# The Computational Structure of Phonological and Phonetic Knowledge

A Dissertation Presented

by

**Scott Nelson**

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

**Doctor of Philosophy**

in

**Linguistics**

Stony Brook University

**2024**

**Stony Brook University**

The Graduate School

**Scott Nelson**

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

**Jeffrey Heinz — Dissertation Advisor**
**Professor, Department of Linguistics**
**and Institute for Advanced Computational Science**

**Marie Huffman — Chairperson of Defense**
**Associate Professor, Department of Linguistics**

**Ellen Broselow**
**Toll Professor, Department of Linguistics**

**Jordan Kodner**
**Assistant Professor, Department of Linguistics**

**Kristine Yu**
**Associate Professor, Department of Linguistics,**
**University of Massachusetts Amherst**

This dissertation is accepted by the Graduate School

Celia Marshik
Dean of the Graduate School

Abstract of the Dissertation

# The Computational Structure of Phonological and Phonetic Knowledge

by

**Scott Nelson**

**Doctor of Philosophy**

in

**Linguistics**

Stony Brook University

**2024**

In this dissertation, I take a computational approach to formalizing the phonetics-phonology interface that provides a baseline for understanding the role that different information plays in generating various observed phenomena. I argue that no matter where one demarcates phonetics and phonology, the computational structure of the interface is an important component that must be formally specified in order to have a robust understanding of the relationship between these two types of knowledge.

The approach itself is based on type theory and provides a way to characterize the phonetics-phonology interface as a series of interacting functions. Clarity in what type of information needs to be represented versus what type can be computed emerges from the careful specification of the structure of these functions.

Following this approach, this dissertation demonstrates that many arguments against separating phonetics and phonology rely on a specific computational structure; what is referred to as the modular feed-forward model. An alternative computational structure, the blueprint model of production, is proposed and is shown to account for various systematic phonetic phenomena that have been used to argue in favor of the combined approach.

Furthermore, this view supports the simplification of long-term memory representations. I show that the computational power required to infer dynamic gestural representations directly from symbolic strings is relatively weak. Consequently, this type of information does not need to be represented and can instead be computed on the fly.

For Benny.
It was always for you.

# Contents

**List of Figures** **vii**

**List of Tables** **ix**

**Acknowledgments** **x**

**1 Introduction** **1**
- 1.1 Outline of the dissertation . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
- 1.2 Preliminaries . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
- 1.3 Computation and the Interface . . . . . . . . . . . . . . . . . . . . . . . . 11
  - 1.3.1 The Interface as Functions . . . . . . . . . . . . . . . . . . . . . . . 11
  - 1.3.2 Interpreting Phonetic Behavior in Relation to Phonological Theory . 17
- 1.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

**2 Phonological and Phonetic Knowledge** **23**
- 2.1 The Relationship between Phonetics and Phonology . . . . . . . . . . . . . 23
- 2.2 Phenomena of Interest . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
  - 2.2.1 Incomplete Neutralization . . . . . . . . . . . . . . . . . . . . . . . 31
  - 2.2.2 Variation in Homophone Duration . . . . . . . . . . . . . . . . . . . 34
  - 2.2.3 Optionality . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
  - 2.2.4 Segmental Representation . . . . . . . . . . . . . . . . . . . . . . . 39
- 2.3 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

**3 The Blueprint Model of Production** **44**
- 3.1 From assembly line to blueprint: function (de)application . . . . . . . . . . 47
- 3.2 Simulations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 51
  - 3.2.1 Expanding the formal definition of the BMP . . . . . . . . . . . . . . 56
  - 3.2.2 Final Devoicing in German . . . . . . . . . . . . . . . . . . . . . . . 65
  - 3.2.3 Tonal Near Merger in Cantonese . . . . . . . . . . . . . . . . . . . . 69
  - 3.2.4 Epenthesis in Arabic . . . . . . . . . . . . . . . . . . . . . . . . . . 73
  - 3.2.5 Homophone Duration Variation in English . . . . . . . . . . . . . . . 76
  - 3.2.6 Optionality in Warao and English . . . . . . . . . . . . . . . . . . . 81
- 3.3 General Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
- 3.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85

**4 The Content of Phonological Representation 87**
4.1 Articulatory Phonology . . . . . 88
4.2 Model-Theoretic Phonology . . . . . 95
    4.2.1 Phonological Representations . . . . . 96
    4.2.2 Phonological Constraints . . . . . 98
    4.2.3 Phonological Transformations . . . . . 100
    4.2.4 Phonological Theory Comparison . . . . . 104
4.3 Translating Between Coupling Graphs and Segmental Strings . . . . . 108
    4.3.1 Representing Coupling Graphs . . . . . 108
    4.3.2 Translating from Strings to Coupling Graphs . . . . . 110
    4.3.3 Translating from Coupling Graphs to Strings . . . . . 123
    4.3.4 Towards a Quantifier-Free Translation . . . . . 128
4.4 Case Study: English past tense . . . . . 129
    4.4.1 Rule-based account . . . . . 130
    4.4.2 Optimality Theory-based account . . . . . 133
    4.4.3 Articulatory Phonology-based Account . . . . . 138
4.5 Summary . . . . . 142

**5 Conclusion 146**
5.1 Summary of Results . . . . . 146
5.2 Looking Ahead . . . . . 148
5.3 Final Thoughts . . . . . 153

**Bibliography 154**

**Appendix 174**

# List of Figures

1.1 The phonetics-phonology interface 12

2.1 Local and Global mappings of input string /$\rtimes$ $...\, cad...\, cad...$ $\ltimes$/ given rule $a \rightarrow b/c\_d$. $\rtimes$ and $\ltimes$ indicate left and right word boundaries. 36
2.2 Different discrete representations of a voiceless alveolar stop used in generative phonology 41

