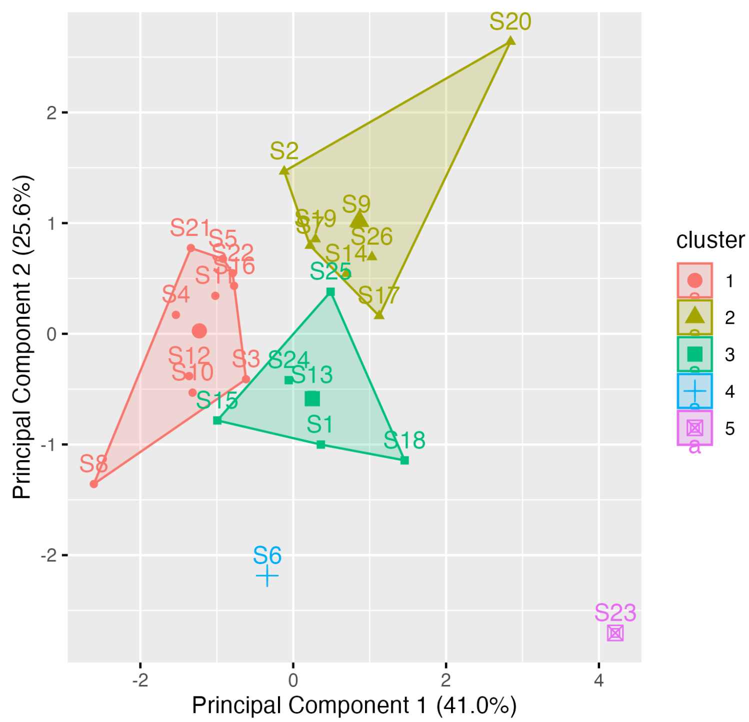

Figure 5.10: Five clusters of participants plotted against principal component 1 and principal component 2.

in Fig. 5.11, from S10 to S8, belong to Cluster 1. We can see in Fig. 5.10 that most of the points within this cluster are grouped closer to the left side of the plot, indicating lower values of PC1, which is most associated with the proportion of innovative [s] responses and the proportion of innovative responses for /c/. The representative speaker of this cluster shown in Fig. 5.12A indeed exhibits a relatively low proportion of innovative [s] responses and no innovative response in /c/.

The next eight participants, from S14 to S9, belong to Cluster 2. Cluster 2 has higher values of PC1 than Cluster 1, which means that the members of Cluster 2 have a relatively high proportion of innovative [s] responses and innovative responses for /c/. This is confirmed when comparing members of Cluster 1 and those of Cluster 2 in Fig. 5.11, and when comparing Fig. 5.12A and Fig. 5.12B.

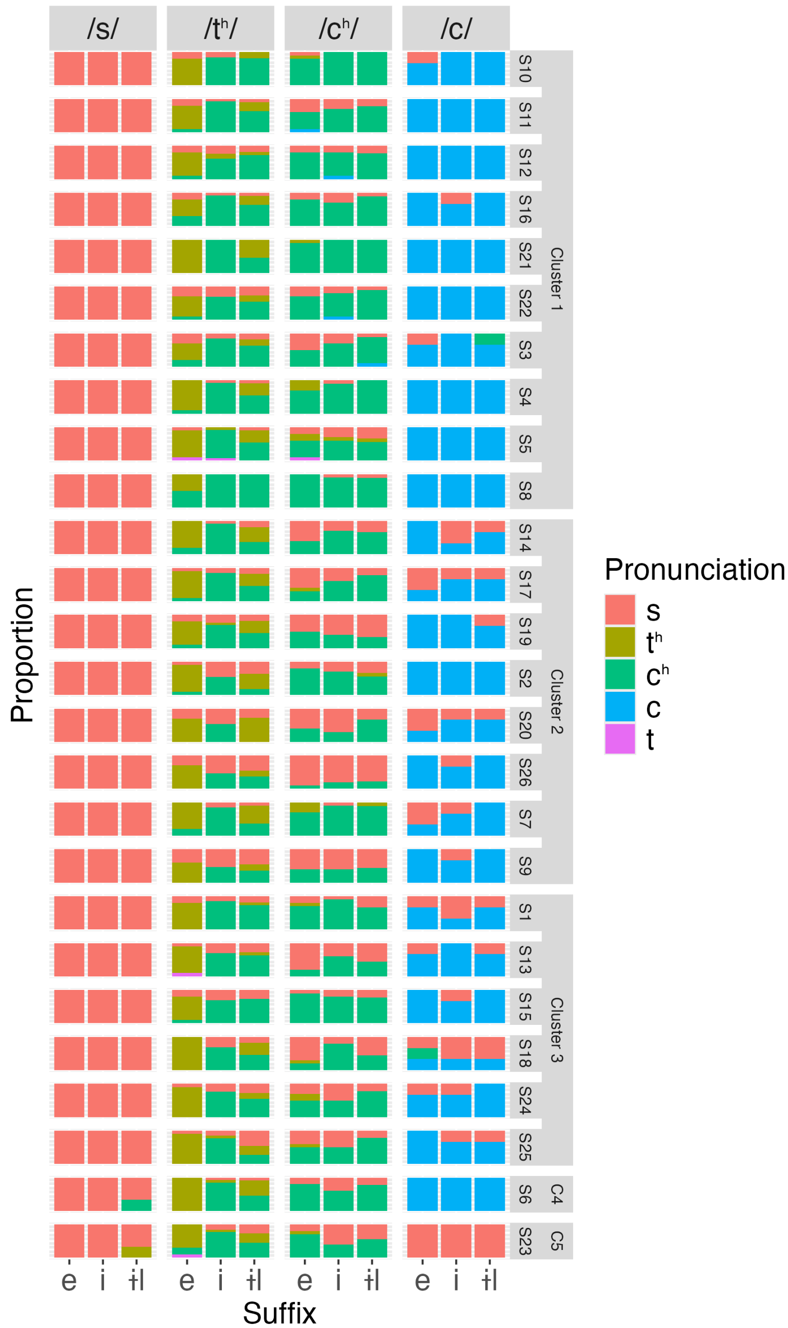

Figure 5.11: Response patterns of individual speakers in each cluster in Experiment 2

Cluster 3, which includes six members (S1 to S25), spans a similar range as Cluster 2 along the PC1 axis. This indicates that Cluster 3 shares similar characteristics with Cluster 2 in terms of innovative [s] responses and innovative responses for /c/. What seems to primarily distinguish

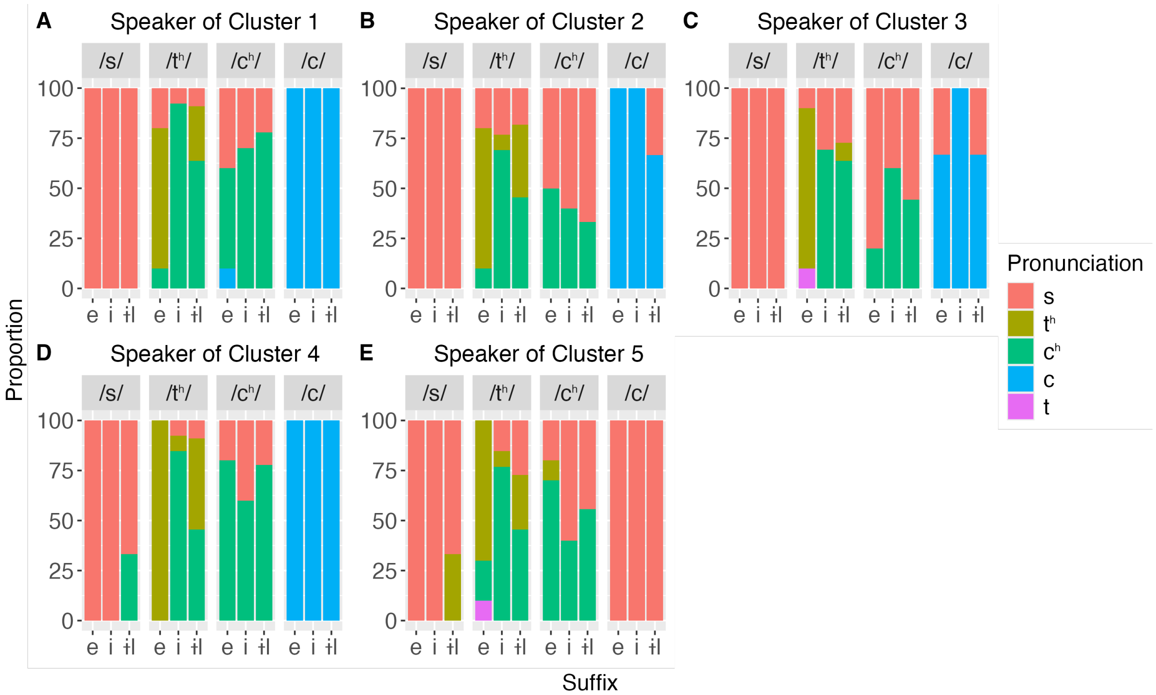

Figure 5.12: Speakers nearest to the centroid of each cluster in Experiment 2

Cluster 3 from Cluster 2 is that the difference in the proportion of innovative responses between /tʰ-e/ and /tʰ-ɨl/ is larger in Cluster 3 than Cluster 2: note that this property contributes considerably to PC2 (Table 5.4), and the two clusters are separated by PC2 (Fig. 5.10). A comparison of Fig. 5.12B and Fig. 5.12C highlights this difference.

Cluster 4 has only one individual (S6) in it, presented in Fig. 5.12D. Cluster 5 also has only one member (S23), presented in Fig. 5.12E. We can see in Fig. 5.10 that these two clusters (individuals) are distinguished from other clusters based on PC2, which is most strongly affected by the proportion of innovative forms for /s/-final stems (along with innovative [cʰ] responses). These two speakers indeed are the only ones who chose non-[s] variants for /s/-final stems. The primary difference between the two is their responses for /c/: while S6 never chose innovative forms for /c/, S23 always provided innovative ([s]) responses for /c/. This is reflected in Fig. 5.10, where these two individuals are distinguished from each other along the PC1 axis, which is strongly affected by the proportion of innovative responses for /c/.

To summarize the results of *k*-means clustering analysis in Experiment 2, the individual participants were grouped into five clusters. We have thus verified using an objective measure that different speaker types indeed exist. Two of them had only one individual in it, suggesting that an individual speaker can be substantially different from all other speakers in a data set. Among the other three clusters, both similarities and differences were found.

5.3.4 Clustering in Experiment 1 and Experiment 2

In Section 4.3.3 and Section 5.3.3, k-means clustering analyses were performed within each experiment. In this section, I examine whether the same individuals tend to cluster together in both experiments. That is, I ask, did participants who showed similar behavior in Experiment 1 pattern similarly in Experiment 2 as well?

To address this question, a contingency matrix was created with the 26 individuals who participated in both experiments (Fig. 5.13), rows and columns representing clusters from Experiment 1 and Experiment 2, respectively. Each cell represents the number of participants clustered into corresponding groups across both experiments. Colors indicate the count of participants, with darker colors representing higher counts. The cells that have the largest number of participants (N=5) are those that belong to Cluster 1 in Experiment 1, and one of Clusters 1, 2, or 3 in Experiment 2. This is expected, because Cluster 1 in Experiment 1 had a larger number of participants than any other clusters, and the great majority of participants belonged to one of Cluster 1, 2 or 3 in Experiment 2. The small sample size (N=26) makes it difficult to assess whether the members of the same cluster in Experiment 1 were more likely to belong to the same cluster in Experiment 2. Among these 15 participants who were assigned to Cluster 1 in Experiment 1, those assigned to Cluster 1 in Experiment 2 may be considered to be consistent across the two experiments, as both Cluster 1 in Experiment 1 and Cluster 1 in Experiment 2 were characterized by a high proportion of canonical forms. In comparison, those who were assigned to Cluster 2 or 3 in Experiment 2 may be considered less consistent, as response patterns of Clusters 2 and 3 are comparatively more innovative. Finally, there were four speakers who belonged to Cluster 3 in Experiment 1 and to Cluster 1 in Experiment 2. These speakers are likely to have produced a high proportion of innovative [c^h] responses in Experiment 1, but did not exhibit such pattern in Experiment 2.

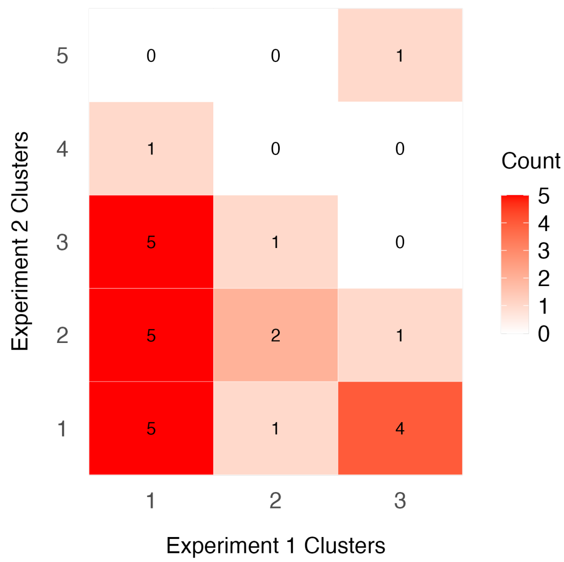

Figure 5.13: Contingency matrix of the clustering in Experiment 1 and Experiment 2

In sum, comparing the cluster membership in Experiment 1 and Experiment 2, some speakers were more consistent than others in their responses across the two experiments. The following section investigates speaker consistency in more detail.

5.3.5 Speaker consistency across Experiment 1 and Experiment 2

This section compares the results of Experiment 1 and Experiment 2, focusing on whether and to what extent speakers were self-consistent. I calculated the Pearson's correlation coefficients between a variable in Experiment 1 and the same variable in Experiment 2. Because there were five variables of interest (see below), I adjusted the p-values to correct for multiple testing by multiplying each of the original p-values by 5.

As can be seen in Fig. 5.14, there was a significant positive correlation between the proportion of canonical forms selected in the two tasks (r=0.58, p<0.01), indicating that speakers who produced more canonical variants in Experiment 1 were more likely to choose canonical pronunciation in Experiment 2 as well. This finding suggests that speakers were largely self-consistent in their responses even though different tasks were employed and the two experiments were separated in time.

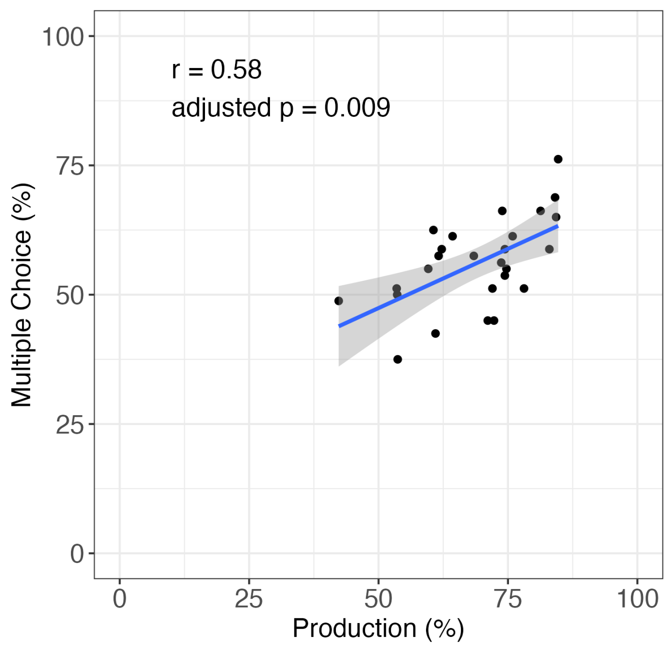

Figure 5.14: Correlations between canonical responses in Experiment 1 and Experiment 2

Further, for the phonological environments examined in Section 4.3.2 and Section 5.3.2, I calculated the correlations between individual speakers' responses in Experiment 1 and 2. Because the proportion of innovative responses in /s/-final stems in Experiment 2 exhibited little individual variability (see Fig. 5.6) to begin with, I did not calculate the correlation for this context. Correlations between the two experiments were thus calculated for (i) the proportion of innovative [s] responses, (ii) the difference in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/, (iii) the proportion of innovative responses for /c/, and (iv) the proportion of innovative [c^h] responses.

First, individual speakers' probability of producing an [s] for non-/s/ in Experiment 1 was significantly correlated with their probability of choosing an [s] for non-/s/ in Experiment 2

(Fig. 5.15A; r=0.57, p<0.05), indicating that those who were more likely to produce a non-canonical [s] in the production test was also more likely to choose a non-canonical [s] response in the multiple choice questions as well. For the difference in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/, there was a non-significant and weak positive correlation (Fig. 5.15B; r=0.14, p=1). Similarly, a non-significant positive correlation was found between individual speakers' probability of providing innovative responses for /c/ in Experiment 1 and that of Experiment 2 (Fig. 5.15C; r=0.42, p=0.17). Finally, a non-significant and weak positive correlation was also found in the probability of providing innovative [c^h] responses for non-/c^h/ in Experiment 1 and that of Experiment 2 (Fig. 5.15D; r=0.27, p=0.88). The variability in the proportion of innovative [c^h] responses in Experiment 2 was relatively small, and so it is unsurprising that the correlation was weak and non-significant.

The non-significance of the positive correlations may be due to the small sample size (N=26), and some of these might have reached significance if the sample size had been larger. Further, the Bonferroni correction applied to adjust the p-values in the calculations above is known to be conservative (Armstrong, 2014; Benjamini & Hochberg, 1995), which also might have led to the non-significant results. With the current sample size and the conservative method used to test for significance, significant positive correlations were still observed between Experiment 1 and Experiment 2 in the proportion of aggregated canonical responses, as well as in the proportion of innovative [s] responses.

As mentioned in Section 5.2.1, there were two different cohorts of participants; Long Delay Group had a 1.5-year interval between the two experiments, while the interval was around 10 days for Short Delay Group. Are participants of Short Delay Group more consistent in their responses than those of Long Delay Group? Or, are members of Long Delay Group not self-consistent at all? To address this issue, I separately investigated the two cohorts' proportion of canonical pronunciations and non-canonical [s] variants in the two experiments (that is, the two variables for which the participants as a whole exhibited consistency; see Fig. 5.14 and Fig. 5.15A). The results showed that positive correlations were found regardless of the interval

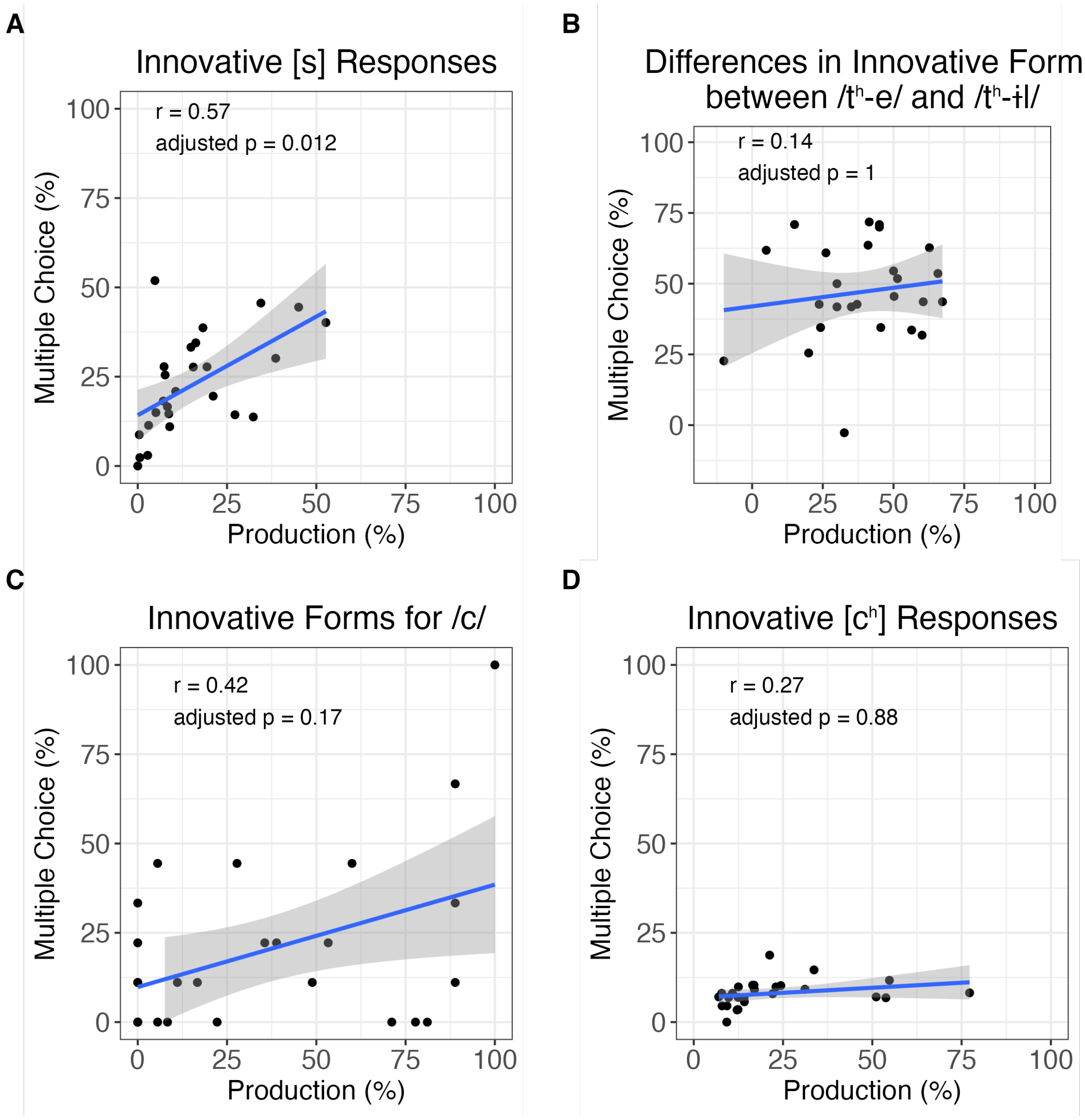

Figure 5.15: Correlations between responses in Experiment 1 and Experiment 2 in four phonological environments

between the two experiments, both in the proportion of canonical variants (Fig. 5.16) as well as in the proportion of [s] for non-/s/ (Fig. 5.17). I did not test the statistical significance of these correlations since the sample sizes are very small. The finding that participants were largely self-consistent across two different tasks, regardless of the length of interval between the two sessions, suggests that inter-speaker variation is not substantially affected by the type of task.