3.1 Visual comparison of the architecture for modular feed-forward models and the BMP. Each box represents a function/module. Solid lines represent the inputs to each function while dashed lines represent the outputs of each function. 47
3.2 Hypothetical Cue space 58
3.3 Simulated cue values for Port and Crawford (1989) results. The left plot shows values for /d/-final words and the right plot shows values for /t/-final words. Symbols + and × represent burst and closure duration, respectively. 67
3.4 Simulated cue values for Yu (2007) tone-sandhi data 73
3.5 Simulated formant values for lexical and epenthetic [i] vowels based on Hall (2013) for a dramatic difference speaker, a small difference speaker, and a no difference speaker 75
3.6 Average duration and log frequency for 17 words and their homophone twin. These data come from the Switchboard corpus (Godfrey et al., 1992). Dashed thin lines connect all homophonous pairs. The thick gray line is the output of a linear model ($\mathcal{I}$) of these points showing a general negative correlation. 77
3.7 Frequency and Duration information for individual tokens of 9 randomly selected homophonous pairs. Each plot represents a single pair. The solid black lines are the predicted linear relationship for that phonological form. Dashed lines indicate 95% confidence intervals. 78

4.1 Gestural score for *mash*. 91
4.2 Coupling graph for *mash*. 92
4.3 Model for the English word *dish* using the segment signature. On the left is the mathematical object itself while on the right is a visualized form of the model. 96

4.4 Model for the English word *dish* using the feature signature. On the left is the mathematical object itself while on the right is a visualized form of the model. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97
4.5 Mapping $cadab$ to $cbdab$. Dashed arrows with overlayed formula indicate which formula describes the output property of a given domain element given its input properties. . . . . . . . . . . . . . . . . . . . . . . . . . . . 104
4.6 Coupling graph model for English word *laughed*. . . . . . . . . . . . . . . . 111
4.7 String model for English word *laughed* ([læft]) . . . . . . . . . . . . . . . . . 112
4.8 Workspace generated by the copy set formula (4.21) for the string to coupling graph transduction. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
4.9 Workspace properties generated by the labeling relations for the string to coupling graph transduction of [læft]. . . . . . . . . . . . . . . . . . . . . . . 117
4.10 Workspace properties generated by the ordering relations for the string to coupling graph transduction of [læft]. . . . . . . . . . . . . . . . . . . . . . . 120
4.11 Workspace properties generated by the licensing for the string to coupling graph transduction of [læft]. . . . . . . . . . . . . . . . . . . . . . . . . . . 122
4.12 Subgraph (dashed) corresponding to the “spine” of the [læft] coupling graph model. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 125
4.13 Visualization of transduction of word “laughed” from the symbolic generative phonology model to the coupling graph articulatory phonology model 127
4.14 Sub-structure that triggers epenthesis based on rule account . . . . . . . . . 132
4.15 Sub-structure that triggers voicing assimilation/devoicing based on rule account . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133
4.16 CLO-CLO banned substructure . . . . . . . . . . . . . . . . . . . . . . . . 140

# List of Tables

3.1 Types . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 48
3.2 Data from Port and Crawford (1989) for neutralized final stops by condition. Ratio indicates the mean value of /d/ divided by the mean value of /t/. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
3.3 Possible planning groupings for the multimorphemic word *marketability* as well as the possible phonetic outputs that each grouping would predict. . . 82

4.1 Truth values for the transformation of $cadab$ to $cbdab$. . . . . . . . . . . . . 103
4.2 Ordering relations for coupling graph model . . . . . . . . . . . . . . . . . 109
4.3 Labeling relations for coupling graph model. . . . . . . . . . . . . . . . . . 109
4.4 Truth values for the gesture labels in the coupling graph interpretation of the string model for [læft]. . . . . . . . . . . . . . . . . . . . . . . . . . . 116
4.5 Example rule derivations for English past tense morpheme alternations. . . 130
4.6 Logical formulas for epenthesis rule . . . . . . . . . . . . . . . . . . . . . . 131
4.7 Logical formulas for voicing assimilation rule . . . . . . . . . . . . . . . . . 132

# Acknowledgments

This dissertation would not have been possible without the labor of countless workers at Stony Brook University. This includes maintenance, custodial, and food service staff, as well as others who I am either forgetting or were hidden from me. Their work keeps the university functioning on a daily basis, which allows me to sit and think about my research without worry. I am forever grateful that they provide the conditions that allow me to do this type of work.

Similarly, the Linguistics Department and Institute for Advanced Computational Science (IACS) staff made my academic pursuits significantly less stressful. I would like to specifically thank Sandra Brennan, Michelle Carbone, and Brian Frank from the Linguistics Department and Chelsey Dollinger, Ann Gardner, Sarena Romano, and Jennifer McCauley from IACS.

I cannot say enough about how grateful I am to have a dissertation advisor like Jeff Heinz. I met Jeff while finishing my Masters degree at Michigan State and his passion for phonological theory was immediately apparent and overwhelmingly infectious. When starting my PhD, Jeff pushed me to work on topics that I was passionate about and I hope this shows in my dissertation. Jeff is also incredibly supportive and proud of all his students. What I appreciated most about him in this regard were the unsolicited affirmations. He was always willing to remind me that what I was doing was important and share why he thought that was the case. That belief helped me greatly and I owe him so much.

The rest of my committee has been extremely helpful in shaping not only this thesis but my academic career up to this point. They deserve all the thanks in the world (and then some more). I TA'd Introduction to Phonetics with Marie Huffman three times while at Stony Brook and she taught me how to be a phonetics *scholar* and not just someone who happened to know a bit about the subject. I am so lucky to have learned about phonological theory from Ellen Broselow. Among many other things, she reinforced in me that there are many different sources of information that help us understand phonological issues. She once responded to a seminar paper I wrote with, “I appreciate the depth and seriousness with which you approach phonological issues.” I carry these words with me and hope that I can continue to live up to them. Jordan Kodner started at Stony Brook during the mid-pandemic online year which meant it took a little bit longer to get to know him. During the final two years of my dissertation, he ended up being a constant conversation partner and it was nice to talk about the Cognitive Science side of my research and, of course, the Zipfian nature of mushroom communication. During the Summer 2023 LSA Summer Institute, I had a nice conversation with Kristine Yu about the tradeoffs between depth- and breadth-based research programs. It was reaffirming talking to someone who similarly combines phonetics, phonology, and computation in their research and I was very happy when she agreed to join my committee as an external member.

I also would like to thank other researchers who have directly or indirectly influenced this thesis. Karthik Durvasula and Eric Baković have been incredibly supportive mentors and friends throughout this process. Karthik was my advisor for both my undergraduate and master’s theses at Michigan State. He shaped my early thinking about the relationship between phonetics and phonology and is still my favorite person to talk to about the subject. But beyond that, he continues to be extremely supportive and helpful with navigating all aspects of my academic career. Eric and I immediately bonded over a shared love of card games but quickly realized we shared many academic interests as well. Eric is the first person from outside my immediate sphere that I can recall taking my work seriously and it has been a real pleasure having someone like him to talk with about phonological theory. There are many other researchers whose comments or conversations shaped my thinking about this dissertation. Particular conversations with Bill Idsardi, Charles Reiss, John Kingston, Kyle Gorman, Jane Chandlee, Adam Jardine, Jason Shaw, and Giorgio Magri stick out in my mind. It has also been nice meeting, discussing research, and in many cases befriending other early career researchers interested in computation, phonology, and phonetics: Yang Wang, Caleb Belth, Huteng Dai, Canaan Breiss, Connor Mayer, Andrew Lamont.

The members of the Linguistics Department at Stony Brook also need to be given a gigantic thank you. The jovial nature of the department made my time on long island so much better than it would have been otherwise. To my cohort – Dakotah Lambert, John David Storment, Michelle Mayro, Chia-Chi Yu, and Liumao Zhong – it was nice spending this part of my life with a lovely group of people. To my officemate throughout my entire time at Stony Brook, JD, I would even say it was...*veeeeeery* nice. There are many other students and postdocs who came before and after me who helped make my time at Stony Brook so special: Neda Taherkhani, Jon Rawski, Anniello De Santo, Andrija Petrovic, Veronica Miatto, Alex Hong-Lun Yeung, Hossep Dolatian, Ayla Karakas, Magda Markowska, Vinny Czarnecki, Kenneth Hanson, Felix Fonseca Quesada, Daniel Greeson, Han Li, Salam Khalifa, Ryan Biks, Sarah Payne, Logan Swanson, Anastasiia Voznesenskaia, and Alina Shabaeva. The faculty at Stony Brook were also incredibly supportive. In no particular order I would like to thank Lori Repetti, Bob Hoberman, Mark Aaronoff, Thomas Graf, John Bailyn, Dan Finer, Jiwon Yun, Jenny Singleton, Tatiana Luchkina, Owen Rambow, and Richard Larson for teaching me how to be a linguist in various ways.

Likewise, the Institute for Advance Computational Science provided a wonderful interdisciplinary atmosphere and additional financial support that helped me with my research. It was nice having the support of and getting to know students from other departments: Khanak Bhargava, Smeet Chheda, Catherine Feldman, Hao Lin, John Murzaku, Adil Soubki, and Gaurav Verma.

I also would like to extend a thank you to the people at Michigan State who inspired me to pursue linguistics further. Karthik Durvasula, Yen-Hwei Lin, Suzanne Wagner, and Marcin Morzycki all impacted me greatly during my undergraduate and master's degrees. I would not be the researcher I am today without their foundational influences.

Many friendships made this time in my life more enjoyable and helped me get through some of the tougher periods. Secret Grief is eternal. Kam and Tyler, our studio session in January 2022 reinvigorated me at a time in my life when I desperately needed it. Derek, Cory, and Shawn, it was nice being able to come back together even if only briefly. Willow, Mike, and Kayla, playing music with you all brings me immense joy. Vinny and Magda, I know I mentioned you above already but having you both in my life as real friends means so much to me. You both have been there for me personally in many different ways and I can't thank you enough. Nalin, thanks for always staying in touch and coming to visit me in Queens! Lindsay DeWitt, Henry Klest, and Lwaxana made living on Long Island a much more enjoyable time. Emily Goodwin inspired me academically and personally more than anyone else over the last year of my life. To all my other friends, thanks for checking in, even when I'm bad at staying in touch.

Finally, I want to thank my family. To my parents, Bob and Kris Nelson, thank you for always supporting my life decisions even when you couldn't quite see the path I was on. Without your love, I would not be here today. To Alex, Sean, Catie, Makála, Leighton, Oliver, and Lyla, I love you all dearly and the beauty you bring into my life has helped me persevere through tough times.

# Chapter 1

# Introduction

The demarcation of phonetics and phonology is a persistent problem within linguistic theory (e.g., Browman and Goldstein, 1986; Ohala, 1990; Pierrehumbert, 1990; Myers, 2000; Scobbie, 2005; Hale and Reiss, 2008; Chabot, to appear). Traditionally, phonological knowledge has been viewed as abstract and discrete while phonetic knowledge has been viewed as concrete and continuous.[1] While this classification scheme has proven to be historically useful, certain types of behavioral data call into question whether or not the split can be made that simply. Some researchers have used these data to argue for the elimination of an underlying discrete computational system, while others have argued for enriching phonological knowledge with continuous representations (Browman and Goldstein, 1986; Bybee, 1999; Pierrehumbert, 2002; Port and Leary, 2005; Gahl, 2008). I will argue instead that discrete phonological knowledge can be maintained while still accounting for the data thought to be problematic. This shows that the previously stated accounts are sufficient to account for the data, but are not necessary to do so.