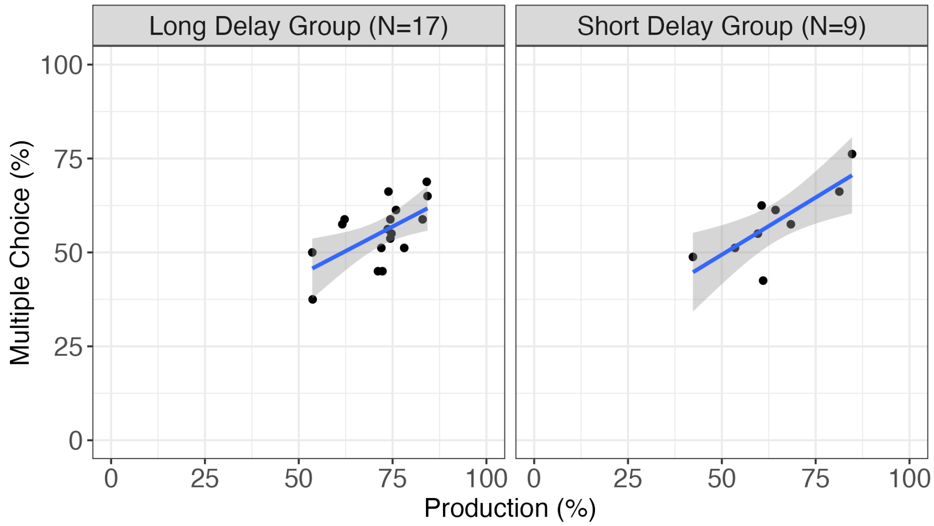

Figure 5.16: Correlations between the proportion of canonical responses in Experiment 1 and Experiment 2, separated by Group

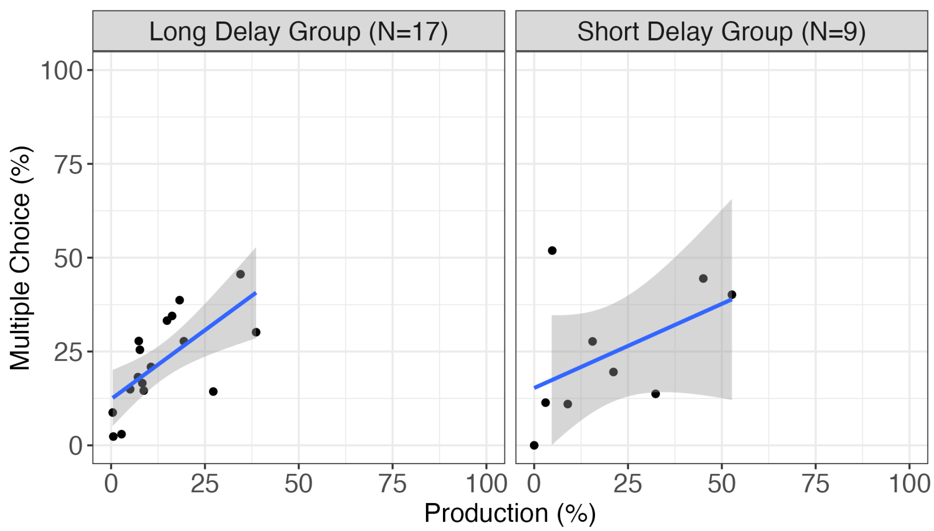

Figure 5.17: Correlations between the proportion of [s] responses for non-/s/ in Experiment 1 and Experiment 2, separated by Group

Next, I examined whether the participant correctly guessed the purpose of the experiment (specifically, Experiment 1, as it was obvious what the multiple choice questions were asking in Experiment 2) affected their consistency. More specifically, it is possible that the observed consistency is driven mostly by those individuals who realized what the experiment was testing?

This is a likely scenario, as they were probably more conscious about their pronunciation. (Recall that this group of participants were found to produce canonical variants more frequently; see Section 4.4). To investigate this possibility, I divided the participants into two groups, one group with those who guessed the purpose of the experiment "correctly" and the other with those who guessed it "incorrectly" or "partially correctly". (Recall from Chapter 4 that individuals who were coded as "partially correct" did not credibly differ in their propensity to produce canonical forms from those who were coded as "incorrect", justifying grouping them into a single category.) As can be seen in Fig. 5.18, both groups were consistent across the two experiments, rejecting the possibility that only those aware of the experiment's purpose were driving the consistency. The finding that even those who were unaware of the purpose exhibited consistency suggests that the consistent responses reflect individuals' genuine linguistic knowledge rather than some meta-linguistic awareness or task effects. This finding must be interpreted with caution, though, due to the small sample sizes in each group (N=19, N=7).

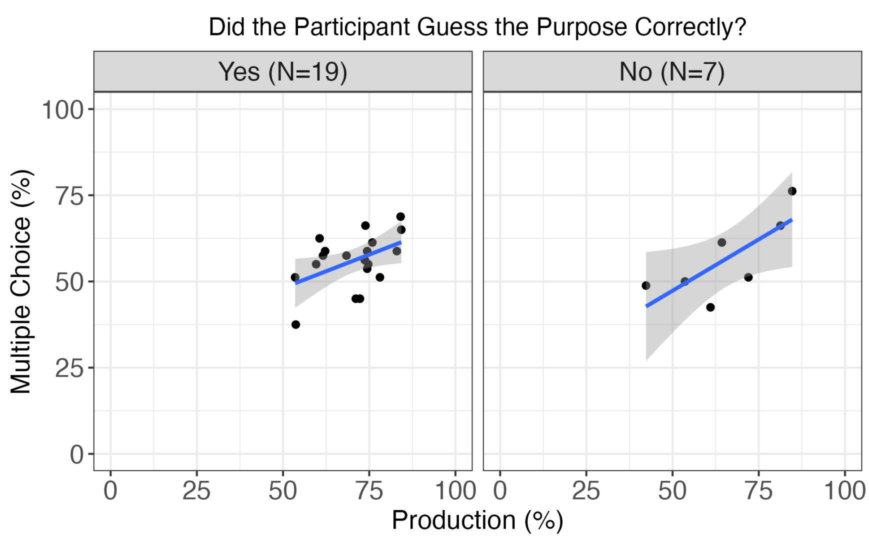

Figure 5.18: Correlations between the proportion of canonical responses in Experiment 1 and Experiment 2, separated by whether the participant correctly guessed the purpose of the experiment

5.4 Results: nonce word test

Two participants were excluded from the analysis, because they frequently produced a pause between stem and the suffix. For instance, in trials in which they were expected to attach [-e] suffix to the stem [net], they produced the stem with an unreleased [t], followed by a slight pause, and the suffix /-e/, rather than the expected [nese], [net^{h}e], [nec^{h}e], [nece], [nete] (or something else). It is hard to regard these responses as inflected forms of stems, and no speakers produce the nouns in this way in real life speech.[2]

Fig. 5.19 shows aggregated responses of the remaining 24 participants in the nonce word test. The responses were predominantly [s], regardless of the suffix (93.8% before [-e], 96.5% before [-i], and 95.3% before [-ɨl]). Among the non-[s] responses, presented in Fig. 5.20, [c^{h}] was the most common (1.9% before [-e], 2.6% before [-i], and 2.8% before [-ɨl]), although comparison among non-[s] responses might not be very meaningful due to their overall low frequencies. We thus see a mismatch between the speakers' responses for the nonce words and those of the real words (see Fig. 4.1 for the speakers' responses in Experiment 1; recall that the nonce word test was also a production task, similar in format to Experiment 1). Crucially, the great majority of the responses were [s] in the nonce word test, while [t^{h}], [c^{h}], [c] and [t] were only minor variants; in the real word test, the non-[s] variants accounted for substantially higher proportions of the responses (although [s] was the most frequent variant.)

At the individual level, I compared the proportions of [s] responses in the nonce words with proportions of [s] responses in the real words (regardless of whether canonical or non-canonical). Fig. 5.21A shows that individuals' production of [s] responses in the novel words are positively correlated with their [s] responses in the real words in Experiment 1. Similarly, Fig. 5.21B demonstrates that they are also positively correlated with individuals' [s] responses in the multiple-choice questions of Experiment 2.[3] The trendlines in both Fig. 5.21A and Fig. 5.21B

[2]Such responses may be regarded as the participants' strategy to avoid producing the inflected form, thereby not having to guess the pronunciation to be used in the inflected form.

[3]The slope of the trendline is more steep in Fig. 5.21B compared with Fig. 5.21A, indicating that the responses in

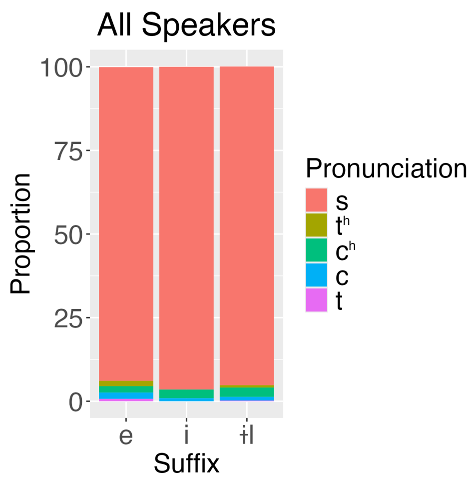

Figure 5.19: Aggregated responses of all speakers in the nonce word test

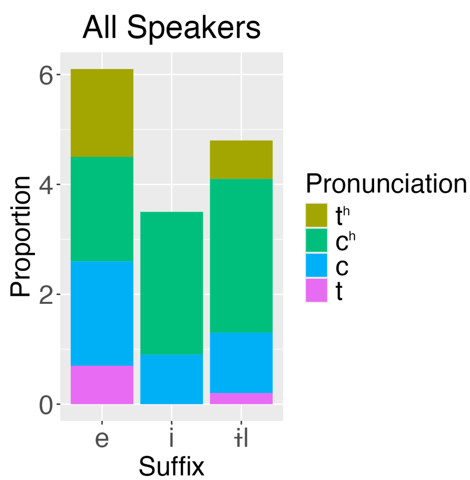

Figure 5.20: Aggregated non-[s] responses of all speakers in the nonce word test

were not steep due to the small degree of individual differences in the nonce word test, resulting from a ceiling effect on the proportion of [s] responses. Among the 24 participants, 12 always produced [s] in the nonce word test, and all but one produced [s] more than 90% of the time. Although it was challenging to reliably determine if individuals' responses for the nonce words could be predicted by their responses for the real words, a trend emerged in which those who

the nonce word test are more closely correlated with those in the multiple choice questions than those in the real word production test. This is likely because both the novel word test and the multiple choice questions task were conducted in the same testing session (Experiment 2).

provided a higher proportion of [s] responses in the real words tended to do so in the novel words as well.

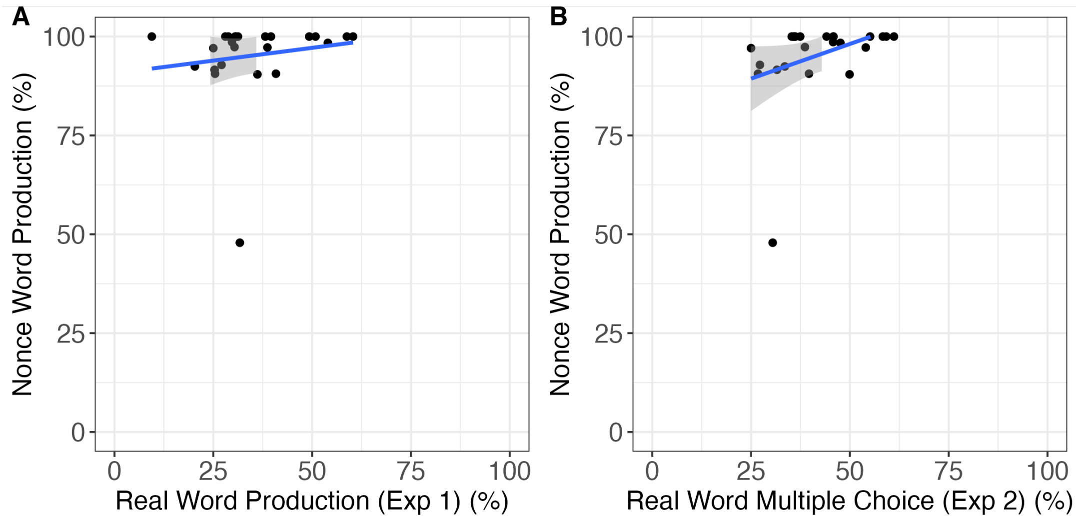

Figure 5.21: Proportion of [s] responses in the nonce words compared with the real words in Experiment 1 (A) and Experiment 2 (B)

5.5 Results: cognitive tests

To investigate the relationship between individuals' cognitive traits and their responses in the experiments, Pearson's correlation coefficients were calculated between one of the three cognitive measures (nonword repetition test scores, Stroop task scores, and AQ scores) and one of the twelve phonological measures (proportion of canonical responses, proportion of non-canonical [s] responses, differences in the proportion of innovative responses between /t^h-e/ and /t^h-ɨl/, proportion of innovative responses in /c/, proportion of innovative responses in /s/, and proportion of non-canonical [c^h] responses, in either Experiment 1 or Experiment 2.) Given the small number of participants who completed the cognitive tests (N=26), this analysis should be considered preliminary. As is shown in Table 5.5, none of the 36 correlations turned out to be statistically significant, when the Bonferroni correction is applied to the *p*-values to correct for multiple testing. However, two observations are noteworthy. First, there was a clear relationship between Stroop task performance and the proportion of canonical forms (where the coefficient was significant before the *p*-value was adjusted to correct for multiple testing.) Second, when scores from each subdomain of the AQ questionnaire were correlated with the phonological measures (not presented in Table 5.5, which only shows the results for total AQ scores), a significant correlation emerged: the attention switching scores were correlated with the proportion of [c] responses for /c/ in Experiment 2. Both of these effects were confirmed in the regression analysis presented in Section 5.6. In the remainder of this chapter, I outline the results on each cognitive measure in more detail.

	canonical	innov. [s]	innov. [t^h-ɨl] - [t^h-e]	innov. in /c/	innov. in /s/	innov. [c^h]
NRT	-0.20 (1.00)	0.11 (1.00)	0.09 (1.00)	0.15 (1.00)	0.05 (1.00)	0.14 (1.00)
Stroop	0.51 (0.29)	-0.55 (0.14)	0.18 (1.00)	-0.29 (1.00)	-0.16 (1.00)	-0.01 (1.00)
AQ	-0.17 (1.00)	0.12 (1.00)	0.09 (1.00)	0.22 (1.00)	0.14 (1.00)	0.14 (1.00)

(a) Experiment 1

	canonical	innov. [s]	innov. [t^h-ɨl] - [t^h-e]	innov. in /c/	innov. in /s/	innov. [c^h]
NRT	0.01 (1.00)	0.09 (1.00)	0.25 (1.00)	0.21 (1.00)	-0.15 (1.00)	0.01 (1.00)
Stroop	0.47 (0.55)	-0.48 (0.51)	-0.27 (1.00)	-0.31 (1.00)	0.25 (1.00)	0.14 (1.00)
AQ	-0.27 (1.00)	0.14 (1.00)	-0.33 (1.00)	0.00 (1.00)	-0.09 (1.00)	0.01 (1.00)

(b) Experiment 2

Table 5.5: Pearson's correlation coefficients between cognitive measures and phonological measures. Numbers in parentheses indicate *p*-values adjusted for multiple testing. (Abbreviations: NRT = Nonword repetition test scores, Stroop = Stroop task scores, AQ = AQ scores, canonical = proportion of canonical responses, innov. [s] = proportion of innovative [s] responses, innov. [t^h-ɨl] - [t^h-e] = difference in the proportion of innovative responses between [t^h-ɨl] and [t^h-e], innov. in /c/ = proportion of innovative responses in /c/, innov. in /s/ = proportion of innovative responses in /s/, innov. [c^h] = proportion of innovative [c^h] responses)

First, individuals' nonword repetition test scores, presented in Fig. 5.22, were not significantly correlated with any of the twelve phonological measures. The correlation coefficients were relatively small, as can be seen in the first rows of the two subtables in Table 5.5.

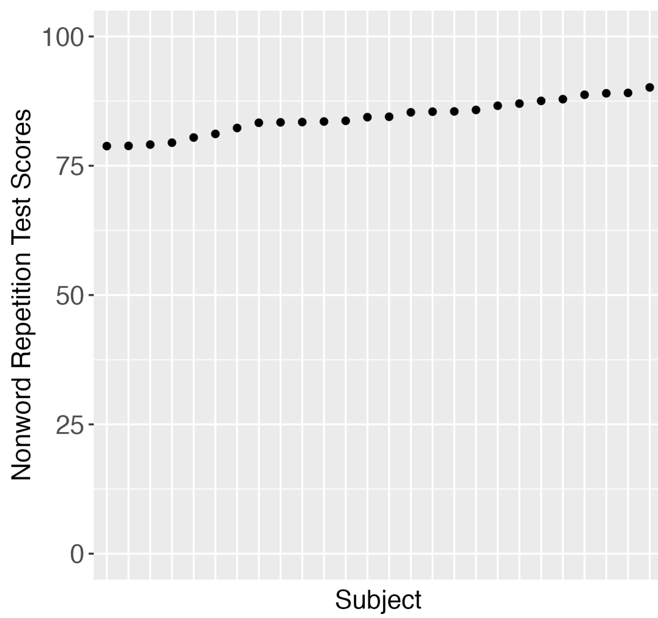

Figure 5.22: Individual participants' nonword repetition test scores

As for the Stroop task scores, I present the results based on the differences in RT, rather than the differences in accuracy, because RT offers a wider range of individual differences. Individuals' accuracy ranged from 97.5% to 100% in the congruent condition, and from 92.5% to 100% in the incongruent condition. It seems that the participants' performance was decent even in the incongruent condition, resulting in a relatively small degree of individual variation. Among the 23 out of 26 participants who had a higher accuracy in the congruent condition (as expected), the difference in accuracy was as low as 0.8% and as high as 7.5%. One participant performed equally well in the two conditions, and two participants had a higher accuracy in the incongruent condition by 1.7%.

Individuals' Stroop task scores, as indicated by the difference in the RT between congruent and incongruent stimuli, were not significantly correlated with any of the phonological measures. However, it is worth noting that the Stroop task scores were positively correlated with the proportion of canonical responses (Fig. 5.23) and negatively with the proportion of non-canonical [s] responses (Fig. 5.24) in both experiments; although the positive coefficients were non-significant (the p-values became greater than 0.05 once adjusted to correct for multiple testing), likely due to the small sample size, they were quite large and consistently observed across the two experiments. The findings indicate that better inhibitory control in individuals is associated with *less* frequent canonical forms in their speech, likely due to more innovative [s] responses. This trend is confirmed by a statistical analysis using a regression model, presented in Section 5.6.

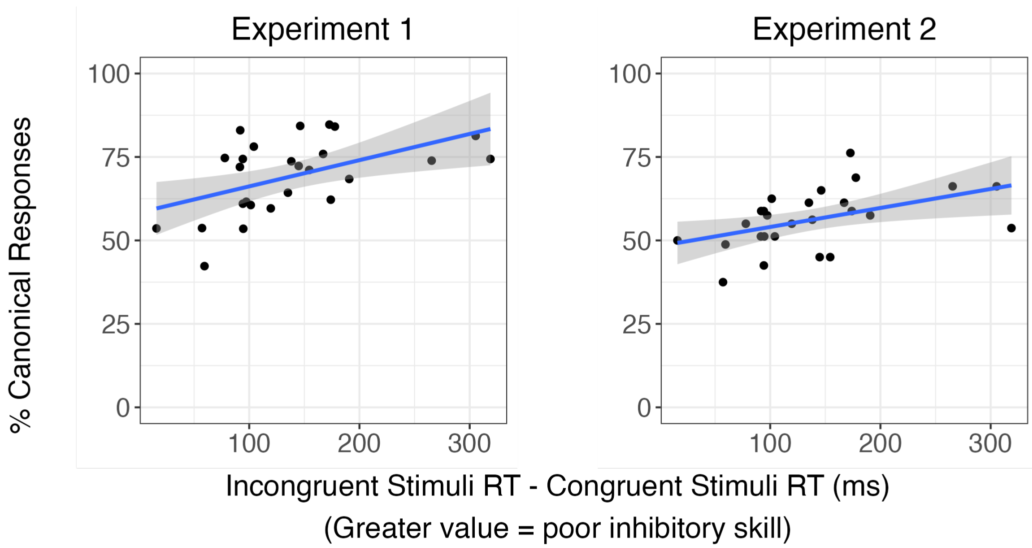

Figure 5.23: Relationship between inhibitory skills, as measured by the difference in the RT between congruent and incongruent stimuli in the Stroop task, and the proportion of canonical responses

Finally, I investigated the relationship between the phonological measures and the total AQ scores, as well as each subscale of AQ scores. No significant correlations were found between individuals' total AQ scores (based on either continuous or binary scoring) and any of the phonological measures, and no discernible trend was observed. For readers' reference, I present individuals' AQ scores in Fig. 5.25, both binary and continuous.