The specific stance I take in this thesis is that the formal structure of the phonetics-phonology interface is an important component for understanding how to differentiate the two types of knowledge. By the phonetics-phonology interface, I specifically mean the translation mechanisms between “the categories of the speaker’s message to the ut-

---

[1] See Pierrehumbert (1990) and Kingston (2019) for extended discussions of this view.

terance's articulatory continuum" and "the auditory continuum to the categories of the listener's recognition of the message's phonological content" (Kingston, 2019). Despite there being two translation mechanisms (production and perception), there is a single interface mediated externally by phonetic signals and internally by phonological representations.

Given this broad characterization, I take a *computational* approach to formalizing the interface: what are the computational properties of the various aspects of the interface? Under this approach, computation itself is the lens used to study the interface and therefore ideas from type theory and functional programming provide a new way of understanding the interface.Production can be viewed as a function from long-term memory to phonetic signal and perception can be viewed as a function from phonetic signal to long-term memory. Based on this view, the questions then become what exactly is the formal type for the long-term memory representations (*how are phonological objects represented?*) and how are the production and perception functions structured (*how is phonological knowledge used?*). In Chapter 3, I answer the second question by discussing the production function from a type-theoretic perspective while in Chapter 4 I answer the first question by using declarative logical programs to show how phonetically rich representations can be reduced to symbolic strings.

This dissertation therefore makes the following contributions. First, it introduces the Blueprint Model of Production in Chapter 3 which is an alternative framework for understanding the integration of phonological and phonetic knowledge. Specifically, I use it to show that it resolves some issues for discrete models of phonological knowledge. While phenomena like incomplete neutralization and variation in homophone duration are thought to be problematic for discrete theories of phonology, I show that the source of the problem may reside in the structure of the phonetics-phonology interface and not in the discreteness of the phonological knowledge itself. That is, when the modular feed-forward model is replaced by the blueprint model of production, these phenomena can be explained while maintaining discrete phonological knowledge. My argument is made abstractly using typed functions but supported concretely with simulations that show exactly how discrete phonological knowledge can lead to systematic gradient phonetic phenomena. This is important because it means the type of evidence that has previously been used to argue against gradient phonological knowledge is not enough to definitively argue against a discrete/categorical theory of phonology.

Second, in Chapter 4, I provide a declarative account of the relationship between segmental representations of lexical items and gestural representations of lexical items. This is done lawfully with model theory and first-order logic. I show exactly how the two types of representations can be translated back and forth without loss of information. Since these translations are bi-directional, insights from each theory are accessible to the other. For example, a phenomenon like the C-center effect (Browman and Goldstein, 1988) can be derived from standard segmental strings and does not need to be stored directly. In the other direction, Articulatory Phonology has struggled with accounting for categorical alternations. The logical translations provided in this chapter also provide a way to translate symbolic phonological processes into ones that operate over gestural representations. I show how this can be done using English past tense morphophonological alternations as a case study. This type of analysis can therefore be used for understanding how to encode categorical alternations within the Articulatory Phonology framework. That being said, the translations provide a notion of equivalency within first-order logic which means the more phonetically rich gestural representations can be inferred from classic segmental sequences and do not need to be directly stored in the mental lexicon.

This dissertation therefore provides a novel formal characterization of the phonetics-phonology interface that highlights its computational properties. Much work on the phonetics-phonology interface describes specific representation structures and specific algorithms, but the goal of this dissertation is to provide a more general framework for thinking about the interface. Furthermore, by being specific about how different types of information interact during language use, it ultimately clarifies what types of patterns the phonological grammar needs to account for. Ultimately, I argue for a phonologically-based phonetics where phonology is central to phonetic implementation and perception but operates autonomously from functional factors.

## 1.1 Outline of the dissertation

The remainder of the dissertation puts into action the computational approach discussed above. The next section of this chapter includes some mathematical preliminaries due to the formal nature of the approach taken. Chapter 1 then ends with a discussion of specific ways in which computation can inform our understanding of the phonetics-phonology interface. Particularly, I discuss how formal computational modeling provides the necessary foundation for interpreting behavioral data as it relates to phonological theory. Much of the discussion here revolves around abstraction and type theory. While this type of theorizing is removed from detailed phonetic measurements, it is nonetheless important to provide a high-level control structure that makes clear how different sources of knowledge lead to the specific phenomena and measurements that have been documented. This chapter therefore provides the philosophical base on which the rest of the dissertation is built.

Chapter 2 provides a background on different views of the relationship between phonological and phonetic knowledge and discusses the empirical phenomena that will be used as case studies in the later chapters. The discussion primarily revolves around issues of discreteness. The main throughline of this dissertation is that discrete phonological knowledge can be maintained while still accounting for systematic continuous phenomena that have been argued to be problematic for discrete phonological theories. Incomplete Neutralization (Port et al., 1981; Port and O'Dell, 1985) is used to argue specifically against a "digital foundation for a formal system of language," suggesting that phonological *computation* could not be discrete (Port and Leary, 2005). Variation in Homophone duration (Gahl, 2008) is similarly used to suggest that phonological *representation* could not be discrete. In this chapter, I lay out the phonetic facts for these phenomena and more to then show in later chapters how the argument against discrete phonological knowledge relies on the modular feed-forward structure of the phonetics-phonology interface. If this assumption about the structure of the interface is changed, the phenomena thought to be problematic for discrete phonological knowledge suddenly become straightforwardly analyzable.

Chapter 3 introduces the blueprint model of production, an alternative to the modular feed-forward model, and provides specific case studies on how discrete phonological knowledge can still account for phonetic data thought to be unaccountable by discrete phonological knowledge. Here I do two things. First, I provide a formal argument for the high-level structure of the interface using function types. The core tenet of generative phonology is that phonology is a function from underlying to surface forms. I show how this tenet can be maintained while simultaneously providing lexical information to the production function by making the phonological grammar an argument to production rather than an intermediary between the lexicon and phonetic production. This subtle move is what allows for discrete phonological knowledge to account for incomplete neutralization and variation in homophone duration. I support my formal argument with simulations that show how discrete knowledge generates continuous behavior in final devoicing in German (Port and Crawford, 1989), tonal alternations in Cantonese (Yu, 2007), and epenthesis in Lebanese Arabic (Gouskova and Hall, 2009; Hall, 2013). I similarly simulate variation in homophone duration (Gahl, 2008) and situate it within the architecture of the blueprint model of production. This chapter provides both abstract and concrete support for an alternative view of the phonetics-phonology interface and highlights the importance of formally clarifying the structure of the interface when interpreting phonetic behavior as it relates to phonological knowledge.

Chapter 4 discusses the nature of phonological segmental representation and shows how symbolic string representations can be considered notational variants of the coupling graph representations used in Articulatory Phonology. This is done by defining what are essentially declarative computer programs based on first-order logic. The logical approach uses model-theoretic structures as a source for defining a lawful relationship between different representation schemes. Coupling graphs being first-order definable from strings provides a notion of equivalency: any distinction that can be made with a string of symbols can be made with a coupling graph and vice versa within the bounds of first-order logic. The consequence for the interface here is that detailed articulatory plans can be inferred directly from discrete representations using a fairly restricted computational process. This further allows for direct theory comparison in terms of phonological processes. In this chapter, I also provide logical characterizations of rule, constraint, and gestural accounts of the English past tense alternation and find that they all require essentially the same computation. One advantage of the logical approach is that it provides a way to translate the structural requirements of alternations from symbolic representations into dynamic coupling graphs. Accounting for these types of phenomena has traditionally been not straightforward within Articulatory Phonology. This approach therefore provides a jumping-off point for seeing what the analyses of these phenomena would have to look like within the dynamical framework.

Chapter 5 concludes, focusing on other directions this type of work can be taken. This includes expanding the approach in this dissertation to determine the structure of the perception function. More concretely it suggests ways in which the ideas argued for in this dissertation can be used to think about other areas like sound change, loanword adaptation, sociolinguistic variation, and learning. In all of these cases, having a computationally formalized theory of the phonetics-phonology interface will provide the necessary conditions to explore phenomena where both low-level and high-level information of sound systems play a role. By being explicit about the theory, we will be better equipped to propose, understand, and interpret empirical studies related to these areas.

## 1.2 Preliminaries

This section provides some basic mathematical preliminaries underlying the computational aspects of the dissertation. The first three parts (sets, relations, functions) will be most helpful for Section 1.3 at the end of this chapter and all of Chapter 3. The last two parts (logic, models) will be most helpful for Chapter 4. Readers familiar with these topics may skip this section.

### Sets

A **set** is a well-defined collection of objects. Objects in this set are called *members* or *elements*. To say that $a$ is an element of the set $A$, we say $a \in A$ ($a$ is in $A$) and to say that it is not an element, we say $a \notin A$ ($a$ is not in $A$). The set with no elements $\varnothing$ is called the **empty set**. The **union** of two sets $A$ and $B$, denoted $A \cup B$, is the set containing all elements that are either in $A$ or in $B$. The **intersection** of two sets $A$ and $B$, denoted $A \cap B$, is the set containing all elements that are only in both $A$ and $B$. Two sets $A$ and $B$ are called **disjoint** if they have no elements in common. Namely, if it is the case that $A \cap B = \varnothing$. The **set difference** of two sets $A$ and $B$ is denoted by $A \setminus B$ (sometimes $A - B$) and represents the set of all elements that are in $A$ but not in $B$.

A set $B$ is called a **subset** of a set $A$ if every element of $B$ is also an element of $A$. This is denoted by $B \subseteq A$. If every element of $B$ is also an element of $A$ and $B \neq A$ then we say $B$ is a **proper subset** of $A$. This is denoted by $B \subset A$. A set $A$ is called a **superset** of a set $B$ if every element of $B$ is also an element of $A$. This is denoted by $A \supseteq B$. If every element of $B$ is also an element of $A$ and $B \neq A$ then we say $A$ is a **proper superset** of $B$. This is denoted by $A \supset B$.

The **Cartesian product** of two sets $A, B$ is $A \times B = \{(a, b) \mid a \in A,\ b \in B\}$, which denotes the set of all ordered pairs where the first element is in $A$ and the second element is in $B$. More generally, the **n-fold Cartesian product** of $n$ sets is $X_1 \times X_2 \times \cdots \times X_n = \{(x_1, x_2 \ldots, x_n) \mid x_1 \in X_1,\ x_2 \in X_2, \ldots,\ x_n \in X_n\}$, which denotes the set of all ordered tuples where the $n$-th element is in $X_n$. The **power set** of a set $A$ is $\mathcal{P}(A) = \{B \mid B \subseteq A\}$, which denotes the set containing all of the subsets of $A$ (including the empty set).