However, once the AQ questionnaire scores were broken down into five subscales, a clear trend of a positive correlation emerged between attention switching scores, assessed using binary scoring, and the proportion of innovative responses for /c/-final stems in the multiple-choice questions. (A positive correlation was also found between attention switching subscale

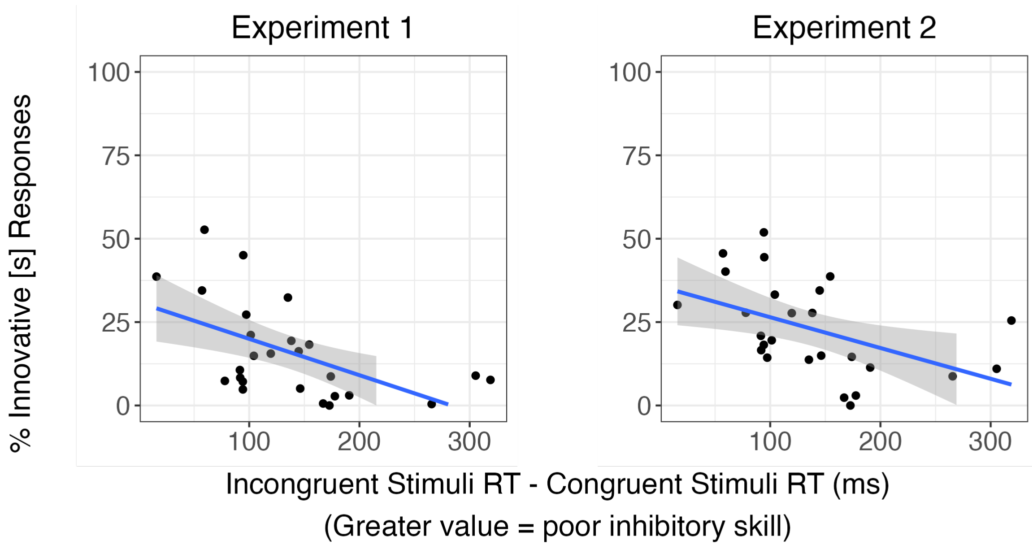

Figure 5.24: Relationship between inhibitory skills, as measured by the difference in the RT between congruent and incongruent stimuli in the Stroop task, and the proportion of innovative [s] responses

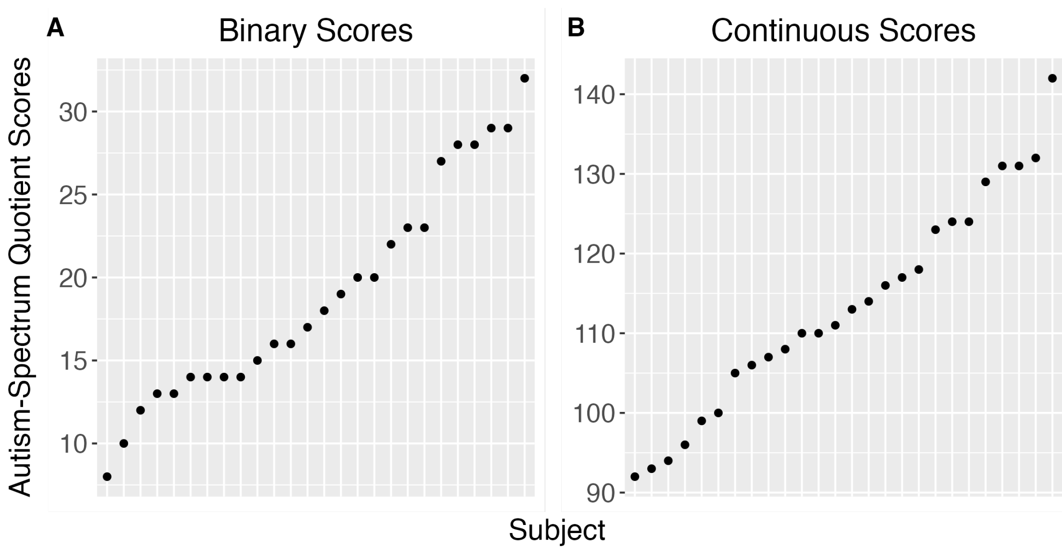

Figure 5.25: Individual participants' Autism-Spectrum Quotient scores based on (A) binary scoring and (B) continuous scoring

assessed based on continuous scoring and the same phonological measure, although it did not reach significance; r=0.36, p=0.069.) As shown in Fig. 5.26A, individuals with higher attention switching scores, indicating poor attention switching and strong focus of attention[4], were more likely to choose an innovative pronunciation for /c/ (r=0.44, unadjusted p<0.05). Then, what are these innovative pronunciations they chose for /c/? Because the majority of the non-[c] responses in /c/ were [s] (see Fig. 5.1), the innovative variant chosen by these individuals must have been [s]. This was indeed the case, as can be seen in Fig. 5.26B; those individuals with poor attention switching or better focusing ability produced innovative [s] responses for /c/ more frequently (r=0.43, unadjusted p<0.05).

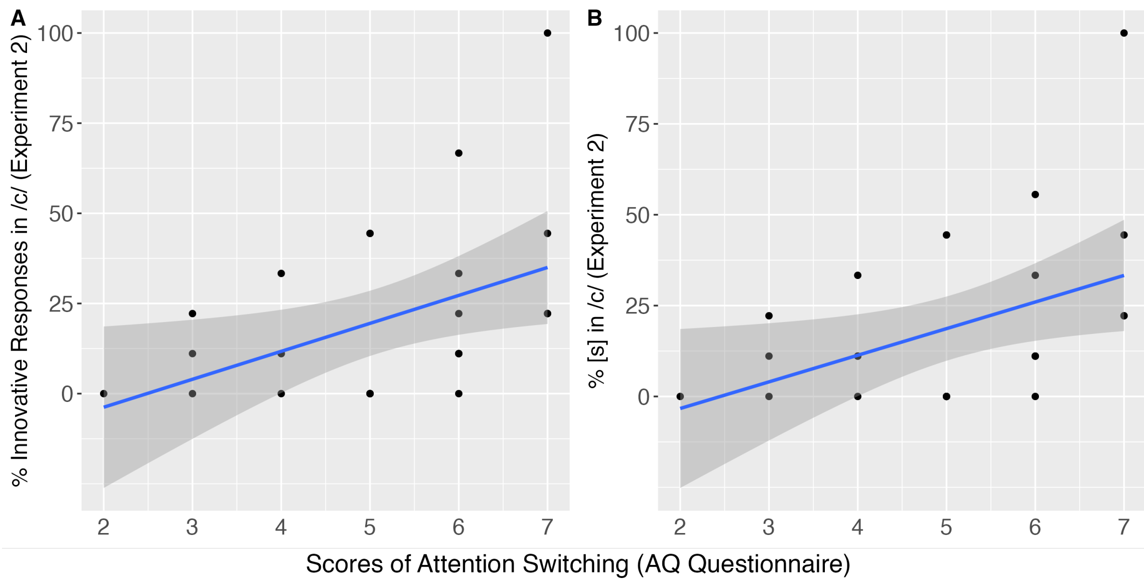

Figure 5.26: Relationship between individuals' responses in /c/ in Experiment 2 and attention switching scores within the AQ scores based on binary scoring

At first glance, it seems counterintuitive that individuals with strong focus more frequently chose innovative forms. This observation may be, however, consistent with the finding that better inhibitory skills are associated with fewer canonical forms, as reported above. It is plausible that strong focus and better inhibitory control are related characteristics, both of which are associated with more frequent innovative forms, and specifically, innovative [s] responses.[5]

[4]Recall that higher AQ scores indicate more autistic traits, in this case, struggling with switching attention.

[5]This hypothesis was not confirmed in the current data, as the correlation between individuals' attention switching scores within the AQ and the Stroop task scores was not significant (r=-0.13, p=0.51). This suggests that the effects

5.6 Statistical testing

As with Experiment 2, I ran a Bayesian logistic regression model to statistically examine the factors that affect the speakers' canonical responses in the multiple choice questions in Experiment 2. The dependent variable was the binary outcome of canonical vs. non-canonical (reference level), in which I categorized the responses ([s], [t^h], [c^h], [c] and [t]) into canonical pronunciation and non-canonical one.

The independent variables are presented in (14), grouped into either categorical or continuous variables. All categorical variables were dummy-coded, and the underlined level is the baseline. All of the categorical variables as well as a subset of the continuous variables listed in (14) were also included in the model reported in Section 4.4 for Experiment 1. Among the continuous variables, STEM FAMILIARITY and NORM IS IMPORTANT were obtained from tasks included only in Experiment 1, not Experiment 2. However, they were deemed relevant to participants' responses in Experiment 2 as well, and thus included in the model reported below. The assumption is that participants' perceptions of stem familiarity and the importance of pronouncing correctly have not changed substantially in the time between Experiment 1 and Experiment 2.

The variables NRT SCORES, STROOP TASK SCORES and the five AQ-related measures (AQ COMMUNICATION, AQ SOCIAL SKILL, AQ ATTENTION SWITCHING, AQ ATTENTION TO DETAIL, AQ IMAGINATION) are included to assess the effect of cognitive traits on the proportion of canonical responses. These measures were not included in the model reported in Section 4.4 for Experiment 1. Instead of using a single variable for the total AQ score, I used scores from each subdomain of the AQ questionnaire to assess the individual contributions of each area.

(14) *Independent variables in the logistic regression analysis (Experiment 2)*

- Categorical variables (underlined = baseline level)

of attention switching and inhibitory skills are distinct; indeed, as described above, attention switching appears to affect only /c/-final stems, and primarily in Experiment 2, whereas inhibitory skills are associated with innovative [s] responses across /t^h/-, /c^h/-, and /c/-final stems, in both experiments.

  - FINAL C: /s/, /tʰ/, /cʰ/, /c/
  - SUFFIX: /-e/, /-i/, /-ɨl/
  - FINAL C × SUFFIX: interaction between stem-final obstruent and suffix
  - RELATIVE FREQ: stem is more frequent (stem > word), stem and word have the same frequency (stem = word), word is more frequent (stem < word)
  - GENDER: female, male

- Continuous variables
  - LOG STEM FREQ: the log of the stem's frequency (in its standalone form)
  - LOG WORD FREQ: the log of the word's frequency
  - LOG STEM FREQ × LOG WORD FREQ: interaction between the log of the stem's frequency and the log of the word's frequency
  - STEM FAMILIARITY: response for "How often do you hear and use this word?" on a Likert scale of 1 (Rarely) to 7 (Very often)
  - AGE: participant's age
  - NORM IS IMPORTANT: response for "How important do you think it is to speak with correct pronunciation?" on a Likert scale of 1 (Not important at all) to 7 (Very important)
  - NRT SCORE: accuracy in the nonword repetition test
  - STROOP TASK SCORE: the difference in the reaction time between incongruent stimuli and congruent stimuli in the Stroop task
  - AQ COMMUNICATION: scores in the *communication* subdomain of the AQ questionnaire
  - AQ SOCIAL SKILL: scores in the *social skill* subdomain of the AQ questionnaire
  - AQ ATTENTION SWITCHING: scores in the *attention switching* subdomain of the AQ questionnaire

- AQ ATTENTION TO DETAIL: scores in the *attention to detail* subdomain of the AQ questionnaire
- AQ IMAGINATION: scores in the *imagination* subdomain of the AQ questionnaire

In addition to these fixed factors, random intercepts for participant and stem were also included.

As with Experiment 1, the model was fit using the *brms* package (Bürkner, 2017, 2018, 2021) in R (R Core Team, 2018). In each of four chains, 10,000 samples were drawn from the posterior distribution over parameter values, with the first 1,000 discarded for warm-up. A weakly informative Student-*t* prior with 5 degrees of freedom was used for the intercept and coefficients (Ghosh et al., 2018). There were no divergent transitions after warm-up. All parameters had an R-hat value of 1, indicating that the model had converged.

Table 5.6 provides coefficient estimates along with their upper and lower bounds of the 95% Credible Interval (CrI). In Table 5.6, parameters whose CrI does not include zero are boldfaced. When the CrI includes zero, I also provide p-direction, a value between 50% and 100%. As mentioned in Section 4.4, a p-direction value close to 50% indicates little evidence for an effect of either direction, and a value close to 100% indicates strong evidence for an effect in that direction.

First, the coefficient for FINAL C (s) was positive and the CrI did not include zero, suggesting that the proportion of canonical responses for /s/-final stems was credibly higher than the reference level /t^h/-final stems, when the suffix is /-e/. The coefficient for FINAL C (c^h) was negative, with the CrI including zero and a p-direction of 94.48%, indicating that the probability of choosing a canonical form was probably lower in /c^h/-final stems than /t^h/-final stems, when the suffix is /-e/. This result is in the opposite direction of Experiment 1, in which the probability of choosing a canonical form was probably higher in /c^h/-final stems than in /t^h/-final stems. The effect of FINAL C (c) was not credible, indicating that /c/-final stems were not different

Parameter	Estimate	Est. Error	95% CrI lower limit	95% CrI upper limit	p-direction
Intercept	0.55	0.48	-0.37	1.49	87.68%
FINAL C (s)	5.04	1.85	1.86	9.12	
FINAL C (c^h)	-0.96	0.61	-2.20	0.23	94.48%
FINAL C (c)	0.15	0.88	-1.58	1.87	57.43%
SUFFIX (i)	0.49	0.24	0.01	0.96	
SUFFIX (ɨl)	-2.41	0.27	-2.95	-1.89	
FINAL C (s) :SUFFIX (i)	1.38	2.58	-2.81	7.39	69.82%
FINAL C (c^h) : SUFFIX (i)	-0.08	0.32	-0.71	0.55	59.56%
FINAL C (c) : SUFFIX (i)	-0.22	0.53	-1.27	0.81	66.06%
FINAL C (s) :SUFFIX (ɨl)	0.31	1.71	-3.37	3.39	59.63%
FINAL C (c^h) : SUFFIX (ɨl)	3.01	0.39	2.26	3.78	
FINAL C (c) : SUFFIX (ɨl)	3.13	0.58	2.00	4.26	
LOG STEM FREQ	0.63	0.35	-0.03	1.33	96.87%
LOG WORD FREQ	-0.28	0.27	-0.81	0.24	85.67%
LOG STEM FREQ : LOG WORD FREQ	0.14	0.34	-0.54	0.80	66.33%
RELATIVE FREQ (stem = word)	0.75	0.88	-0.96	2.49	80.91%
RELATIVE FREQ (stem < word)	1.21	0.37	0.49	1.94	
STEM FAMILIARITY	0.08	0.11	-0.15	0.30	75.13%
AGE	0.05	0.20	-0.34	0.44	59.77%
GENDER (male)	0.77	0.41	-0.06	1.58	96.63%
NORM IS IMPORTANT	0.45	0.18	0.08	0.81	
NRT SCORES	0.07	0.15	-0.23	0.37	67.92%
STROOP TASK SCORES	0.28	0.16	-0.05	0.60	95.46%
AQ COMMUNICATION	-0.38	0.32	-1.01	0.24	89.34%
AQ SOCIAL SKILL	0.37	0.33	-0.26	1.01	88.01%
AQ ATTENTION SWITCHING	-0.35	0.20	-0.75	0.05	95.91%
AQ ATTENTION TO DETAIL	-0.04	0.18	-0.39	0.31	58.86%
AQ IMAGINATION	-0.01	0.19	-0.38	0.36	51.95%

Table 5.6: Results of the Bayesian logistic regression analysis of Experiment 2

from /t^h/-final stems in their proportion of canonical responses. SUFFIX (i) credibly predicted the outcome with a positive coefficient, which is expected due to the obligatory palatalization process in /t^h-i/. SUFFIX (ɨl) credibly predicted the outcome with a negative coefficient, confirming the suffix effect in which [t^h] (or, the canonical form of /t^h/ in the present context) is preferred before /-e/ than before /-ɨl/.

As for the interaction terms between FINAL C and SUFFIX, FINAL C (c^h) : SUFFIX (ɨl) exhibited a credible effect with a positive coefficient. This means that while the probability of choosing a canonical form was probably lower in /c^h/-final stems than /t^h/-final stems when the suffix is /-e/, the probability was higher in /c^h/ than /t^h/ when the suffix is /-ɨl/, as can be seen in Fig. 5.27. This again confirms the suffix effect reported in the literature (e.g. Jun, 2010; Kang, 2007). Additionally, FINAL C (c) : SUFFIX (ɨl) also had a credible effect with a positive coefficient. Given that /c/-final stems and /t^h/-final stems did not differ in their proportion of canonical responses in the /-e/ suffix, this credible interaction suggests that the likelihood of being chosen in the canonical form is higher in /c/-stems compared to /t^h/-stems when the suffix is /-ɨl/. This is confirmed in Fig. 5.27.

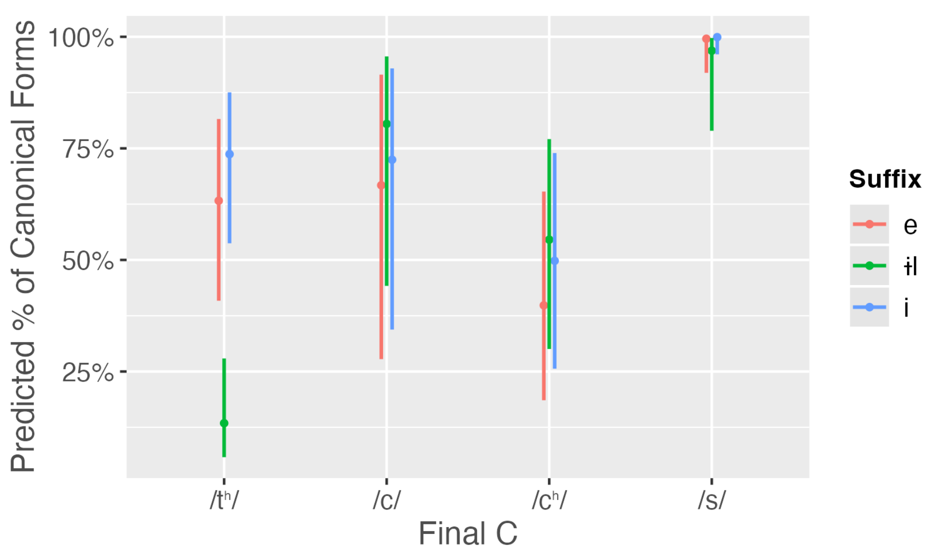

Figure 5.27: Interaction of stem-final obstruent and suffix in Experiment 2

Next, a number of frequency measures credibly predicted the probability of choosing a canonical form. First, LOG STEM FREQ had a positive coefficient with a p-direction of 96.87%, suggesting that higher stem frequency is associated with higher likelihood of choosing a canonical form. LOG WORD FREQ had a negative coefficient, with the CrI including zero. The p-direction was 85.67%, suggesting moderate evidence that higher word frequency is probably associated with lower likelihood of choosing a canonical form. This is contrary to the expectation outlined in Section 4.4 that a word would be more likely to be realized with a canonical form when the stem frequency is low and the word frequency is high. The interaction between LOG STEM FREQ and LOG WORD FREQ did not reliably predict the outcome. It was also found that RELATIVE FREQ (stem < word) credibly predicted the proportion of canonical forms, with positive coefficients and the CrI not including zero. This finding suggests when word frequency is higher then the stem frequency, the participants were more likely to choose canonical forms. Individuals' rating of stem familiarity did not reliably predict the probability of choosing a canonical form.

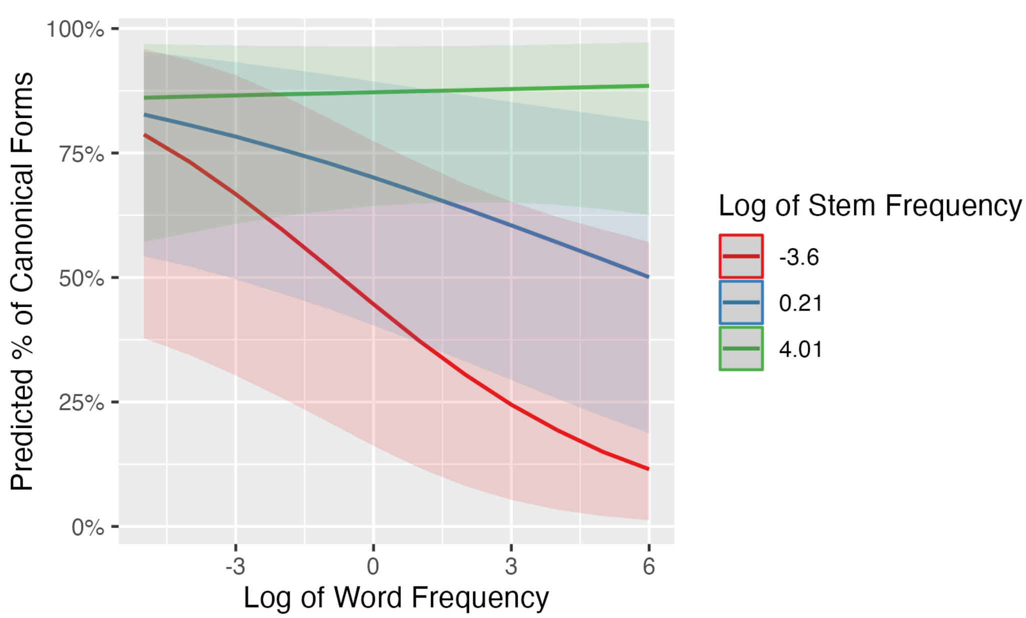

Figure 5.28: Interaction of log of word frequency and log of stem frequency in Experiment 2

Finally, a number of individual-level measures credibly predicted the probability of choosing a canonical form. As with Experiment 1, there was no evidence that AGE explained the individual's probability of choosing a canonical form. Unlike Experiment 1, however, GENDER credibly predicted the probability of choosing a canonical form with a positive coefficient and a p-direction of 96.63%. Specifically, male participants were more likely to choose a canonical form than female participants. Another divergence from findings of Experiment 1 is that NORM IS IMPORTANT, individuals' perceived importance of pronouncing correctly, also predicted the proportion of canonical forms with a positive coefficient with its CrI not including zero. Among the cognitive measures, NRT SCORES did not reliably explain the likelihood of choosing a canonical form. STROOP TASK SCORES had a positive coefficient, with its CrI including zero and a p-direction of 95.46%. This finding confirms the observation in Section 5.5 that better inhibitory control is associated with lower likelihood of choosing a canonical form. (Recall that a higher value for this variable indicates poorer inhibitory control.) Among the scores of each subdomain of the AQ questionnaire, only AQ ATTENTION SWITCHING reliably predicted the proportion of canonical forms. It had a negative coefficient, a CrI that includes zero, and a p-direction of 95.91%. Higher AQ ATTENTION SWITCHING scores, indicating *poor* attention switching, or *better* focus, were associated with *fewer* canonical forms. This confirms the correlational analyses provided in Section 5.5.