### Relations

A **relation** with arity $n$ is a subset of the Cartesian product $X_1 \times \cdots \times X_n$. A **binary relation** $R$ between sets $X$ and $Y$ is therefore an element of $\mathcal{P}(X \times Y)$. This is denoted as $R \subseteq X \times Y$ and may also be written as either $xRy$ or $R(x, y)$ to mean $(x, y) \in R$. There are certain properties that a relation $R$ may hold over a set $X$: **reflexive** - for all $x \in X$, $xRx$; **irreflexive** - for all $x \in X$, it is not the case that $xRx$; **symmetric** - for all $x, y \in X$, if $xRy$ then $yRx$; **antisymmetric** - for all $x, y \in X$, if $xRy$ and $yRx$ then $x = y$; **asymmetric** - for all $x, y \in X$, if $xRy$ then it is not the case that $yRx$; **transitive** - for all $x, y, z \in X$, if $xRy$ and $yRz$ then $xRz$; **connected** - for all $x, y \in X$, if $x \neq y$ then $xRy$ or $yRx$.

### Functions

A **function** is a special type of relation $f \subseteq X \times Y$ such that $\forall x, y, y'.((x, y) \in f \wedge (x, y') \in f) \rightarrow y = y'$. More colloquially, a function is a relation where one element in a set $X$ can be related to *at most one other element* in a set $Y$. A function $f$ is **undefined** for an element $x \in X$ if there is no pair $(x', y) \in f$ such that $x = x'$, otherwise $f$ is **defined** for $x$. When $f$ is defined for $x$ we write $f(x) = y$ or sometimes $x \mapsto_f y$. Equivalently, a function $f$ may be referred as a *map* and written as $f$ maps $x$ to $y$. If $\forall x \in X.(x, y) \in f$ then we say that $f$ is a **total function**, otherwise $f$ is a **partial function**.

The set $X$ is called the **domain** and the set $Y$ is called the **co-domain** (or range) of the function $f : X \rightarrow Y$. Relatedly, the **image** of $f$ is the set $\{f(x) \in Y \mid x \in X \wedge f(x) \text{ is defined}\}$ and the **pre-image** of $f$ is the set $\{x \in X \mid f(x) \text{ is defined}\}$. A function

$f : X \to Y$ is **injective** or *one-to-one* if for all $y \in Y$, the pre-image $f^{-1}(y)$ has at most one element. A function $f : X \to Y$ is **surjective** or *onto* if for all $y \in Y$, the pre-image $f^{-1}(y)$ is non-empty. A function $f : X \to Y$ is **bijective** if it is both injective and surjective.

Given two functions $f : X \to Y$ and $g : Y \to Z$ such that the domain of $g$ is the co-domain of $f$, we can define the **composition** of $f$ and $g$ as a single function $g \circ f : X \to Z$. this is equivalent to first applying $f$ to an input $x$ and then applying the output of $f$ to $g$, or $g(f(x))$. Function composition is *associative*: $h \circ g \circ f = h \circ (g \circ f) = (h \circ g) \circ f$. It is not necessarily *commutative*: $g \circ f$ is not guaranteed to be the same as $f \circ g$.

### Logic

A **proposition** is a statement that is either true or false, but cannot be both. Logics are built from two main components: a *syntax* and a *semantics*. The syntax defines the form that terms of the language take, and the semantics defines how those forms are interpreted. In the syntax of a logical language, there are atoms which build up to make larger elements called **well formed formulas** (or wffs). In propositional logic, these atoms are propositions $P, Q, R, \ldots$. The syntax for propositional logic is defined as follows. Suppose $\phi, \psi$ are atomic formulas, then $\phi$ is a well formed formula (wff), $\neg\phi$ is a wff, and $(\phi \wedge \psi)$ is a wff. Common logical connectives such as $\vee$, $\to$, and $\leftrightarrow$ can all be defined from these three statements.

While the syntax provides a way to combine elements, an **assignment** or **interpretation** function is required to give the propositions meaning. Let $\mathbb{P}$ be the set of all propositions and let $i : \mathbb{P} \to \{0, 1\}$ be a function from propositions to truth values that assigns a truth value to a given proposition. If $P \in \mathbb{P}$ is a proposition, then $i(P) = 1$ means $P$ is true and $i(P) = 0$ means $P$ is false. A binary **satisfaction** relation $\vDash$ between truth valuations and formulas specifies when a formula defined by the syntax is true or false. The expression $i \vDash \phi$ means that $\phi$ is true under $i$, or $i$ models/satisfies $\phi$. The expression $i \nvDash \phi$ means that $\phi$ is false under $i$, or $i$ does not model/satisfy $\phi$.

First-order logic extends Propositional Logic in the following ways. First, **first-order variables** $x, y, z$ are added which range over elements of a given domain. Second, **quantification** is added which provides a way to infer properties that range over the domain. There are two quantifiers that are used: the universal quantifier ($\forall$) and the existential quantifier ($\exists$). The universal quantifier is often read as "for all" and requires a property to hold for *all* elements in the domain. The existential quantifier is often read as "there exists" and requires a property to hold for *at least one* element in the domain. Third, **equality** ($=$) is added which provides a way to determine whether two variables identify the same element in the domain.

### Models

A **model signature** $\Sigma$ is a collection of symbols for the functions, relations, and constants that are used to describe structures. A **$\Sigma$-structure** $A$ contains a set called the **domain**, as well as **denotations** for each symbol in $\Sigma$. A denotation for a constant is a single element of the domain of $A$. A denotation for a function symbol of arity $k$ is a $k$-ary function on the domain of $A$. A denotation for a relation symbol of arity $k$ is a $k$-ary relation on the domain of $A$.

A **logical language** in first-order logic is defined by combining the first-order symbols with a specific model signature $\Sigma$. A **$\Sigma$-formula** is any logical formula where all the non-logical symbols are drawn from $\Sigma$. A $\Sigma$-formula is **satisfiable** if it evaluates to `True` under *some* assignment of the variables in the domain. A $\Sigma$-formula is **valid** if it evaluates to `True` under *every* assignment of the variables in the domain. A $\Sigma$-formula like $Qx(\phi)$ (where $Q$ is a quantifier, $x$ a variable, and $\phi$ a well-formed formula) is called a **$\Sigma$-sentence** given that $\phi$ contains no variables other than $x$. We say that $\phi$ is in the scope of $Qx$ and any instance of $x$ in $\phi$ are **bound** by it. Any variables in $\phi$ other than $x$ are said to be **free**. The full definition of a $\Sigma$-sentence is any $\Sigma$-formula with no free variables.

Suppose we have a $\Sigma$-structure $A$ and a $\Sigma$-sentence $\phi$. If $\phi$ evaluates to `True` then we write $A \vDash \phi$ to mean $A$ satisfies/models $\phi$. If $\phi$ evaluates to `False` then we write $A \nvDash \phi$ and say $A$ does not satisfy/model $\phi$. For a fixed $\Sigma$-structure $A$, the set of $\Sigma$-sentences that it satisfies is the **theory** of $A$: $\{\phi \mid A \vDash \phi\}$.[2] For a fixed $\Sigma$-sentence $\phi$, the set of $\Sigma$-structures that satisfy it is the **spectrum** of $\phi$: $\{A \mid A \vDash \phi\}$. For a fixed set of $\Sigma$-sentences $T$, the set of $\Sigma$-structures that satisfy *all* $\Sigma$-sentences in $T$ is the **model** of $T$: $\{A \mid \forall \phi \in T[A \vDash \phi]\}$.

## 1.3 Computation and the Interface

This section contains two parts. The first part explains the computational view of the interface in the form of function types. This leads to formal clarity for the types of questions that are asked and answered throughout the dissertation. The second part of this section provides a philosophical foundation for why the abstract computational approach is taken in this dissertation. Ultimately, formal clarity at the theoretical level can lead to novel interpretations of old results while also providing new pathways for further experimentation.

### 1.3.1 The Interface as Functions

While Kingston's (2019) definition of the phonetics-phonology interface uses articulatory and auditory continua to describe endpoints of the interfaces, they are both lawfully related to an acoustic continuum. With this in mind, the phonetics-phonology interface can be described as the bi-directional relationship between the mental phonological representations and the physical acoustic speech signal. The title of Halle's (2003) collection of papers, *From memory to speech and back*, arguably best captures this relationship. The translation from memory to speech is the production process, the translation from speech back to memory is the perception process, and memory is the phonological representation of individual morphemes stored in the lexicon. This entire process is what will be

---

[2] These definitions are taken from Bhaskar (2023).

referred to as the phonetics-phonology interface throughout the dissertation. Figure 1.1 below displays the interface in visual form.

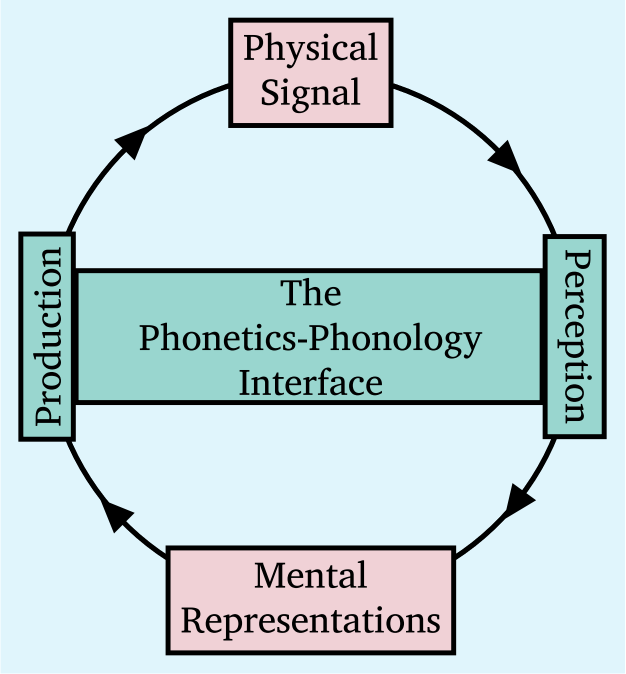


Figure 1.1: The phonetics-phonology interface

Given this formulation of the interface, three core questions must be answered to provide a full characterization. The first question is what is the nature of the phonological representations stored in long-term memory? These representations act as both the starting point for the production process and the ending point for the perception process and therefore provide a mental link between the two. Furthermore, the structure of the representational space has consequences for the computations involved in production and perception. Under many approaches, phonological representations do not contain all of the information present in the phonetic signal. Consequently, claims about the interface are ultimately claims about data expansion and data compression. The second question is what is the nature of the computation from phonological representations to phonetic signals? In other words, how is the production process characterized such that an impoverished phonological representation can be expanded into a rich phonetic signal? The third and final question is what is the nature of the computation from phonetic signal to phonological representations? This is a question about the characterization of the perception process and how a rich phonetic signal can be compressed into an impoverished phonological representation. One could also imagine a fourth question that characterizes the phonetic signal, but this will be assumed to be the measurable acoustic signal and therefore not pursued further. Before expanding on these three main questions, it is necessary to say more about the other aspect of the dissertation: computation.

The term computation in this dissertation refers explicitly to both the theory of computation as well as Marr's (1982) computational level of analysis. The theory of computation is a subfield of computer science and mathematics that studies what and how things can be computed using abstract computing devices (Hopcroft et al., 2006; Sipser, 2013; Sannella et al., 2021). Marr (1982), in his work on vision, proposed three levels for understanding any machine that carries out an information processing task. At the highest level of abstraction is the computational level which describes the goal of a computation, why it is appropriate, and what the logic and strategy for carrying it out is. At lower levels of abstraction are the representation and algorithmic level which describes how the computation is implemented and what types of representations are used to do the computation, and the implementation level which describes how the process is physically implemented.