5.7 Discussion

As stated at the beginning of this chapter, the main goals of Experiment 2 were (i) to examine the consistency of individuals' responses across different tasks and over time, (ii) to explore how speakers produce obstruent variants in nonce words, and (iii) to perform a preliminary analysis of the relationship between the speakers' variant choices and their cognitive traits. Each of these points are addressed in order.

5.7.1 Systematicity in individual differences

Most importantly, the participants who completed both Experiment 1 and Experiment 2 were consistent in their responses. Those who frequently produced canonical forms (Experiment 1) were more likely to choose canonical forms as their pronunciation in multiple choice questions (Experiment 2). Further, speaker consistency was also observed in the more specific phonological environments. Notably, those who had a higher proportion of [s] responses for /t^h/-, /c^h/- or /c/-final stems in Experiment 1 were more likely to do so in Experiment 2 as well. Note that the presence of such non-canonical [s] is well documented in the literature (Jun, 2010; Jun & Lee, 2007; Albright, 2008; Hayes, 1998). The findings suggest that speaker variation is systematic; if it was mere random noise, we would not have observed individual differences in the well-known statistical tendency or individuals' self-consistent behavior across the two experiments.

Before inferring that the observed individual differences reflect the speakers' true linguistic knowledge, I consider a number of possibilities that they are artifacts. First, is it possible that the participants were forced/pressured to behave similarly in the two experiments? They were not told explicitly about what the experiments were testing, or that the two experiments are related. Some of them, especially those who participated in Experiment 2 after 10 days of completing Experiment 1, might have realized that the study was longitudinal. It is unlikely, however, that these participants remembered their responses in Experiment 1 at the time of completing Experiment 2, since they had to produce 104 different target words (30 different stems) in Experiment 1. Even within Experiment 1, individuals produced different variants for the same words in block 1 and block 2. The observation of within-speaker variation across the two blocks gives little credence to the possibility that participants memorized their previous responses and used it in a later task.

Another possibility is that the observed individual differences more reflect variability in how individuals perform experimental tasks than they reflect variability in linguistic knowledge. For instance, some participants may be more familiar or comfortable with being presented with auditory stimuli, and thus provide responses that are closer to their spontaneous speech. I consider this possibility unlikely. The two experiments employed different tasks, although they both aimed to access the participants' pronunciation for the stem-final obstruents. Experiment 1 asked them to produce an inflected form without any written presentation of the stem or the inflected form, while Experiment 2 presented five written forms of pronunciations and asked them to choose their own pronunciation for the target word, also presented in written form. Moreover, to minimize the effects of individuals' familiarity with experiment procedures or how successfully they understood the tasks, I allowed participants to take as much time as they wanted to complete the tasks. Furthermore, as mentioned in Section 4.5, participants were given the flexibility to choose their preferred time and location for participating in the experiment to ensure convenience and comfort. Therefore, it is hard to think that the observed across-speaker variability is primarily due to task performance differences rather than genuine differences in their linguistic knowledge.

Now that we can plausibly regard the individual variation as stemming from differences in linguistic knowledge, it is worth considering the extent to which this knowledge reflects differences in the lexicon versus grammar. This question can be addressed by investigating whether the process is productive or not, asking participants to inflect a nonce word. Before discussing the implications of the nonce word test results on the sources of differences in linguistic knowledge, I review two observations from Experiment 1 that bear on the possible differences in the individuals' mental lexicon.

First is individual participants' familiarity rating of stems in Experiment 1. It is possible that the more familiar the stem is to an individual, the more likely they are to recall its standard or lexical form. In the statistical testing presented in Section 4.4, we found little evidence that stem familiarity rating predicted the probability of being produced with a canonical form. A caveat is that there are several measures of word and stem frequency included in the model, which could obscure the effect of stem familiarity. For instance, it is possible that individuals' sensitivity to the relative frequency between word and stem (that is, whether the stem is used in the suffixed form more frequently than in the unsuffixed form) might explain individual differences in the probability of producing a canonical form. This can be tested by including by-subject random slope for relative frequency between word and stem in the model.

Another task of Experiment 1 that may tap into speakers' lexical knowledge is the spelling test, in which the speakers were auditorily presented with the unsuffixed forms of the target stems and spelled them. Recall that the Korean orthography represents lexical or underlying forms of the stem-final obstruents. I explored the possibility that the target stems are represented with different forms in individuals' lexicon; for instance, some speakers might (incorrectly) remember '/juc^h/' 'a type of traditional game', an infrequent stem, as /jus/, (and their spelling would reflect this,) providing more [s] responses for this stem. Scoring of the spelling test, presented in (15), revealed that there was only a small degree of variation in individuals' accuracy. 22 out of 35 participants in Experiment 1 spelled all stems correctly, and 8 participants spelled only one stem incorrectly. This variability is far smaller than the individual differences in the variant frequencies observed in the experiments.

(15) Accuracy in the spelling test

Accuracy	**Number of participants**
100% (30/30)	22 (62.9%)
96.7% (29/30)	8 (22.9%)
93.3% (28/30)	2 (5.7%)
90.0% (27/30)	2 (5.7%)
86.7% (26/30)	1 (2.9%)

Thus, when accuracy in the spelling test was used as a proxy for underlying representations of the final obstruents, we found no evidence that individual variability in lexical representations can explain the variation in pronunciation.

5.7.2 Productivity

A nonce word test was conducted to offer more insights into the source of individual differences. If we observe similar individual differences in the nonce word test, for instance, those who produced a high proportion of [s] in Experiment 1 also produced [s] frequently in the nonce word test, we could argue that the inter-speaker variation stems from differences in the speakers' probabilistic grammar. The degree of inter-speaker variation observed in the real word production test in Experiment 1 was not reproduced in the nonce word production test in Experiment 2, as [s] was the predominant response in the majority of the participants' nonce word production. This pattern in the wug test does not match the lexical frequencies of stem-final obstruents. Although /s/ is the most frequent among the stem-final obstruents in the lexicon, the probability of producing an [s] for the nonce words is disproportionately higher than the frequency of /s/ in the lexicon or the proportion of [s] variants in the real words production.

To help understand the reason that participants in the current study did not frequency-match, it is useful to review a previous wug test in Jun (2010), which conducted a well-formedness judgment task using a set of noun stems in Korean. In Korean, certain bound stems combine with the verb stem *ha* 'to do' to form compound verbs (e.g. /pisɨs-ha-ta/ 'similar-do-DECLARATIVE'). Jun (2010) employed topicalized forms of these verbs, e.g. /pisɨs-ɨn ha-ta/ 'similar-TOPIC do-DECLARATIVE', asking native Korean participants to rate five different realizations of such words, e.g. [pisɨ**s**-ɨn], [pisɨ**t^h**-ɨn], [pisɨ**c^h**-ɨn], [pisɨ**c**-ɨn], and [pisɨ**t**-ɨn]. This study can be considered a type of wug test, because the topicalized forms of these stems sound perfectly acceptable to the native speakers yet they are rarely used in actual speech. The relative ratings of each of these five variants matched their lexical frequencies, that is, [s] was rated the highest, [c] and [t] were rated the lowest, and [t^h] and [c^h] received an intermediate rating.

This frequency matching behavior shown in Jun (2010) was not replicated in the present study. I believe that the discrepancy can be attributed to the design of the two studies. Whereas Jun (2010) provided the participants with the possible pronunciations of the suffixed forms in written form (similar to the multiple choice questions of the present study), the participants in the current wug test were not provided with a list of possible pronunciations. When asked to produce the suffixed form only once, the participants seemed to have skewed towards their most preferred form, which gave rise to predominant [s] responses when the responses were aggregated. Additionally, despite efforts to create nonce stems that sound well-formed as Korean stems, participants might have perceived the stems as foreign-sounding. This could increase the likelihood of producing [s] for the stem-final obstruent in the suffixed form, as Koreans typically inflect loanwords ending with a coronal obstruent with [s]; for example, final coronal obstruents in English loanwords like *David* and *cut* are pronounced as [s] (e.g. *David*-NOM → [teipis-i], *cut*-ACC → [k^hʌs-ɨl]; Jun & Lee, 2007). Further, a limitation in the experimental design in which the number of target stems (N=25; ending in [t], potentially produced as [s] when suffixed) exceeded the number of filler stems (N=15; ending in a sonorant) might have amplified the number of [s] responses. Producing a few [s] variants in the beginning of the test might have primed their later responses, especially in individuals who tend not to switch their pronunciations.

Despite the rather small degree of variance in the individuals' responses in the nonce words, a tendency was observed in which those who provided fewer [s] responses in the real words also produced relatively fewer [s] responses in the novel words. This is evidence suggesting that at least some amount of individual differences can be traced to differences in the grammar, rather than lexicon. This argument can be corroborated by the aforementioned wug test (Jun, 2010), in which speakers' well-formedness ratings closely matched lexical statistics.

Finally, I would like to discuss the finding that among the minor variants the speakers produced for the nonce words, namely [t^h], [c^h] and [c], the most frequent variant was the [c^h]. This is in line with the results of the production test in Experiment 1 in that [c^h] was the most frequent non-[s] variant in the real words as well. Preference for [c^h] is also attested by the finding that when /s/ is produced with a non-canonical variant, it is always [c^h]. This is consistent with previous studies showing that /c^h/ is the second most frequent coronal obstruent in the lexicon, following /s/, and that [c^h] is the second most preferred variant for both real and wug stems (Jun, 2010).

5.7.3 Relationship between cognition and phonology

In addition to investigating the speakers' pronunciation of the obstruents in the existing and novel words, Experiment 2 also assessed individuals' cognitive traits using three tasks: nonword repetition test to assess phonological working memory, Stroop task as a proxy for inhibitory skills, and AQ questionnaire to measure autistic traits. I first discuss the cognitive measures that were found to reliably predict the proportion of canonical forms, and then the ones not closely linked to the individuals' choice of canonical variants.

First, individuals' Stroop task performance credibly predicted their proportion of canonical variants. The smaller RT difference between congruent and incongruent stimuli, which indexes better inhibitory skill, was associated with less frequent canonical responses. At first glance, this result seems counterintuitive, since we would plausibly expect an individual with good inhibitory skills to inhibit all the non-standard variants and try to produce or choose the canonical forms. In fact, we found otherwise, i.e. those with better inhibitory control preferred *non*-canonical forms. Why is this the case?

This seemingly confusing pattern can probably be explained by referring to the finding that better inhibitory control is associated with a higher proportion of innovative [s] responses. Assuming that speakers prefer [s] because /s/ is the most common among the stem-final obstruents in the lexicon, I speculate that individuals with strong inhibitory skills are more acutely aware of this lexical statistic or can more actively utilize this lexical pattern. It is possible that, during the experiment, these individuals were better at suppressing the non-[s] variants (possibly because they are aware that non-[s] variants are infrequently used in the language). Alternatively, during language acquisition, their superior inhibitory skills may have enabled them to internalize lexical statistics more effectively. It remains unclear at what point the influence of inhibitory control takes effect.

Although we are not certain about when and how exactly the effect of inhibitory skills influences individuals' phonological behavior, what it *does* suggest is that the attested inter-speaker variation does not merely reflect differences in their propensity to conform to normative pronunciation rules. If varying degrees of conservatism primarily modulate individuals' frequency of producing or choosing canonical forms, we would expect to observe more canonical forms among those with better inhibition: the better the inhibitory skills, the more successfully an individual can suppress non-standard forms. What we find is contrary, rejecting conformativity to prescriptive rules as the primary source of individual variation.

Among the subscales of the AQ scores, attention switching had a credible effect on the proportion of canonical forms. Specifically, individuals with poor attention switching and strong focus of attention were more likely to choose innovative forms for /c/ in Experiment 2. This finding is counterintuitive, as one might anticipate that individuals with better focusing abilities would provide more canonical responses, potentially due to greater awareness of the experiment's purpose or an increased effort to choose the 'correct' pronunciation of the word. In fact, those individuals provided more non-canonical responses, specifically innovative [s] responses for /c/. As with the effect of inhibitory skills, it is not certain how exactly attention switching influences individuals' responses. It is possible that strong focus leads one to acquire lexical patterns better than others (which happens prior to participating in experiments), and/or that it helps them to focus on the task and better access their lexical knowledge during the experiment. Additionally, individuals with high attention switching scores may not frequently alternate between pronunciations, resulting in consecutive [s] responses.

Note that attention switching and the frequency of canonical responses were correlated only in the multiple choice questions (Experiment 2), not in the production task (Experiment 1), as outlined in Section 5.5. This difference may be due to the nature of the multiple choice questions, which make participants more conscious of their responses (pronunciations) and may provide more opportunities for attention switching or focus to affect the responses.

Having discussed the factors that reliably predict the proportion of canonical forms, I now turn to the null results. Cognitive measures other than inhibitory skills and one subscale of the AQ scores did not affect the probability of choosing a canonical form. The null results themselves are not surprising, as research on the relationship between cognition and language is still in their beginning stage, and we are far from fully understanding the link between the two. In fact, a few recent studies cast doubt on the relationship between cognitive or personality traits and linguistic behavior (Hall-Lew et al., 2021; Wade, 2022). These null results merit further discussion, as a link between language and cognition may exist but could be obscured by methodological limitations.

First, the sample size was relatively small; only 26 participants were included in the calculation of correlation coefficients. This is smaller than the sample sizes typically recommended for correlation studies(Cohen, 1992; Wilson Van Voorhis & Morgan, 2007). Some of the correlations might have reached statistical significance if a larger number of participants were recruited, such as the one between Stroop task performance and the probability of producing non-canonical [s], which exhibited a clear trend. (This relationship was confirmed in the regression analysis presented in Section 5.6.)

Second, it is possible that the cognitive tasks employed in the present study did not accurately measure the constructs that they intended to measure. For instance, in the Stroop task employed in the present study, the participants were asked to identify the color of the text by pressing a key, rather than saying it aloud as in the original version of the task (Stroop, 1935). The individuals' accuracy may have been affected by their ability to control the keys, rather than only measuring the construct of interest, namely the ability to inhibit irrelevant dimension of the stimuli. It is recommended, therefore, that multiple tasks should be employed to measure a single construct (James et al., 2018). For example, to measure working memory span, one should carry out reading, operation and listening span tests. Using multiple measures to assess a single construct will also allow for statistical models, such as structural equation modeling, that can extract what is common across those tasks in order to remove task effects and measure a “purer” underlying construct (latent variable). Since our understanding of how individual cognitive differences relate to linguistic behaviors is still in its infancy, and considering we have only a preliminary grasp of which traits to measure, employing a variety of tests is necessary. This aligns with the earlier point on the necessity of increasing sample sizes, as testing multiple measures requires adequate statistical power to avoid false positives.

Putting the methodological considerations aside, it is also possible that inter-speaker variation in phonology really cannot be predicted by differences in individuals’ cognitive traits. As mentioned in Chapter 1, while many studies attribute the across-speaker variation in phonetics and phonology to individual differences in these domain-general cognitive styles or personality traits, as reviewed in (Yu & Zellou, 2019), recent studies suggest that not all the variance can be explained by these differences (Hall-Lew et al., 2021), or that very little of it can (Wade, 2022). It is possible, therefore, that the link between domain-general cognitive skills and language ability is indeed absent in the present study, suggesting that language may not extensively rely on domain-general cognitive functions. It is also plausible that numerous studies investigated such a link and found null results, which may have remained unpublished. Therefore, it remains uncertain whether inter-speaker variation in phonology can be reliably predicted by differences in individuals’ cognitive traits.

The present study only reports preliminary findings on the relationship between cognition and language, and more research is needed to elucidate to what extent language draws on domain-general systems and to what extent it is modular (Fodor, 1983). Importantly, future research needs to advance our methods for testing how variation in phonology can be explained by differences in cognition, by recruiting a larger number of participants and a larger set of behavioral tests. Some may critique the current analyses as being only exploratory (some may even consider them as fishing expedition), correlating various cognitive measures with multiple linguistic measures in the hope of finding some significant correlation. However, given our limited understanding of whether individuals’ phonological patterns can be explained by domain-general cognitive traits, such exploratory analyses are necessary. These initial investigations allow for more targeted future research, ultimately enhancing our understanding of the relationship between cognition and language.

CHAPTER 6

Modeling the inter-speaker variation

6.1 Introduction

In this section, I model the experimental data of individual speakers. I fit a Maximum Entropy (MaxEnt) grammar to the corpus frequencies, which are assumed to approximate the language input the learners receive. I first establish the constraints needed to explain the speakers' responses, and fit the grammar to the learning data multiple times, each representing an individual speaker's learning. The input data is consistent across all learning simulations, but I use different bias terms for certain constraints across different iterations of learning. (I use the term 'bias' because this parameter, which I vary to produce variability in the output, is typically used to encode some prior learning biases that speakers bring to learning; however, as explained in the remainder of this chapter, the parameter in the present study does not strictly represent the prior learning biases that individuals possess before language learning.) This models a scenario where individual differences arise because speakers have differently biased grammars, even though they receive the same language input. In Section 6.6, I consider the potential influence of input differences on the observed variation across speakers and stipulate that differences in input cannot account for the variability.

Specifically, I aim to model the three speakers (S1, S2, and S3) identified as the closest to the centroid of each cluster from Experiment 1, as determined by k-means clustering (Fig. 4.13). I chose to model individual differences observed in the production study in Experiment 1 rather than those of multiple choice questions in Experiment 2 for three reasons. First, the production task (volunteering an inflected form of a noun stem) is more similar to how speakers use their grammars in their real lives, compared to answering multiple choice questions (choosing the written pronunciation that they think best represents their own pronunciation among a number of options). Second, individual participants provided a greater number of responses in the production test (N=104) compared with the multiple choice questions (N=80). Third, because a greater number of participants were included in the k-means clustering in Experiment 1 (N=35) compared to Experiment 2 (N=26), the clustering results of Experiment 1 may reflect a more accurate characterization of speaker groups of Korean. Therefore, I give more credence to the response patterns of Experiment 1 than those of Experiment 2.

In modeling the speakers' responses, I excluded [t-e], [t-i] and [t-ɨl] variants, which accounted for only 0.05% of the responses. If we were to explain occurrences of [t] responses, we would have to include many more constraints than are established in Section 6.3 below, which will dramatically increase processing time.

6.2 Maximum Entropy grammar

The variable realization of the stem-final obstruents was analyzed within the framework of Maximum Entropy (MaxEnt) grammar (Smolensky, 1986; Goldwater & Johnson, 2003). In a variable phonological process, a MaxEnt grammar can model statistical tendencies observed in the variation by assigning numerical weights to Optimality-Theoretic (OT) constraints. The weights are calculated to maximize the probability of the observed frequencies in the data.

In the MaxEnt framework, one can specify a target or expected value (μ) for each constraint weight. A non-zero μ value indicates that the constraint is expected to have an effect on the output. One can also specify a penalty for constraint weights deviating from this default value, represented by σ. Smaller values of σ impose greater penalties for deviating from the μ's and require more data to shift the weights away from these target weights. When non-zero μ values are specified, I set the σ at 1, which can be considered a small value, so that the learned weights do not substantially deviate from μ values. This will ensure that different sets of μ's for individual speakers will indeed result in different output frequencies.

All learning simulations were conducted in *maxent.ot* package (Mayer et al., 2024) in R (R Core Team, 2018).

6.3 Constraints

I employ OT constraints to explain the production patterns observed in the experiment results, as outlined below.

6.3.1 ANTICORRESPONDENCE constraints

The stem-final obstruents considered in this study are involved in alternation between [t], the neutralized variant in the final position of the unsuffixed form, and one of [s], [t^h], [c^h], [c] or [t] when followed by a vowel-initial suffix. I employ a set of ANTICORRESPONDENCE constraints (Hayes, 1999), which actively require that morphemes alternate in particular ways. ANTICORRESPONDENCE constraints are rather ad hoc, but are useful in explaining complex and language-specific alternation patterns. ANTICORRESPONDENCE is defined as follows.

- ANTICORRESPONDENCE: If morpheme μ appears with shape X in a particular context C, it must appear with shape X′ in a distinct context C′ (Hayes, 1999).

In the current case, "context C" would correspond to the unsuffixed form, and "context C′ " to the suffixed form. "Shape X" will always be [t], while "shape X′ " can be any of [s], [t^h], [c^h], or [c]. For instance, ANTICORRESPONDENCE(t, _#, s, _e) enforces that [t] in the final position of an unsuffixed form alternates with [s] before the suffix [-e], penalizing the candidates in which the stem-final obstruent is not [s] before [-e], as illustrated in (16); ANTICORRESPONDENCE(t, _#, t^h, _ɨl) enforces an alternation between [t] in an unsuffixed form and [t^h] in an [-ɨl]-suffixed form. Note that ANTICORRESPONDENCE refers to surface forms, rather than to underlying forms, and thus are able to explain prevalence of some variant (e.g. [s]) in all contexts, unaffected by the type of underlying form of the obstruent.

(16) *An illustration of an* ANTICORRESPONDENCE *constraint*

/... t^h-e/	ANTICORR(t, _#, s, _e)
a. ... t^h-e	*
b. ... s-e	
c. ... c^h-e	*
d. ... c-e	*

A total of 11 ANTICORRESPONDENCE constraints were included in the constraint set, presented in (17). There are 12 logically possible constraints relevant for the current data, each requiring one of the four obstruent variants ([s], [t^h], [c^h], [c]) before one of the three suffixes (/-e/, /-i/, /-ɨl/). I did not include ANTICORRESPONDENCE(t, _#, t^h, _i) in the constraint set, however, because [t^h] is banned before the /-i/ suffix due to the palatalization process; if we had included ANTICORRESPONDENCE(t, _#, t^h, _i) in the learning simulation, it would have received a (near-)zero weight.

(17) ANTICORRESPONDENCE *constraints*

- ANTICORRESPONDENCE(t, _#, s, _e): [t] in the final position of an unsuffixed form should appear with [s] before the [-e] suffix.
- ANTICORRESPONDENCE(t, _#, s, _i): [t] in the final position of an unsuffixed form should appear with [s] before the [-i] suffix.
- ANTICORRESPONDENCE(t, _#, s, _ɨl): [t] in the final position of an unsuffixed form should appear with [s] before the [-ɨl] suffix.
- ANTICORRESPONDENCE(t, _#, t^h, _e): [t] in the final position of an unsuffixed form should appear with [t^h] before the [-e] suffix.
- ANTICORRESPONDENCE(t, _#, t^h, _ɨl): [t] in the final position of an unsuffixed form should appear with [t^h] before the [-ɨl] suffix.

- ANTICORRESPONDENCE(t, _#, c^h, _e): [t] in the final position of an unsuffixed form should appear with [c^h] before the [-e] suffix.
- ANTICORRESPONDENCE(t, _#, c^h, _i) : [t] in the final position of an unsuffixed form should appear with [c^h] before the [-i] suffix.
- ANTICORRESPONDENCE(t, _#, c^h, _ɨl): [t] in the final position of an unsuffixed form should appear with [c^h] before the [-ɨl] suffix.
- ANTICORRESPONDENCE(t, _#, c, _e): [t] in the final position of an unsuffixed form should appear with [c] before the [-e] suffix.
- ANTICORRESPONDENCE(t, _#, c, _i): [t] in the final position of an unsuffixed form should appear with [c] before the [-i] suffix.
- ANTICORRESPONDENCE(t, _#, c, _ɨl): [t] in the final position of an unsuffixed form should appear with [c] before the [-ɨl] suffix.

As a formal mechanism to explain the alternation, we could alternatively adopt stochastic rules that enforce the observed alternations based on Albright's Paradigm Learning Model (Albright, 2002a,b, 2005, 2008) by using minimal generalization learner (Albright & Hayes, 2003). For instance, Jun (2010) employed this model to explain the speakers' preference for certain stem-final variants in different suffix vowel contexts. I adopt the constraint-based model MaxEnt grammar, as opposed to rules. This model offers a mathematically well-defined framework for handling conflicting constraints and assigning probabilities to outcomes. Additionally, in the phonological variation literature, varying bias terms are often assigned to certain constraints to explain why speakers learn some patterns more readily than others (Hayes et al., 2009; Wilson, 2006). In modeling inter-speaker variation, we could make use of this apparatus to demonstrate that individuals learn phonological patterns differently because they have different degrees of biases to the same constraints and as a result, their production patterns match the lexicon to varying degrees.

6.3.2 Faithfulness constraints

Next, I posit faithfulness constraints that require an underlying form of a stem-final obstruent (the variant that is represented in the orthography) to be preserved in its corresponding output form. Specifically, a family of *MAP-IO constraints are proposed in (18) that ban an obstruent in the input from being mapped to another obstruent variant in the output. For instance, *MAP-IO($t^h \rightarrow s$) prevents /t^h/ in the input from surfacing as [s] in the output, as in (19). (Unlike *MAP constraints proposed in Zuraw, 2013, which compare surface forms, *MAP-IO is meant to regulate mappings between underlying and surface forms.)

(18) *MAP-IO *constraints*

- *MAP-IO($s \rightarrow t^h$): /s/ in the input must not correspond to [t^h] in the output.
- *MAP-IO($s \rightarrow c^h$): /s/ in the input must not correspond to [c^h] in the output.
- *MAP-IO($s \rightarrow c$): /s/ in the input must not correspond to [c] in the output.
- *MAP-IO($t^h \rightarrow s$): /t^h/ in the input must not correspond to [s] in the output.
- *MAP-IO($t^h \rightarrow c^h$): /t^h/ in the input must not correspond to [c^h] in the output.
- *MAP-IO($t^h \rightarrow c$): /t^h/ in the input must not correspond to [c] in the output.
- *MAP-IO($c^h \rightarrow s$): /c^h/ in the input must not correspond to [s] in the output.
- *MAP-IO($c^h \rightarrow t^h$): /c^h/ in the input must not correspond to [t^h] in the output.
- *MAP-IO($c^h \rightarrow c$): /c^h/ in the input must not correspond to [c] in the output.
- *MAP-IO($c \rightarrow s$): /c/ in the input must not correspond to [s] in the output.
- *MAP-IO($c \rightarrow t^h$): /c/ in the input must not correspond to [t^h] in the output.
- *MAP-IO($c \rightarrow c^h$): /c/ in the input must not correspond to [c^h] in the output.

(19) *An illustration of a* *MAP-IO *constraint*

/... t^h-e/	*MAP-IO($t^h \rightarrow$ s)
a. ... s-e	*
b. ... t^h-e	
c. ... c^h-e	
d. ... c-e	

In a traditional OT analysis, IDENT constraints would penalize changes in the features of input obstruents in the output. For example, since /t^h/ is [-delayed release] and [s] is [+delayed release], a change from /t^h/ to [s] would violate IDENT(delayed release). While the IDENT constraints are more commonly used, the family of *MAP-IO constraints produced a closer match between observed and predicted frequencies.[1] While *MAP-IO constraints may be considered language-specific and somewhat ad hoc than IDENT constraints, they can better reproduce the variant frequencies produced by the speakers in the experiment.

6.3.3 Palatalization

In addition to the constraints introduced so far, I also included a constraint *T-I that penalizes [t^h] and [t] before [i]- or [j]-initial suffixes. Specifically, this constraint will ban [t^h-i] in the current analysis (because I chose not to include the [t] variant, and there is no [j]-initial suffix employed in the experiments.)

(20) *A constraint that bans [t^h-i]*

- *T-I: [t^h] and [t] are banned before [i]- or [j]-initial suffixes.

[1] The correlation between observed and predicted frequencies was very high in both sets of constraints, indicating a near-perfect match between the observed and predicted frequencies regardless of the type of faithfulness constraints used. Specifically, the correlation coefficient range from 0.97 (S3) to 1 (S1 and S2) for *MAP-IO constraints (all p's < 0.001), and from 0.95 (S2) to 0.99 (S1) for IDENT constraints (all p's < 0.001). Crucially, however, the model with IDENT constraints had a severe overprediction problem in all speakers (S1-S3): it often predicted greater than zero probabilities for the forms with zero observed frequencies. Thus, I decided to use *MAP-IO constraints.

6.4 Learning data

The learning data fed to the model was the corpus frequencies, which is assumed to represent the (presumably homogeneous) linguistic input that the speakers receive. As was the case in the analysis of the corpus study (see Chapter 3), the observed frequencies of an input type were a combination of type and token frequencies (token-weighted type frequencies). Each word type contributes a total of one unit, divided between different obstruent variants, according to the token frequency of each variant. Let us take /t^h-ɨl/ as an example. There were 16 tokens of /pat^h-ɨl/ 'field-ACC', 10 of which was realized with [s], 1 as [t^h], and 5 as [c^h]. Therefore, this word contributed 10/16=0.625 to the [s] candidate of the input type /t^h-ɨl/, 1/16=0.062 to the faithful [t^h] candidate of /t^h-ɨl/, and 5/16=0.312 to the [c^h] candidate of /t^h-ɨl/.

6.5 Learning simulations

The goal of the modeling is to fit the grammar proposed in Section 6.3 to the corpus, and predict the three speakers' production patterns using the learned weights. The simulation is conducted three times, each representing one speaker's learning, and the model is provided with different bias terms for relevant constraints each time it is run. By doing so, it would be possible to obtain different output frequencies. The assumption is that although individuals receive the same input, individual differences arise because speakers have differently biased grammars.

Prior to this main modeling procedure, I first ensured that the constraints established in Section 6.3 are adequate to explain the observed data. To do this, I provided the model with (token) frequencies of an individual speaker[2], and fit the grammar to this individual data. If the model prediction closely matches the learning data, we can say that the constraint set can

[2] Recall that token-weighted type frequencies rather than token frequencies were used to create learning data from the corpus data to prevent a few types from dominating the patterns (see Section 6.4). However, in the experiment from which individual speakers' frequencies were obtained, all words were presented an equal number of times. In this context, token-weighted type frequencies are irrelevant, and token frequency does not carry the disadvantage of allowing a small number of frequent items to dominate the results.

adequately explain the variation. When constraints are valid, they should be able to capture the patterns in the learning data, and the model should be able to make predictions that closely resemble the learning data. As shown in Fig. 6.1, the model successfully captured the major patterns of observed frequencies for all speakers, indicating that the constraints are valid.

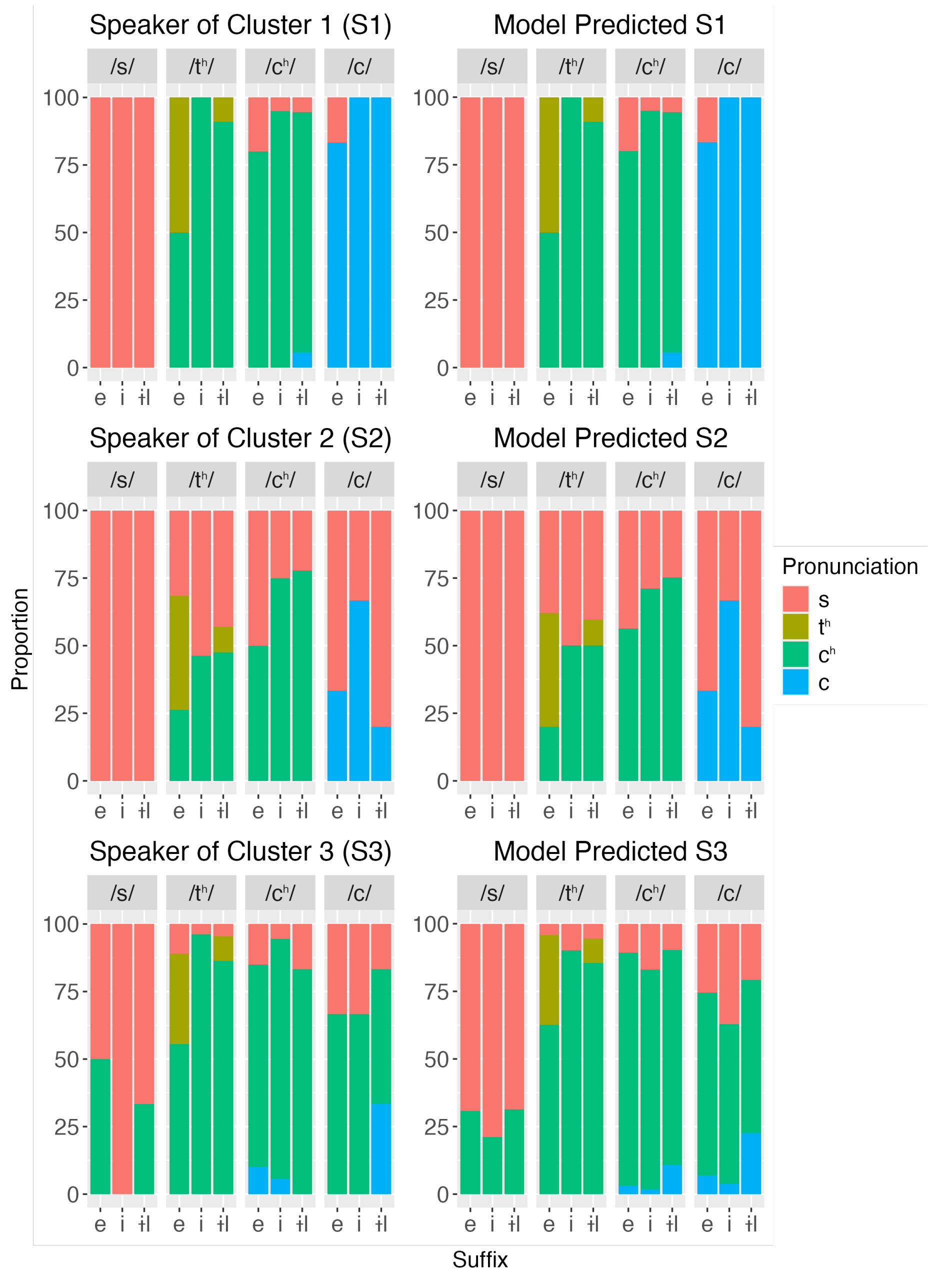

Figure 6.1: Observed and predicted frequencies when the learning data was the speakers' production frequencies in the experiment

Next, I use the corpus frequencies as the learning data and try to predict the individual speakers' output frequencies. As mentioned earlier, I vary the expected weight μ of relevant constraints across different simulations, representing prior bias given to the weights. This means that the speakers have certain prior expectations or biases for the constraint weights, represented as μ parameter of the model. Once those biases are exposed to the lexicon, we get an end state that is a compromise between the biases and the lexicon.

What should be the μ values for constraints in individual speakers' grammars? To answer this question, I referred to the constraint weights learned from the learning simulation described above. Because these weights were found to successfully predict the speakers' output frequencies, it is reasonable to use these weights as the target or expected weight of the constraints. I refer to these sets of weights as "baseline μ's" (these values will be modified later, hence the name "baseline"), which are presented in Table 6.1.

When the models were provided with the corpus frequencies as the learning data and the set of μ values presented in Table 6.1 as well as a σ of 1 for all constraints, their output frequencies did not closely resemble the speakers' production patterns. In Fig. 6.2, the leftmost panels (A1, B1 and C1) present actual speaker's production patterns and the center panels (A2, B2 and C2) present frequencies predicted by these models. Most notably, for Speaker of Cluster 1 and 3 (S1 and S3), the model substantially over-predicted the [s] responses. For Speaker of Cluster 2 (S2), who is characterized by relatively high proportions of [s] responses, the model performed relatively well, except that it slightly over-predicted [s] responses in /c^h/ contexts. The learning simulations thus indicate that even when the constraints were assigned μ values that are the same as weights learned when the grammar was fitted to individual speakers' data, the models, when using the corpus data as the input data, were not able to yield output frequencies that approximate the individual speakers' actual frequencies.

To achieve output frequencies that resemble speakers' response patterns, the μ values were adjusted multiple times. Fig. 6.3 demonstrates three adjustments made to the baseline μ values to achieve model-predicted frequencies that approximate Speaker 1's production. The panel

Constraint	Speaker 1	Speaker 2	Speaker 3
*MAP-IO(s → t^h)	38.7	15.0	15.1
*MAP-IO(s → c^h)	49.2	23.9	0.4
*MAP-IO(s → c)	55.8	19.7	16.1
*MAP-IO(t^h → s)	53.7	16.2	6.2
*MAP-IO(t^h → c^h)	2.9	4.8	3.1
*MAP-IO(t^h → c)	50.7	26.0	23.3
*MAP-IO(c^h → s)	9.4	12.2	2.5
*MAP-IO(c^h → t^h)	39.1	21.7	13.4
*MAP-IO(c^h → c)	23.3	26.2	0.3
*MAP-IO(c → s)	24.8	13.5	2.2
*MAP-IO(c → t^h)	51.1	21.7	18.0
*MAP-IO(c → c^h)	52.0	31.6	0.8
ANTICORR(t, _#, t^h, _e)	14.4	2.3	2.2
ANTICORR(t, _#, c^h, _e)	17.4	6.4	5.9
ANTICORR(t, _#, s, _e)	25.4	18.4	6.3
ANTICORR(t, _#, c, _e)	2.2	4.2	2.8
ANTICORR(t, _#, c^h, _i)	5.4	4.7	8.3
ANTICORR(t, _#, s, _i)	11.9	16.0	9.3
ANTICORR(t, _#, c, _i)	10.1	3.2	4.8
ANTICORR(t, _#, t^h, _ɨl)	6.6	2.5	1.4
ANTICORR(t, _#, c^h, _ɨl)	11.9	9.0	6.7
ANTICORR(t, _#, s, _ɨl)	18.5	20.1	7.1
ANTICORR(t, _#, c, _ɨl)	32.4	5.1	5.0
*T-I	27.4	17.0	10.9

Table 6.1: Baseline μ values for constraints for individual speakers

labeled as “S1, Baseline μ’s” is the same as A2 of Fig. 6.2, and the panel labeled as “S1, Step 3” is the same as A3 of Fig. 6.2. Because the model with the baseline μ’s Table 6.1 over-predicted [s] responses and under-predicted [c^h] responses (compare A1 and A2 in Fig. 6.2), I first lowered the μ values for the three ANTICORR constraints that encourage [s] responses (ANTICORR(t, _#, s, _e), ANTICORR(t, _#, s, _i) and ANTICORR(t, _#, s, _ɨl)) by about 5 to 10 (Step 1). As a result of this adjustment, the proportion of [s] responses decreased (“S1, Step 1” of Fig. 6.3), as intended. However, the proportion of [c^h] responses was still lower than the observed data.

In Step 2, I first raised the μ values for the three ANTICORR constraints that encourage [c^h]

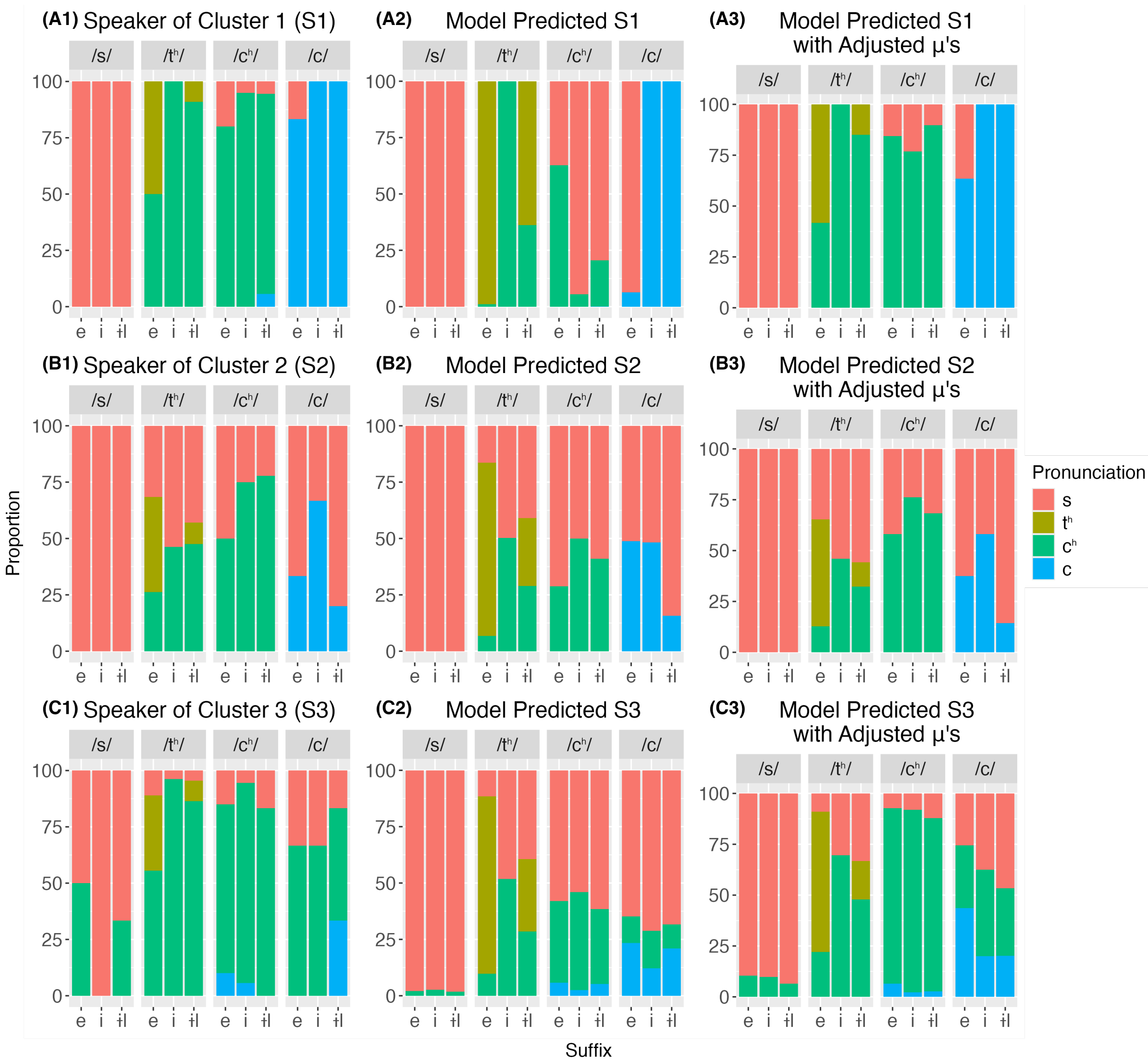

Figure 6.2: Observed frequencies of individual speakers (A1, B1, C1), model-predicted frequencies when μ values were the same as weights learned in a model trained on individual data (A2, B2, C2), and model-predicted frequencies when μ values were adjusted (A3, B3, C3).

responses, ANTICORR(t, _#, c^{h}, _e), ANTICORR(t, _#, c^{h}, _i) and ANTICORR(t, _#, c^{h}, _ɨl, by about 10. I also lowered the μ of *MAP-IO(t^{h} → c^{h}). As shown in "S1, Step 2" of Fig. 6.3, the proportion of [c^{h}] in /t^{h}/ did indeed increase from Step 1 to Step 2 in both /t^{h}/ and /c^{h}/. However, the proportion of [c^{h}] in /t^{h}/ was still lower than the observed proportions.

To address this problem, in Step 3, I lowered the μ values for ANTICORR(t, _#, t^{h}, _e) and

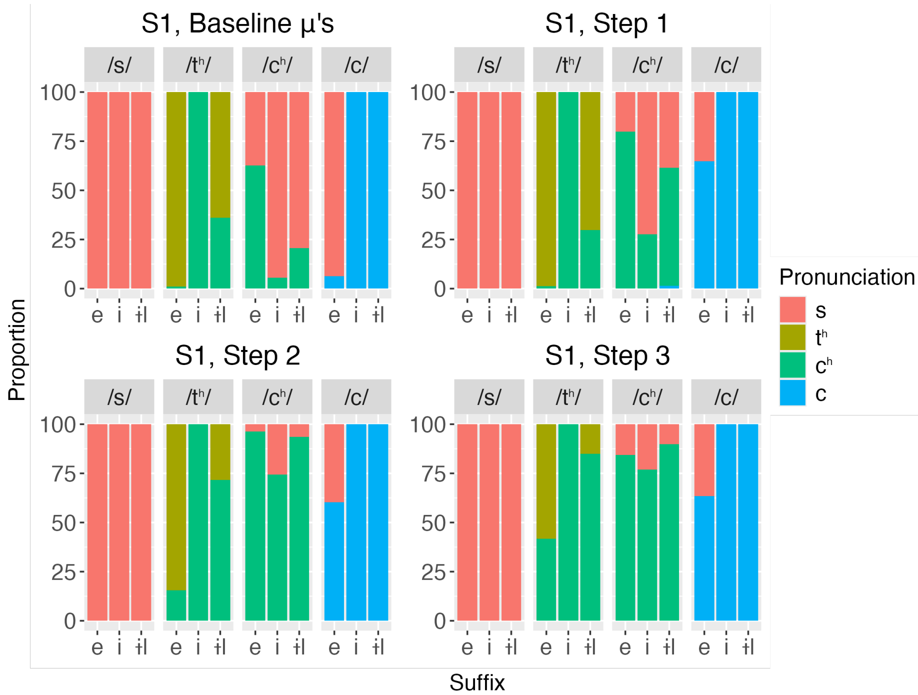

Figure 6.3: Adjusting μ's for the model to match Speaker 1's response

ANTICORR(t, _#, t^{h}, _ɨl) by around 5 to 10. With a lower μ for these two constraints, the model now predicted a higher proportion of [c^{h}] for /t^{h}-e/ and /t^{h}-ɨl/, as shown in "S1, Step 3" of Fig. 6.3. This is very similar to Speaker 1's production patterns, presented in A1 of Fig. 6.2.

S2 and S3 were modeled in the same way: the speaker's observed frequencies were input into a model, which assigned weights to the constraints. I modified these weights, through trial and error, to be provided to the subsequent model as μ values, along with the corpus frequencies as the input data. This second model predicted frequencies that closely matched the speaker's observed frequencies. As shown in Fig. 6.4, the correlation between observed and model-predicted frequencies were reasonably high in all speakers. The learned weights are presented in Table 6.2.

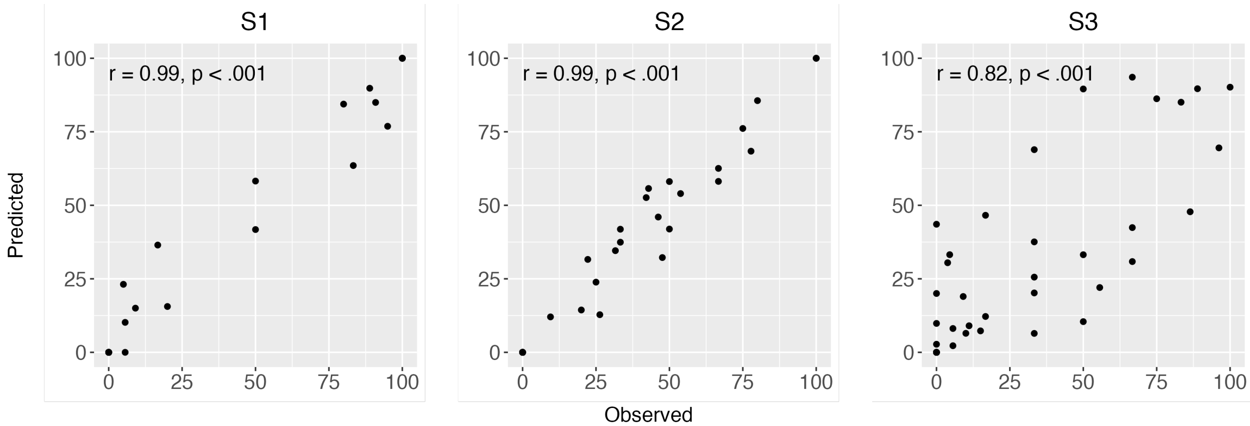

Figure 6.4: Correlation between observed and model-predicted frequencies in the final models

Constraint	**Speaker 1**	**Speaker 2**	**Speaker 3**
*MAP-IO(s → t^h)	38.7	15.0	15.1
*MAP-IO(s → c^h)	49.2	23.9	4.6
*MAP-IO(s → c)	55.8	19.7	16.1
*MAP-IO(t^h → s)	44.8	14.1	3.9
*MAP-IO(t^h → c^h)	13.3	1.1	5.4
*MAP-IO(t^h → c)	50.6	25.9	23.2
*MAP-IO(c^h → s)	4.1	14.3	0.0
*MAP-IO(c^h → t^h)	39.0	21.6	13.3
*MAP-IO(c^h → c)	23.3	26.2	0.4
*MAP-IO(c → s)	20.5	13.3	0.3
*MAP-IO(c → t^h)	51.1	21.7	18.0
*MAP-IO(c → c^h)	52.0	31.6	2.6
ANTICORR(t, _#, t^h, _e)	7.1	4.9	1.2
ANTICORR(t, _#, c^h, _e)	20.1	4.6	5.4
ANTICORR(t, _#, s, _e)	22.5	18.6	3.0
ANTICORR(t, _#, c, _e)	2.5	4.8	3.2
ANTICORR(t, _#, c^h, _i)	9.4	3.9	8.3
ANTICORR(t, _#, s, _i)	12.2	17.1	5.9
ANTICORR(t, _#, c, _i)	8.0	4.2	5.0
ANTICORR(t, _#, t^h, _ɨl)	1.8	4.6	1.5
ANTICORR(t, _#, c^h, _ɨl)	16.9	6.7	7.8
ANTICORR(t, _#, s, _ɨl)	18.8	20.2	5.9
ANTICORR(t, _#, c, _ɨl)	29.9	5.2	4.8
*T-I	26.9	16.5	10.4

Table 6.2: Weights learned from the final models

6.6 Discussion

To summarize the results of the learning simulations, using the corpus frequencies as the training data and a set of μ's that are specific to individuals (along with a small σ), the model was able to arrive at individual speakers' probabilistic grammar. Crucially, the μ values were drawn, with several adjustments, from the weights learned in a separate model trained with the speakers' observed frequencies. When these weights were used as the μ's without any modifications, the model-predicted frequencies did not match the speakers' behavior. This is not surprising, because the model learns a weight that is partway between the μ values and the weights motivated by the lexicon. For example, if the speaker has a weight of 3 for a certain constraint in their grammar and this value is provided as a μ, the learned weight in that model should be a compromise between μ=3 and the weight supported by the learning data. If the learning data suggests a weight of 1, the learned weight would be a value between 1 and 3 (e.g. 2), not 3. This results in the discrepancy between the model prediction (which would be based on a weight of 2) and the observed pattern of the speaker (which would be based on 3).

Before discussing the implications of the modeling procedures, I would like to provide some comments on the use of μ values in the present study. In this study, the μ values for individual speakers' grammars are initially estimated based on their production outputs. This differs from the typical use of μ values in most studies, where they represent prior biases for constraint weights, assumed to be innate, universal, and often phonetically grounded. In the current modeling, they function as a parameter in the model that generates variation, but not necessarily representing biases that individuals possess prior to learning. Further, the source of the μ values is uncertain; these values could potentially reflect differences in cognitive traits, pressure toward normative pronunciation, or language-related experience (the model does not have room for explicitly encoding these sources of variability). The μ values are a way to encapsulate multiple sources of individual variation. Since differences in learning data are unlikely to account for the degree of variability observed in the experiment (as discussed below), it is the variation in μ values that drives the differences in the model's outputs. These μ values remain descriptive, merely capturing variation without explaining why or how individual learners differ. Moreover, there is no principled method for determining the appropriate μ values for each speaker; instead, they are found through adjustments based on trial and error.

The modeling procedures demonstrate that each speaker's end grammar is a compromise between weights supported by the lexicon, and individual factors encapsulated in the speakers' μ values. What do the results tell us about the source of inter-speaker variation? One interpretation is that individuals within a speech community are exposed to the same input but across-speaker differences arise because speakers have varying μ values. An alternative interpretation, also consistent with the current modeling, is that the speakers actually have the same μ values, but because their actual input data differed, the μ values we obtain when we assume they were all exposed to the same input data have to be different. As a concrete example, suppose that in reality two speakers (Speaker A and B) have μ=3 for some constraint. During language acquisition, Speaker A is exposed to data that motivate a weight of 5 for this constraint, and they end up with actual weight between 3 and 5 (e.g. 4). Speaker B is exposed to data that motivate a weight of 1, and they learn a weight between 1 and 3 (e.g. 2). If we try to model both speakers' acquisition using a corpus that suggests a weight of 3, we'll have to give Speaker A a higher μ and Speaker B a lower μ, even though in reality they had the same μ (but different training data). In this scenario, inter-speaker variation arises due to differences in the input the individuals receive.

The model failed to predict individual speakers' response patterns even when the set of weights learned from the speaker's actual data were given as μ's (unadjusted). The most notable discrepancy was that the model substantially over-predicted [s] responses in S1 and S3. (For S2, the proportion of [s] was only slightly higher in the model prediction compared to the observed data.) The apparent reason for such a discrepancy between model predictions and speakers' behavior in the experiment is that the input data (corpus frequencies) has a great amount of [s] tokens but the biases were not strong enough to suppress [s] variants in the output to the degree that the speakers (dis)favored them. We must then consider why the speakers deviated from the lexicon by producing far fewer [s] variants. I speculate that this could be due to the design of the stimuli; the stimuli set did not accurately reflect the lexical distributions of stem-final obstruents in that stems ending in /s/ accounted for only 13.3% (4/30) of the target stems. Such underrepresentation of /s/-final stems in the experiment might have reduced the speakers' propensity to produce [s] responses.

Another possibility is that the corpus data do not accurately represent the language input that speakers use to learn grammar. Specifically, speakers may base their learning on either token frequency or type frequency, rather than on token-weighted type frequency as implemented in the current modeling (see Section 6.4). To test whether either token frequency or type frequency more closely approximates the input that some speakers are exposed to, I examined whether the model's tendency to over-predict [s] is mitigated when the learning data are based on token or type frequency.[3] As shown in Fig. 6.5, when the learning data is created based on token or type frequency, the proportion of [s] variants is still higher in the model prediction compared with responses of S1 and S3.

Yet another possible input type is frequencies obtained from child-directed speech. As child-directed speech serves as direct language input to the learners, it could predict the speakers' behavior better than data obtained from adult-to-adult interaction. Experiments conducted in Ko & Jun (2024) found that the proportion of canonical forms was higher in child-directed speech than adult-directed speech (because mothers used high-frequency words, which are more likely to be produced with canonical variants compared to low-frequency words, in their speech directed to children.) It is therefore possible that response patterns of some speakers such as S1,

[3] In the learning data based on token frequency, I entered the raw count of all instances in which each pronunciation was used. For instance, words with a stem ending in /t^h/ and the /-e/ suffix appeared 410 times as [t^h], 12 times as [s], and 4 times as [c^h]. These counts were used as the observed frequencies for the candidates [t^h], [s], and [c^h] for the input type /t^h-e/. For type frequency, if a particular pronunciation occurred for a word at least once, it was counted as one. Regardless of how often a word with a specific pronunciation appeared, it contributed a count of one to the type frequency for that pronunciation. In the example of the input type /t^h-e/, the frequency for the candidate [s] was 4, because there were four word types (/pat^h-e/, /pjʌt^h-e/, /pak*at^h-e/, /sot^h-e/) that were produced with an [s] at least once.

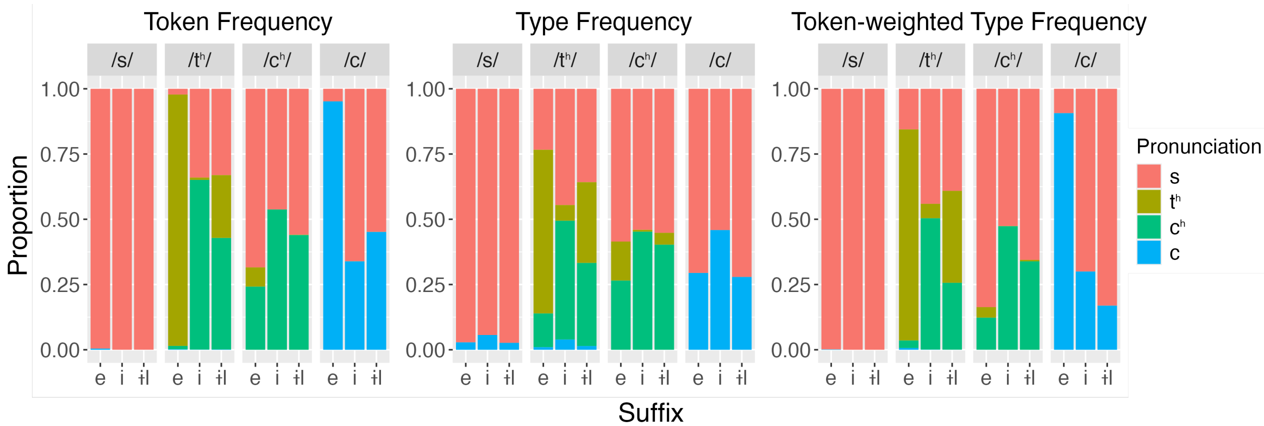

Figure 6.5: Model-predicted frequencies based on different learning data

who produced relatively higher proportions of canonical forms, can be better predicted with learning data based on child-directed speech.

To test this possibility, I used the Ko corpus of Korean child-directed speech (Ko et al., 2020) to extract all words consisting of a stem ending in a coronal obstruent and a vowel-initial suffix. Although to the best of my knowledge the Ko corpus is the largest Korean child-directed speech corpus available, consisting of approximately 70,000 words, there were only 41 such words. Only 26 of these contained one of the three suffixes employed in the Experiment 1 and 2 (that is, /-e/, /-i/, /-ɨl/); all the /t^h/-final stems were suffixed with /-e/, all the /c^h/-final stems with /-i/, and none of these had /c/-final stems. Therefore, it seems difficult to learn from child-directed speech how words like /...t^h-ɨl/, /...c^h-e/, and /...c-i/, etc. should be pronounced. The small number of target tokens in the child-directed speech corpus then makes it difficult to use them as learning data.[4]

Finally, to test the effect of input differences especially with regard to the speaker variation in the production of [s] responses, I conducted a series of learning simulations with training

[4]This problem is probably not specific to the Ko corpus used in the current analysis; Albright (2008) notes that obstruent-final nouns only account for 18% of the Korean nouns (cf. sonorant-final 42%, vowel-final 39%). Further, among these obstruent-final nouns, those ending in a *coronal* obstruent, the focus of the current study, are underrepresented compared to those ending in a labial or velar obstruent. Therefore, the insufficient learning data seems to be a general challenge that a child learning the inflectional paradigm faces.

data varying in the frequency of [s]. The aim was to see if the model is able to match S1 and S3's proportion of [s] responses when the the training data were manipulated to have a lower proportions of [s] variants. As shown in Fig. 6.6, only when the frequencies of [s] were reduced to 1/10 of their actual frequencies (e.g. the frequency of [s] for the input type /s-e/ changed from 8.99 to 0.899, the frequency of [s] for the /t^h-e/ changed from 1.63 to 0.163, and so on) did the model-predicted [s] frequencies became closer to those of S1 and S3. I am not sure if it is feasible that one speaker is exposed to ten times as many [s] responses as another speaker. Future research is necessary to more closely investigate the effect of input difference on individual differences in phonology.

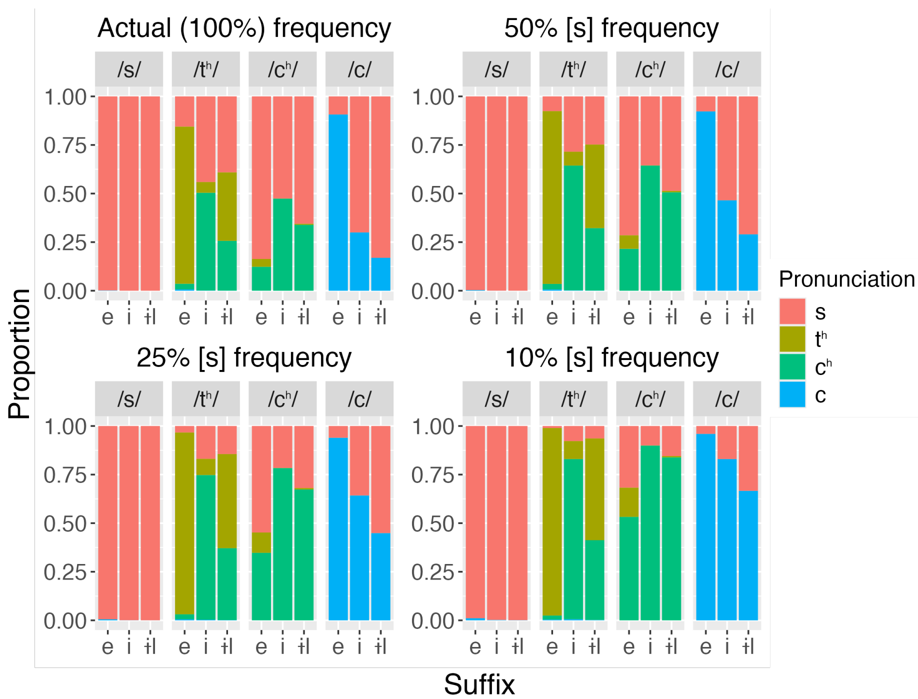

Figure 6.6: Model predictions when the frequencies of [s] in the input are manipulated

CHAPTER 7

Discussion and conclusion

In this dissertation, I investigated inter-speaker variation in the realization of coronal obstruents in the final position of noun stems in Korean. I first conducted a corpus study to understand the statistical distribution of these variants. Following this, I carried out two longitudinally designed experiments with different tasks to examine whether systematic and reliable individual differences could be identified and to explore the sources of this variability. The results showed a substantial degree of variation across speakers in each experiment, and the participants were consistent in their choice of variants across the two experiments. Finally, I modeled individual speakers' production patterns using a MaxEnt grammar, assigning different bias terms to the constraints for each speaker. In the remainder of this chapter, I discuss consistency and systematicity of individual differences, as well as the potential sources of individual variability as reflected in the experimental and modeling results.

7.1 Systematicity of inter-speaker variation

One view of language learning holds that individuals who received roughly the same linguistic input converge on the same linguistic knowledge (grammar), and that individual differences arise due to random variation or differences in factors related to performance or language use (Chomsky, 1975). Although individual differences are commonly reported in the literature, relatively few studies have examined whether the observed differences truly reflect stable and inherent traits of the individuals and their linguistic systems (Kingston et al., 2015; James et al., 2018; Durvasula, 2024). A few previous studies directly tested the reliability of individual differences in linguistic behavior, either by testing the same set of participants two times (Han et al., 2016; Kong & Edwards, 2016; Wade et al., 2021) or by calculating split-half correlations, checking for consistency between the two halves of the test items (James et al., 2018). These studies argue that the most compelling evidence for individual differences reflecting characteristic traits of individuals comes from studies that evaluate whether individuals show consistent behavior across different occasions.

The current longitudinal study contributes to this limited body of research. Through two different experimental tasks, I showed that speakers were consistent in their choice of variants over time (e.g. those who provided more innovative [s] responses in Experiment 1 were more likely to choose innovative [s] responses in Experiment 2). The finding that individual speakers were consistent within themselves across two experiments, which employed different tasks and were separated by up to 1.5 years, attests to the reliability of inter-speaker variation. Consequently, I reject the possibility that individual differences can be dismissed as mere random errors.

Another aspect of individual differences in phonology that suggests its systematicity is that they are explained by other traits of individuals. As will be outlined below, a number of individual-level characteristics were associated with individuals' probability of producing or choosing canonical pronunciations, as well as innovative responses in certain phonological environments (e.g. [s] responses for /c/-final stems). This adds to the evidence that individual variation in phonology is not random but is instead systematically related to specific individual traits.

7.2 Sources of individual differences

This study investigated several potential sources of the observed inter-speaker variation in phonology: demographic factors (specifically, age and gender), pressure to conform to standard pronunciation (including the speaker's belief in the importance of producing normative pronunciation and whether the speaker knows what they are being tested on), cognitive traits (phonological working memory, inhibitory skills, autistic traits), differences in input data, differences in the lexicon (as reflected in the spelling test and perceived familiarity of the test items), as well as differences in phonological grammar. In the following subsections, I discuss each factor separately in terms of whether they explained the variable realization of stem-final obstruents, while acknowledging that these factors are in fact interrelated; for instance, sociolinguistic background can influence the type of language input one receives, which in turn affects the grammar an individual learns.

7.2.1 Demographics

A survey at the end of the experiment sessions collected information about the speaker age and gender. The effect of age on the proportion of canonical forms was not credible in either Experiment 1 (production task) or in Experiment 2 (multiple choice questions). One potential reason for the null effect may be that the variability in the participant age was relatively small to begin with – all participants were in their 20s or 30s except two, who were in their 40s. Since the variation shown in the realization of stem-final obstruents is a diachronic analogical change, it is plausible to think that younger speakers tend to produce more innovative variants. Similar to the findings of the present study, however, Choi (2004) found limited age-dependent variation in the proportion of canonical pronunciation of stem-final /t^h/ and /c^h/. The difference among the four age groups examined (40s, 50s, 60s and 70s) was only small.

The effect of gender was found in the multiple choice questions in Experiment 2, but not in the production test in Experiment 1. More specifically, male speakers were more likely to choose the canonical form than female speakers when they were asked to choose their pronunciation among several written options, but not when they were asked to produce the inflected form of the stem. The gender difference reported in the literature, namely that female speakers tend to be more innovative compared to male speakers in their speech patterns, was evident in the present study only when the purpose of the experiment was explicit and the participants were fully aware of which variant they were choosing. In comparison, Choi's (2004) survey, in which speakers were in their 40s or older, did not find a gender difference, although they used a written survey (similar to the multiple choice questions in this study). All in all, it seems that gender difference is only evident in younger speakers' responses.

Another sociolinguistic variable that could have influenced the pronunciation of stem-final obstruents is education level. Although the present study did not formally collect data on participants' education levels, my informal interactions with them suggest that most, if not all, participants were college graduates or higher, indicating minimal variability in this regard. Additionally, the spelling test included in Experiment 1, which can serve as a proxy for education level, showed little variability in scores. As a result, it was not possible to test the effect of education level in this study. We might expect that individuals with higher education levels are more familiar with prescriptive language rules and, therefore, more likely to provide canonical responses. However, Choi (2004) found this to be only partially true. Among the five groups of respondents (no schooling, elementary school graduates, middle school graduates, high school graduates, and college graduates), the highest proportion of canonical form selection was, as anticipated, in the highest-education group (college graduates). Surprisingly, the next highest proportion was observed in the middle group (middle school graduates). The findings must be interpreted with caution, though, since education level was somewhat confounded with age in Choi (2004). In any case, the effect of education level does not seem as straightforward as one might assume.

7.2.2 Conservativity in pronunciation

Since the variants produced for stem-final obstruents can be categorized into canonical and non-canonical forms, I investigated whether individuals' tendency to conform to normative pronunciations could explain the variation between speakers. One measure of this conservatism was the strength of an individual's belief in the importance of correct pronunciation. While this factor did not predict the likelihood of producing a canonical form in the production task (Experiment 1), it did reliably predict the probability of choosing a canonical form in the multiple choice questions (Experiment 2). Thus, we have limited evidence that pressure toward normative pronunciation influences individuals' speech patterns. Only when participants were explicitly asked to choose their pronunciation among written representations of the alternants, likely making them more conscious of their responses, did the pressure for standard pronunciation reliably affect the results.

Participants' awareness of the study's purpose in Experiment 1 can serve as another metric for conservatism in pronunciation, as individuals are more likely to provide what they believe is the correct form when they know what they are being tested on. Indeed, those who correctly guessed the purpose of the experiment produced canonical forms more frequently than those who did not. In sum, the findings suggest that the tendency to follow the prescriptive rules of pronunciation can explain the inter-speaker variation in a limited way.

7.2.3 Cognitive traits

The variable realizations of stem-final obstruents were explained by a subset of the cognitive traits examined: inhibitory skills and one aspect of autistic traits, specifically, attention switching or strong focus. As discussed in Section 5.7, it is not certain whether the cognitive characteristics influenced how individuals learn phonological grammar prior to coming to the experiment (solid lines in Fig. 1.1, repeated below as Fig. 7.1) or how they used their phonological knowledge during the experiment (dashed lines in Fig. 7.1). However, the findings do suggest that the observed inter-speaker variation reflects inherent characteristics of the individuals, as these differences in cognition credibly predict such variation.

Phonological working memory capacity and other autistic traits (attention to detail, communication, imagination and social skills) did not reliably explain the variability in phonological behavior. The null effects also warrant discussion. In recent decades, research has increasingly explored the relationship between language and cognition (see Yu & Zellou, 2019 for a review). While evidence is accumulating that phonological behavior can be predicted by domain-

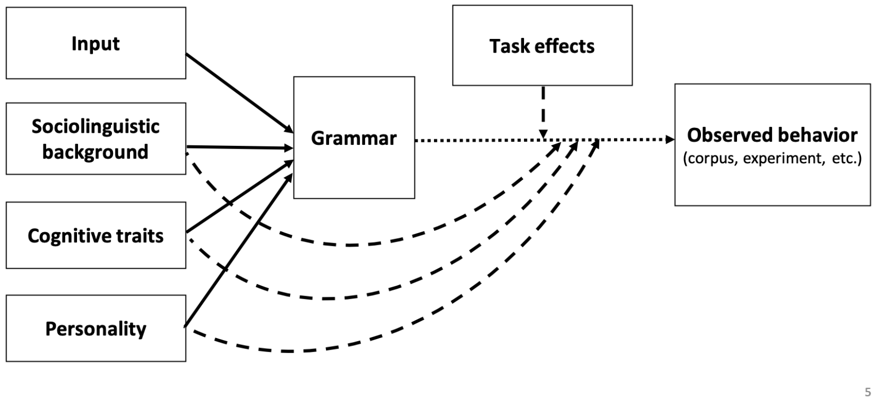

Figure 7.1: Factors influencing individual speakers' phonological behavior

general cognitive traits, some studies have reported no significant relationship between the two (Kong & Edwards, 2016; Kapnoula et al., 2017; Hall-Lew et al., 2021; Wade, 2022). One reason for the lack of agreement on whether there exists a relationship between linguistic and cognitive traits may be that different studies investigate different linguistic behavior and cognitive measures. For instance, speech perception may be more reliably predicted by non-linguistic characteristics of individuals than production is (Lev-Ari & Peperkamp, 2014; Yu & Zellou, 2019); additionally, individual differences in certain phonetic measures might be explained by specific cognitive traits but not others, or even by a particular subscale of the AQ score but not the total score (see Section 5.5). As we test a wider range of linguistic variables and cognitive measures and build a more comprehensive understanding of these relationships, we can begin to theorize why certain cognitive traits might predict specific linguistic behaviors. We can then formulate more targeted linking hypotheses for future studies, enabling an insight into the mechanisms that underlie the link between cognitive traits and linguistic behavior.

Although null results do not necessarily indicate that a relationship does not exist, they are important and should be taken seriously, as null results tend to go underreported in the research community. Consequently, the relationship between linguistic and domain-general cognitive behavior might be overrepresented. This study offers only preliminary insights into the relationship between cognition and language, highlighting the need for further research to clarify the extent to which language relies on domain-general systems (Fodor, 1983). Crucially, future research should aim to refine our methods for examining how phonological variation can be explained by cognitive differences, by involving a larger sample of participants and employing a broader range of behavioral tests.

7.2.4 Language input

One potential source of individual differences that the present study was not able to directly examine is the language input that individuals received. Even within a speech community where the individuals speak the same dialect (the participants spoke Seoul Korean), it is possible that individuals were exposed to different portions of the language, leading them to learn different grammars. This is likely, especially given that learning data on the noun stems ending in a coronal obstruent is sparse in the language (Albright, 2008) and that sparse learning data often results in across-speaker variation (Niyogi & Robert, 1997; Reilly, 2007; Han et al., 2016, 2007). In Chapter 6, I demonstrated that the MaxEnt model can predict individual differences in the proportion of [s] responses when the frequencies of [s] forms in the learning data are varied. This suggests that input differences could account for across-speaker variation, although it remains unclear whether the variability introduced in the input data during the modeling process reflects the variability in the language input that individuals actually receive.

7.2.5 Lexicon

Individual differences in the lexicon and how effectively individuals access it may also influence their pronunciation of stem-final obstruents. There are several ways in which variations in lexical entries and lexical access can affect phonological behavior. First, exposure to different input data may result in individuals possessing different lexical entries. For example, if someone has never encountered the [c^h] variant for /s/-final nouns, they are less likely to store [c^h] as a variant for such nouns and, consequently, less likely to produce this variant in the experiment. Additionally, individuals may be exposed to different amounts of /s/-final stems in their learning data, which could also affect their composition of lexical entries. Even when individuals are exposed to identical learning data, cognitive traits can lead some individuals to learn lexical statistics better than others. Furthermore, individuals may vary in how successfully they utilize their lexical knowledge during the experiment.

It is not practical to accurately assess individual differences in the lexicon and how successfully individuals access it, so the present study did not directly investigate differences in lexical knowledge. However, two components of Experiment 1 do explore aspects of individuals' lexical knowledge and how they access it, which may potentially affect their responses for existing words. One task involves providing the spelling of the stems, which is what the participants believe to be the normative underlying representation, or what they think others' underlying representation is. The other involves rating stem familiarity, where stems perceived as familiar to the speaker may be retrieved more readily during the experiment. Little evidence was found to support the influence of either factor. Participants performed at ceiling in the spelling test, indicating minimal variation in what they think the normative or lexical representation of each stem is. Additionally, stem familiarity was not a credible predictor of the proportion of canonical forms in either experiment. Thus, although it is plausible to think that individual differences in lexicon-related factors could affect pronunciation, we found no experimental evidence directly showing the effect of these differences.

A nonce word test was conducted in Experiment 2 to preclude the influence of lexicon and explore the effect of grammar, testing the hypothesis that the individual differences arise solely due to differences in the lexicon. The findings provide evidence, albeit weak, that individual differences are not solely attributable to differences in the lexicon. Although there was only a small degree of variability in the speakers' responses for the nonce words (likely due to the design of the task), they were somewhat related to individuals' responses for the existing words, suggesting some degree of productivity. In Jun's (2010) well-formedness rating task, the average ratings for the obstruent variants in the wug words matched the relative frequency of stems ending in those obstruents in the corpus. This finding also indicates that the variable pattern is productive (at least when responses are considered in aggregate; individual differences were not examined in that study). I discuss the results of the nonce word test again in Section 7.2.7

7.2.6 Task effects

Individuals may vary in their familiarity with the experimental setup; for instance, some might feel more nervous than others when using a computer or pressing the *Start recording* button. This could have led those individuals' behavior in the experimental sessions to deviate substantially from how they typically produce the variants in their real lives outside the testing session. Among the studies focusing on individual differences, only a few directly considered the potential effect of familiarity with the testing situation or failure to understand the experimental task (e.g. Dąbrowska, 2012; Bennett & Braver, 2020); these studies concluded that the observed variation stems from differences in the linguistic knowledge rather than differences in how successfully individuals perform a given task.

To minimize these task-related effects in the present study, the experiments were designed to be as undemanding as possible. Participants completed the experiments at a time and place convenient for them, were provided with practice items to get familiarized with the task, and were given unlimited time to respond in both the production task and multiple choice questions. These were efforts to minimize the influence of performance-related factors in individual differences such as memory capacity, attention spans, access to lexicon, and other physiological or cognitive constraints, allowing the participants to access their lexical or grammatical knowledge as successfully as possible. As participants performed at ceiling in using the correct suffix in the production experiment, one can safely assume that the task was sufficiently easy. Therefore, similar to the arguments of previous studies mentioned above, I argue that the inter-speaker variability reflects individuals' underlying knowledge about their language rather than how they adapt differently to the experimental environments.

7.2.7 Phonological grammar

All of the factors mentioned thus far are typically regarded as extragrammtical factors, although they may influence how individuals learn and use the grammar. Is there evidence that the observed individual differences stem from differences in the phonological grammars? While it is not feasible to fully tease apart extragrammatical and grammatical components involved in speech production and comprehension, compelling evidence for the role of grammar should come from individuals' behavior in wug tests. The working assumption is that if individuals are consistent within themselves across an experiment employing real words and another one employing nonce words, and the task was designed so that differences in task performance are minimized, one could argue that the observed inter-speaker variation in the real words reflect differences in their grammatical knowledge. The present study found that those individuals who produced [s] frequently for real words also exhibited a higher proportion of [s] responses for nonce words, suggesting that inter-speaker variation in phonology is at least partially rooted in differences in grammar. This challenges the view that such variation arises solely from factors related to performance or language use.

If there are differences in phonological grammar, represented by the μ values in the modeling, then where do they come from? Among the factors that could influence individuals' grammar learning identified in Fig. 7.1 (solid lines), the present study discussed the potential influence of input difference, sociolinguistic variables (specifically, demographics and pronunciation conservatism) and cognitive traits. As for potential differences in learning samples, I speculated that the variability in the input may explain some but not all variation in the learning outcome. Among the demographic factors, it was hard to test the age effect, and gender effect was limited in its scope. Tendency to conform to normative pronunciation explained the individuals' probability of choosing the canonical form, as one might expect. A subset of the cognitive traits examined can also explain some of the variance.

If there are remaining portions of the differences in grammar that cannot be explained by the factors just mentioned, what are their origins? One speculation is that there are other extra-grammatical factors that were not (directly) tested in the current study that influence learning of phonological grammar, e.g. age, education level, personality traits, or cognitive traits not employed in the current study. Another speculation, not incompatible with the first one, is that there is variability in predispositions for language learning. In this scenario, individuals might be born with different innate aptitudes for internalizing specific rules of a language. While this study has identified the sources of variation in individual speech patterns as much as possible, it may have only addressed a small portion of this variation.

Some of the factors just discussed are also likely to influence how speakers *use* their grammatical knowledge during the experiment, as indicated by the dashed lines at the bottom of Fig. 7.1. In fact, it is unclear whether the factors that credibly predict variation influence language learning prior to the experiment (e.g., speakers with stronger focus learn lexical statistics more effectively) or exert their effects during the experiment (e.g., the same individuals focus better on the task and can better retrieve their grammatical knowledge). I leave it to future research to differentiate these two possibilities.

7.3 Conclusion

In this dissertation, I have argued that individual differences in speech patterns are stable, systematic, and reflective of true variability in the underlying linguistic knowledge that individuals possess. As discussed, an individual's speech pattern results from numerous interdependent factors. As argued at the beginning of this dissertation, understanding the nature and sources of inter-speaker variation in phonology should be central to phonological theory and analysis, as it can provide insights into the structure of phonological grammars and other linguistic phenomena. Future research should aim to further test the reliability of individual differences across a wider range of phonological processes and to pinpoint the origins of this variability.

Bibliography

Adams, Anne-Marie & Susan E. Gathercole. 1996. Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology Section A* 49(1). 216–233. doi:10.1080/713755610.

Albright, Adam. 2002a. *The Identification of Bases in Morphological Paradigms*: University of California, Los Angeles dissertation.

Albright, Adam. 2002b. Islands of Reliability for Regular Morphology: Evidence from Italian. *Language* 78(4). 684–709. doi:10.1353/lan.2003.0002.

Albright, Adam. 2005. The Morphological Basis of Paradigm Leveling. In Laura J. Downing, Tracy Alan Hall & Renate Raffelsiefen (eds.), *Paradigms in Phonological Theory*, 17–43. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199267712.003.0002.

Albright, Adam. 2008. Explaining Universal Tendencies and Language Particulars in Analogical Change. In Jeff Good (ed.), *Linguistic Universals and Language Change*, 144–181. Oxford University Press. doi:10.1093/acprof:oso/9780199298495.003.0007.

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161. doi:10.1016/S0010-0277(03)00146-X.

Allen, J Sean, Joanne L Miller & David DeSteno. 2003. Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America* 113(1). 544–552.

Armstrong, Richard A. 2014. When to use the Bonferroni correction. *Ophthalmic and Physiological Optics* 34(5). 502–508. doi:10.1111/opo.12131.

Arthur, Dana T., Michael T. Ullman & F. Sayako Earle. 2021. Declarative Memory Predicts Phonological Processing Abilities in Adulthood. *Frontiers in Psychology* 12. 658402. doi:10.3389/fpsyg.2021.658402.

Baayen, R.Harald, Ton Dijkstra & Robert Schreuder. 1997. Singulars and plurals in dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language* 37(1). 94–117. doi:10.1006/jmla.1997.2509.

Baddeley, Alan, Susan Gathercole & Costanza Papagno. 1998. The phonological loop as a language learning device. *Psychological Review* 105(1). 158–173. doi:10.1037/0033-295X.105.1.158.

Baker, Adam, Diana Archangeli & Jeff Mielke. 2011. Variability in american english s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23(3). 347–374. doi:10.1017/S0954394511000135.

Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin & Emma Clubley. 2001. The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders* 31(1). 5–17. doi:10.1023/A:1005653411471.

Barrick, Murray R. & Michael K. Mount. 1991. The big five personallity dimensions and job performance: a meta-analysis. *Personnel Psychology* 44(1). 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x.

Bayles, Andrew, Aaron Kaplan & Abby Kaplan. 2016. Inter- and intra-speaker variation in French schwa. *Glossa: a journal of general linguistics* 1(1). 1–30. doi:10.5334/gjgl.54.

Beddor, Patrice Speeter. 2009. A Coarticulatory Path to Sound Change. *Language* 85(4). 785–821. doi:10.1353/lan.0.0165.

Beddor, Patrice Speeter. 2012. Perception grammars and sound change. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: Perception, production, and social factors*, 37–55. Philadelphia, PA: John Benjamins.

Benjamini, Yoav & Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1). 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Bennett, William G. & Aaron Braver. 2020. Different speakers, different grammars: Productivity and representation of xhosa labial palatalization. *Phonological Data and Analysis* 2(6). 1–29. doi:10.3765/pda.v2art6.9.

Bishop, Jason. 2021. Exploring the similarity between implicit and explicit prosody: Prosodic phrasing and individual differences. *Language and Speech* 64(4). 873–899. doi:10.1177/0023830920972732. PMID: 33238799.

Bishop, Jason, Adam J. Chong & Sun-Ah Jun. 2015. Individual differences in prosodic strategies to sentence parsing. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the International Congress of Phonetics Sciences 18*, Glasgow, UK: The University of Glasgow.

Bishop, Jason & Darlene Intlekofer. 2020. Lower working memory capacity is associated with shorter prosodic phrases: Implications for speech production planning. In *Proceedings of Speech Prosody 10*, 191–195. Tokyo, Japan.

Bissell, Marie. 2021. Automatic phonetic classification of vocalic allophones in Tol. *Proceedings of the Linguistic Society of America* 6(1). 403. doi:10.3765/plsa.v6i1.4977.

Bornstein, M. H., M. O. Haynes & R. M. Painter. 1998. Sources of child vocabulary competence: A multivariate model. *Journal of Child Language* 25. 367–393.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. doi:10.18637/jss.v080.i01.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411. doi:10.32614/RJ-2018-017.

Bürkner, Paul-Christian. 2021. Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software* 100(5). 1–54. doi:10.18637/jss.v100.i05.

Chodroff, Eleanor & Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61. 30–47. doi:10.1016/j.wocn.2017.01.001.

Choi, Hyewon. 2004. *Phyocwun palum silthay cosa 3 [A survey of standard pronunciation 3]*. Seoul: National Institute of Korean Language.

Chomsky, Noam. 1975. *Reflections on Language*. NewYork, NY: Pantheon.

Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112(1). 155–159.

Constantino, John N. & Richard D. Todd. 2003. Autistic Traits in the General Population: A Twin Study. *Archives of General Psychiatry* 60(5). 524–530. doi:10.1001/archpsyc.60.5.524.

Conway, A. R., N. Cowan & M. F. Bunting. 2001. The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review* 8(2). 331–335. doi:10.3758/bf03196169.

Cristia, Alejandrina. 2001. Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of The Acoustical Society of America* 129. 3271–80.

Cronbach, Lee J. 1957. The two disciplines of scientific psychology. *American Psychologist* 12. 671–684.

Cronbach, Lee J. 1975. Beyond the two disciplines of scientific psychology. *American Psychologist* 30(2). 116–127. doi:10.1037/h0076829.

Crowne, Douglas P. & David Marlowe. 1960. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24(4). 349–354.

Dąbrowska, Ewa. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2(3). 219–253.

Davidson, Lisa, Shmico Orosco & Sheng-Fu Wang. 2021. The link between syllabic nasals and glottal stops in American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1). 1–23. doi:10.5334/labphon.224.

de Lacy, Paul. 2009. Phonological evidence. In Steve Parker (ed.), *Phonological argumentation: Essays on evidence and motivation*, 43–78. Sheffield, UK: Equinox Publications.

Do, Young Ah. 2013. *Biased learning of phonological alternations*: Massachusetts Institute of Technology dissertation.

Dollaghan, Chris & Thomas F. Campbell. 1998. Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research* 41(5). 1136–1146. doi:10.1044/jslhr.4105.1136.

Dufva, Mia, Pekka Niemi & Marinus J. M. Voeten. 2001. The role of phonological memory, word recognition, and comprehension skills in reading development: From preschool to grade 2. *Reading and Writing: An Interdisciplinary Journal* 14(1-2). 91–117. doi:10.1023/A:1008186801932.

Durvasula, Karthik. 2024. Karthik durvasula: Individual variation? `https://karthikdurvasula.gitlab.io/posts/2024-02-14-individual-variation/`.

Eckert, Penelope & Sally McConnell-Ginet. 2003. *Language and Gender*. New York: Cambridge University Press.

El Bouzaïdi Tiali, Samuel, Elsa Spinelli, Fanny Meunier, Richard Palluel-Germain & Marcela Perrone-Bertolotti. 2021. Influence of homophone processing during auditory language comprehension on executive control processes: A dual-task paradigm. *PLOS ONE* 16(7). e0254237. doi:10.1371/journal.pone.0254237.

Ellis Weismer, Susan, J Bruce Tomblin, Xuyang Zhang, Paula Buckwalter, Jan Gaura Chynoweth & Maura Jones. 2000. Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research* 43(4). 865–878.

Ettlinger, Marc, Ann R. Bradlow & Patrick C. M. Wong. 2014. Variability in the learning of complex morphophonology. *Applied Psycholinguistics* 35(4). 807–831. doi:10.1017/S0142716412000586.

Eysenck, Hans J. 1966. Personality and experimental psychology. *Bulletin of the British Psychological Society* 19. 1–28.

Eysenck, Hans J. 1997. Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology* 73(6). 1224–1237.

Festman, Julia, Antoni Rodriguez-Fornells & Thomas F. Münte. 2010. Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions* 6(1). 5. doi:10.1186/1744-9081-6-5.

Fischer, John. 1958. Social influence in the choice of a linguistic variant. *Word* 14. 47–56.

Fodor, Jerry A. 1983. *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Foulkes, Paul & Gerard Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34(4). 409–438. doi:10.1016/j.wocn.2005.08.002.

French, L. M. 2006. *Phonological working memory and L2 acquisition: A developmental study of children learning French in Québec*. Lewiston, NY: Edwin Mellen Press.

Gathercole, Susan E. & Alan D. Baddeley. 1989. Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language* 28(2). 200–213. doi:10.1016/0749-596X(89)90044-2.

Gathercole, Susan E. & Alan D. Baddeley. 1990. The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology* 81(4). 439–454. doi:10.1111/j.2044-8295.1990.tb02371.x.

Gernsbacher, Morton A., Kathleen R. Varner & Mark E. Faust. 1990. Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(3). 430–445. doi:10.1037/0278-7393.16.3.430.

Ghosh, Joyee, Yingbo Li & Robin Mitra. 2018. On the use of cauchy prior distributions for bayesian logistic regression. *Bayesian Analysis* 13(2). 359–383. doi:10.1214/17-BA1051.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the stockholm workshop on variation within optimality theory*, 111–120.

Gollan, Tamar H., Tiffany Sandoval & David P. Salmon. 2011. Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science* 22(9). 1155–1164. doi:10.1177/0956797611417002.

Gordon, Peter C., Jennifer L. Eberhardt & Jay G. Rueckl. 1993. Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology* 25(1). 1–42. doi:10.1006/cogp.1993.1001.

Grabowski, Emily & Jennifer Kuo. 2023. Comparing K-means and OPTICS clustering algorithms for identifying vowel categories. *Proceedings of the Linguistic Society of America* 8(1). 5488. doi:10.3765/plsa.v8i1.5488.

Gupta, Prahlad. 2003. Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology Section A* 56(7). 1213–1236. doi:10.1080/02724980343000071.

Ha, Seunghee. 2020. Developmental patterns of nonword repetition in preschool children. *Commun Sci Disord* 25(4). 921–928. doi:10.12963/csd.20767.

Hall-Lew, Lauren, Patrick Honeybone & James Kirby. 2021. Individuals, communities, and sound change: An introduction. *Glossa: a journal of general linguistics* 6(1). 1–17. doi: 10.5334/gjgl.1630.

Han, Chung-hye, Jeffrey Lidz & Julien Musolino. 2007. V-Raising and Grammar Competition in Korean: Evidence from Negation and Quantifier Scope. *Linguistic Inquiry* 38(1). 1–47. doi: 10.1162/ling.2007.38.1.1.

Han, Chung-hye, Julien Musolino & Jeffrey Lidz. 2016. Endogenous sources of variation in language acquisition. *Proceedings of the National Academy of Sciences* 113(4). 942–947. doi: 10.1073/pnas.1517094113.

hankwuk cengsin mwunhwa yenkwuwen [The Academy of Korean Studies]. 1995. *Hankwuk-pangencalyocip 1: Kyengkitophyen [Korean Dialectal Data 1: Gyeonggi-do].*

Hay, Jennifer. 2003. *Causes and consequences of word structure.* London: Routledge.

Hayes, Bruce. 1998. On the richness of paradigms, and the insufficiency of underlying representations in accounting for them.

Hayes, Bruce. 1999. Phonological Restructuring in Yidiɲ and its Theoretical Consequences. In Ben Hermans & Marc Oostendorp (eds.), *The Derivational Residue in Phonological Optimality Theory*, 175–205. Amsterdam: John Benjamins.

Hayes, Bruce, Kie Zuraw, Peter Siptar & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822–863.

Hoekstra, Rosa A., Meike Bartels, Danielle C. Cath & Dorret I. Boomsma. 2008. Factor Structure, Reliability and Criterion Validity of the Autism-Spectrum Quotient (AQ): A Study in Dutch

Population and Patient Groups. *Journal of Autism and Developmental Disorders* 38(8). 1555–1566. doi:10.1007/s10803-008-0538-x.

Hwang, Mina. 2015. Nonword repetition of typically developing children. *Commun Sci Disord* 20(3). 374–385. doi:10.12963/csd.15254.

James, Ariel N., Scott H. Fraundorf, Eun-Kyung Lee & Duane G. Watson. 2018. Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language* 102. 155–181. doi:10.1016/j.jml.2018.05.006.

Janda, Laura A. 2013. Quantitative methods in cognitive linguistics: An introduction. In Laura A Janda (ed.), *Cognitive linguistics: The quantitative turn*, 1–32. Berlin & Boston: De Gruyter Mouton.

Jarosz, Gaja. 2019. Computational Modeling of Phonological Learning. *Annual Review of Linguistics* 5(1). 67–90. doi:10.1146/annurev-linguistics-011718-011832.

John, Oliver P., Eileen M. Donahue & Robert L. Kentle. 1991. *The Big Five Inventory—versions 4a and 54.* Berkeley: Institute of Personality and Social Research, University of California, Berkeley.

Johnson, Keith, Peter Ladefoged & Mona Lindau. 1993. Individual differences in vowel production. *The Journal of the Acoustical Society of America* 94(2). 701–714. doi:10.1121/1.406887.

Jun, Jongho. 2010. Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* 19(2). 137–179. doi:10.1007/s10831-010-9058-8.

Jun, Jongho & Jeehyun Lee. 2007. Multiple stem-final variants in Korean native nouns and loanwords. *Eoneohag* 47. 159–187.

Jun, Sun-Ah & Jason Bishop. 2015. Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech* 58(4). 459–473. doi:10.1177/0023830914563368.

Kang, Beom-Mo & Hung-Gyu Kim. 2004. *Hankwuke hyengtayso mich ehwi sayong pintouy pwunsek2 [Frequency analysis of Korean morpheme and word usage2]*. Seoul: Institute of Korean Culture, Korea University.

Kang, Eunji, Ho-Young Lee & Juwon Kim. 2004. The phonetic realization of syllable codas in Korean. *Malsori* 49. 1–30.

Kang, Yoonjung. 2003. Sound changes affecting noun-final coronal obstruents in Korean. In William McClure (ed.), *Japanese/ Korean Linguistics 12*, 128–139.

Kang, Yoonjung. 2005. The Emergence of the Unmarked in an Analogical Change.

Kang, Yoonjung. 2007. Frequency effects and regularization in Korean noun variations.

Kapnoula, E. C., M. B. Winn, E. J. Kong, J. Edwards & B. McMurray. 2017. Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of experimental psychology. Human perception and performance* 43(9). 1594–1611.

Kassambara, Alboukadel & Fabian Mundt. 2020. *factoextra: Extract and visualize the results of multivariate data analyses*. `https://CRAN.R-project.org/package=factoextra`. R package version 1.0.7.

Kim, Hyunsoon. 2001. A phonetically based account of phonological stop assibilation. *Phonology* 18. 81–108.

Kingston, J., S. Rich, A. Shen & S. Sered. 2015. Is perception personal? In *Proceedings of the 18th International Congress of the Phonetic Sciences*, International Phonetics Association.

Ko, Eon-Suk, Jinyoung Jo, Kyung-Woon On & Byoung-Tak Zhang. 2020. Introducing the ko corpus of korean mother–child interaction. *Frontiers in Psychology* 11. doi:10.3389/fpsyg.2020.602623.

Ko, Eon-Suk & Jongho Jun. 2024. Phonological variation in child-directed speech is modulated by lexical frequency. *Journal of Child Language* 51(2). 288–313. doi:10.1017/S0305000923000466.

Ko, Kwang-Mo. 1989. Explaining the noun-final change t>s in Korean. *Eoneohag* 11.

Koenig, Laura L. 2000. Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research* 43(5). 1211–1228. doi:10.1044/jslhr.4305.1211.

Kong, Eun Jong & Jan Edwards. 2016. Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics* 59. 40–57. doi:10.1016/j.wocn.2016.08.006.

Kong, Eun Jong & Hyunjung Lee. 2018. Attentional Modulation and Individual Differences in Explaining the Changing Role of Fundamental Frequency in Korean Laryngeal Stop Perception. *Language and Speech* 61(3). 384–408. doi:10.1177/0023830917729840.

Kormos, Judit & Anna Sáfár. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11(2). 261–271. doi:10.1017/S1366728908003416.

Labov, William. 1963. The social motivation of a sound change. *Word* 19(3). 273–309.

Labov, William. 1966/2006. *The social stratification of english in new york city.* Cambridge University Press.

Labov, William. 1972. *Sociolinguistic patterns.* University of Pennsylvania press.

Labov, William. 1994. *Principles of Linguistic Change, vol. 1: Internal Factors.* Oxford, UK:: Blackwell.

Labov, William. 2001. *Principles of Linguistic Change, vol. 2: Social Factors.* Oxford, UK: Blackwell.

Lai, Wei, Lacey Wade & Meredith Tamminga. 2022. Individual differences in simultaneous perceptual compensation for coarticulatory and lexical cues. *Linguistics Vanguard* 8(1). 29–39. doi:doi:10.1515/lingvan-2020-0040.

Lee, Eun Ju & Hyun Sub Sim. 2003. Phonological Memory in the Nonword Repetition of Children: A Comparison of Functional Phonologically Disordered and Normal Children. *Communication Sciences and Disorders* 8(2). 127–145.

Lee, Michael D. & Michael R. Webb. 2005. Modeling individual differences in cognition. *Psychonomic Bulletin & Review* 12(4). 605–621. doi:10.3758/BF03196751.

Lev-Ari, Shiri. 2016. Studying individual differences in the social environment to better understand language learning and processing. *Linguistics Vanguard* 2(s1). 1–10. doi:10.1515/lingvan-2016-0015.

Lev-Ari, Shiri. 2018. The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology* 71(10). 2249–2260. doi:10.1177/1747021817739865. PMID: 30226426.

Lev-Ari, Shiri & Sharon Peperkamp. 2013. Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics* 41(5). 320–331. doi:10.1016/j.wocn.2013.06.002.

Lev-Ari, Shiri & Sharon Peperkamp. 2014. The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics* 47. 36–46. doi:10.1016/j.wocn.2014.09.001.

Liu, Huei-Mei, Patricia K. Kuhl & Feng-Ming Tsao. 2003. An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science* 6. F1–F10.

Lundström, Sebastian, Zheng Chang, Maria Råstam, Christopher Gillberg, Henrik Larsson, Henrik Anckarsäter & Paul Lichtenstein. 2012. Autism Spectrum Disorders and Autisticlike

Traits: Similar Etiology in the Extreme End and the Normal Variation. *Archives of General Psychiatry* 69(1). 46–52. doi:10.1001/archgenpsychiatry.2011.144.

Macqueen, J. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 281–297.

Mayer, Connor. 2020. An algorithm for learning phonological classes from distributional similarity. *Phonology* 37. 91–131. doi:10.1017/S0952675720000056.

Mayer, Connor, Kie Zuraw & Adeline Tan. 2024. *maxent.ot: Perform phonological analyses using maximum entropy optimality theory*. `https://CRAN.R-project.org/package=maxent.ot`. R package version 1.0.0.

Mielke, Jeff, Adam Baker & Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92(1). 101–140. doi:10.1353/lan.2016.0019.

Mount, Michael K., Murray R. Barrick, Steve M. Scullen & James Rounds. 2005. Higher-order dimensions of the big five personality traits and the big six vocational interest types. *Personnel Psychology* 58(2). 447–478. doi:10.1111/j.1744-6570.2005.00468.x.

National Institute of Korean Language. 2022. NIKL Korean Dialogue Corpus (audio) 2020(v.1.3). `https://corpus.korean.go.kr`.

Newman, Rochelle S. 1997. *Individual differences and the link between speech perception and speech production*: State University of New York at Buffalo Doctoral dissertation.

Newman, Rochelle S. 2003. Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America* 113(5). 2850–2860. doi:10.1121/1.1567280.

Newman, Rochelle S., Meredith L. Rowe & Nan Bernstein Ratner. 2016. Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language* 43(5). 1158–1173. doi:10.1017/S0305000915000446.

Nitsan, Gal, Arthur Wingfield, Limor Lavie & Boaz M Ben-David. 2019. Differences in Working Memory Capacity Affect Online Spoken Word Recognition: Evidence From Eye Movements. *Trends in Hearing* 23. 1–12. doi:10.1177/2331216519839624.

Niyogi, Partha & Berwick Robert. 1997. A dynamical systems model for language change. *Complex Systems* 11. 161–204.

Peterson, Gordon E & Harold L Barney. 1952. Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America* 24(2). 175–184.

R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. `https://www.R-project.org/`.

Reilly, Ehren M. 2007. Choosing just the right amount of over-application in Texistepec Popoluca. In Leah Bateman, Michael O'Keefe, Ehren M. Reilly & Adam Werle (eds.), *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III*, Amherst, MA: GLSA.

Renwick, Margaret E. L. & D. Robert Ladd. 2016. Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Laboratory Phonology* 7(1). 19. doi:10.5334/labphon.17.

Revelle, William, Joshua Wilt & David M. Condon. 2011. Individual differences and differential psychology: A brief history and prospect. In *The Wiley-Blackwell handbook of individual differences*. The Wiley-Blackwell Handbooks of Personality and Individual Differences., 3–38. Hoboken, NJ, US: Wiley Blackwell. doi:10.1002/9781444343120.ch1.

Rousseeuw, Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20. 53–65. doi:10.1016/0377-0427(87)90125-7.

Sampson, Geoffrey R. 2005. Quantifying the shift towards empirical methods. *International Journal of Corpus Linguistics* 10.

Sanchez, Kauyumari, Jennifer Hay & Elissa Nilson. 2015. Contextual activation of australia can affect new zealanders' vowel productions. *Journal of Phonetics* 48. 76–95. doi:https://doi.org/10.1016/j.wocn.2014.10.004. The Impact of Stylistic Diversity on Phonetic and Phonological Evidence and Modeling.

Sasisekaran, Jayanthi. 2013. Nonword repetition and nonword reading abilities in adults who do and do not stutter. *Journal of Fluency Disorders* 38(3). 275–289. doi:https://doi.org/10.1016/j.jfludis.2013.06.001.

Smolensky, Paul. 1986. Information processing in dynamical systems: Foundations of Harmony Theory. In James L. McClelland, David E. Rumelhart & PDP Research Group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, vol. 2: Psychological and biological models*, 390–431. Cambridge, MA: MIT Press.

Snyder, Cathryn, Michelle Cohn & Georgia Zellou. 2019. Individual variation in cognitive processing style predicts differences in phonetic imitation of device and human voices. In *Proceedings of Interspeech 2019*, 116–120. doi:10.21437/Interspeech.2019-2669.

Speciale, Giovanna, Nick C. Ellis & Tracey Bywater. 2004. Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics* 25(2). 293–321. doi:10.1017/S0142716404001146.

Stevens, Mary & Jonathan Harrington. 2014. The individual and the actuation of sound change. *Loquens* 1(1). e003. doi:10.3989/loquens.2014.003.

Stewart, Mary E. & Mitsuhiko Ota. 2008. Lexical effects on speech perception in individuals with "autistic" traits. *Cognition* 109(1). 157–162. doi:10.1016/j.cognition.2008.07.010.

Stroop, J. R. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18(6). 643–662. doi:10.1037/h0054651.

Sumner, Meghan, Seung Kyung Kim, Ed King & Kevin B. McGowan. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4. doi:10.3389/fpsyg.2013.01015.

Taler, Vanessa, Geoffrey P. Aaron, Lauren G. Steinmetz & David B. Pisoni. 2010. Lexical neighborhood density effects on spoken word recognition and production in healthy aging. *The Journals of Gerontology: Series B* 65B(5). 551–560. doi:10.1093/geronb/gbq039.

Theodore, Rachel M., Joanne L. Miller & David DeSteno. 2009. Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America* 125(6). 3974–3982. doi:10.1121/1.3106131.

Thorndike, Robert L. 1953. Who belongs in the family? *Psychometrika* 18(4). 267–276. doi:10.1007/BF02289263.

Ullman, Michael T. 1999. Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes* 14(1). 47–67. doi:10.1080/016909699386374.

Ullman, Michael T. 2004. Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92(1). 231–270. doi:https://doi.org/10.1016/j.cognition.2003.10.008. Towards a New Functional Anatomy of Language.

van der Schuit, Margje, Eliane Segers, Hans van Balkom & Ludo Verhoeven. 2011. How cognitive factors affect language development in children with intellectual disabilities. *Research in Developmental Disabilities* 32(5). 1884–1894. doi:10.1016/j.ridd.2011.03.015.

Wade, Lacey. 2022. Experimental evidence for expectation-driven linguistic convergence. *Language* 98(1). 63–97.

Wade, Lacey, Wei Lai & Meredith Tamminga. 2021. The Reliability of Individual Differences in VOT Imitation. *Language and Speech* 64(3). 576–593. doi:10.1177/0023830920947769.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30(5). 945–982.

Wilson Van Voorhis, Carmen R. & Betsy L. Morgan. 2007. Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology* 3(2). 43–50. doi:10.20982/tqmp.03.2.p043.

Yang, Yoonhee, Dongsun Yim, Shinyoung Kim & Jiyun Han. 2013. The relationship among receptive vocabulary, non-word repetition, and quick incidental learning in preschoolers with and without delay in vocabulary development. *Commun Sci Disord* 18(4). 379–391. doi: 10.12963/csd.13079.

Yu, Alan C. L. 2010. Perceptual Compensation Is Correlated with Individuals' "Autistic" Traits: Implications for Models of Sound Change. *PLoS ONE* 5(8). e11950. doi:10.1371/journal.pone.0011950.

Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In Alan C. L. Yu (ed.), *Origins of Sound Change: Approaches to Phonologization*, 201–227. Oxford, UK: Oxford Univ. Press.

Yu, Alan C. L. 2021. Toward an individual-difference perspective on phonologization. *Glossa: a journal of general linguistics* 6(1). 1–24. doi:10.5334/gjgl.661.

Yu, Alan C. L., Carissa Abrego-Collier & Morgan Sonderegger. 2013. Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and "Autistic" Traits. *PLoS ONE* 8(9). e74746. doi:10.1371/journal.pone.0074746.

Yu, Alan C. L., Robert McAllister, Nicholas Mularoni & Carol K. S. To. 2024. Brief Report: Atypical Temporal Sensitivity in Coarticulation in Autism: Evidence from Sibilant–Vowel Interaction in Cantonese. *Journal of Autism and Developmental Disorders* doi:10.1007/s10803-024-06258-w.

Yu, Alan C. L. & Georgia Zellou. 2019. Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics* 5(1). 131–150. doi:10.1146/annurev-linguistics-011516-033815.

Zee, Tim, Louis ten Bosch & Mirjam Ernestus. in preparation. Morphological effects in speech reduction are speaker specific and may partly originate from the words' most frequent phonological context .

Zellou, Georgia. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61. 13–29. doi:10.1016/j.wocn.2016.12.002.

Zlatin, Marsha A. 1974. Voicing contrast: perceptual and productive voice onset time characteristics of adults. *The Journal of the Acoustical Society of America* 56(3). 981–994.

Zsiga, Elizabeth & Maria Gouskova. 2006. On the status of voiced obstruents in Tswana: Against *ND. In *Proceedings of NELS*, 721–734.

Zuraw, Kie R. 2000. *Patterned exceptions in phonology*: University of California, Los Angeles Doctoral dissertation.

Zuraw, Kie R. 2013. *MAP constraints.