The perspective taken in this dissertation relies on the function type each module computes. This notation follows from Pierce (2002) which derives from the lambda calculus (Church, 1932, 1933). What follows is a basic introduction to function types. A function maps one or more elements in a set $A$ to elements in a set $B$ such that each $a$ in $A$ maps to at most one element in $b$ in $B$. For a function $f$ that maps elements from set $A$ to set $B$ we write $f :: A \rightarrow B$. A phonology function (or $P$ for short) could therefore be written as $P :: UR \rightarrow SR$. In prose this means "the phonology function $P$ maps URs to SRs." Note the phonology function $P$ is agnostic as to the particulars of the representations of $UR$ and $SR$. For example, they could be continuous, discrete, or some combination.

$P :: UR \rightarrow SR$ simply means that the phonology module takes a UR-type thing and returns an SR-type thing.

Functions with more than one argument are written similarly. Addition can be thought of as a function with two arguments: $add(x)(y) = x + y$. Its function type would then be written as: $add :: (\mathbb{R} \rightarrow (\mathbb{R} \rightarrow \mathbb{R}))$. In the remainder of the dissertation, the brackets will be removed and therefore when reading function types with multiple arrows, one way of thinking about it is that everything to the left of the rightmost arrow is an argument and everything to the right of the rightmost (non-bracketed) arrow is the output. The function type of $add$ can therefore be understood as a map from two real numbers to a single real number.

Two other concepts will be useful for understanding the current approach to the interface: higher-order functions and the notion of function application. Functions like the ones described above are first-order functions. These are contrasted with higher-order functions. A higher-order function is a function that either takes as an input another function or returns a function as its output. An example of a higher-order function that takes a function as part of its input is the $map$ function.

Given two inputs $f$ and $\vec{x}$, where $f$ is a function of type $f :: X \rightarrow Y$ that takes things of type $X$ as its input, and $\vec{x}$ is an array of length $n$ that contains $x$'s $[x_1, \ldots, x_n]$, $map(f)(\vec{x})$ applies function $f$ to every individual element of $x \in \vec{x}$ and returns the array $[f(x_1), \ldots, f(x_n)]$. To give a concrete example, consider the function $add1(x) = x + 1$ and the array of integers $[-23, 1, 9, 307]$. If we were to provide both of these as the input to the $map$ function, we would end up with $map(add1, [-23, 1, 9, 307]) = [-22, 2, 10, 308]$. The $map$ function is not limited to numerical data types/functions and works just as well over strings. For example, for all strings $w$, let $redup(w) = ww$. Then $map(redup, [a, ba, cab]) = [aa, baba, cabcab]$. To summarize, the function type of $map$ is given by $map :: (X \rightarrow Y) \rightarrow [X] \rightarrow [Y]$.

This brings us to a discussion of *function application*. Function application is the act of applying a specific function to an argument, but it can also be thought of as a higher-order function itself. The two arguments for function application would be one of type $X$ and the other of type $X \to Y$ (i.e. a function that maps $X$ type things to $Y$ type things). Given these two arguments, it would output something of type $Y$. For the overall type we would therefore write $function\text{-}application :: X \to (X \to Y) \to Y$. The notion of function application is important for the analysis in Chapter 3 because it provides a way to relate a classic view of the interface to the new proposal outlined in this dissertation.

Typed functions relate directly to Marr's (1982) computational level because the function type partially describes the goal of a computation. For example, a typed function of the form $A :: B \to C$ can be read as "$A$ is a function that maps $B$ type things to $C$ type things. Relatedly, the goal of any information processing device ($A$) is to take some information ($B$) and process it into something new ($C$).[3] It becomes more interesting when $B$ gets decomposed into its composite parts. Since the other aspect of Marr's (1982) computational level is describing the logic and strategy for carrying out the computation, it is necessary to describe how all of the different types of information that make up $B$ work together towards the outcome $C$.

It is now possible to show how the notation of addition discussed above relates to the common $add(x, y) = x + y$ notation using the notion of *currying*. To explain, consider the fact that since functions in general can return functions (to be discussed further shortly), functions with multiple arguments do not need to be given all the arguments at once. If fewer than the totality of arguments is given, then a *function* is returned. The type for addition at this point is: $add :: \mathbb{R} \to \mathbb{R} \to \mathbb{R}$. This can be thought of as the uncurried version of $add' :: (\mathbb{R}, \mathbb{R}) \to \mathbb{R}$. Whereas $add$ requires two arguments to be fully saturated,

---

[3] The use of "goal" here does invoke certain teleological explanations. Marr uses "goal" in his original formulation of the computation level, but a neutral reading of the term could place it as being synonymous with "final state" in the sense that there is a specific type of information that will be output given that the final state or "goal" has been reached. In some ways, this is quite similar to Aristotle's explanation of *natural teleology* with the classic example being the goal of an acorn is to grow into an oak tree, but this is an intrinsic property of acorns and not necessarily a property of how acorns should be used. In the same way, the goal of an information process task may be to output information simply because that is intrinsically what it has evolved to do and not to satisfy any higher-level goals or properties.

$add'$ takes a single argument which is a pair of real numbers. It is always possible to convert between a function that takes one input as a pair of arguments and a higher-order function that takes multiple arguments. This conversion is called *currying* (Curry, 1980). Currying itself can be thought of as a higher-order function, which takes an uncurried function like $add$ and returns the curried version like $add'$. The type signature of currying is thus $curry :: ((A, B) \rightarrow C) \rightarrow (A \rightarrow B \rightarrow C)$. The argument of the $curry$ function is a function mapping $(a, b)$ pairs to $c$-type things. The output of the $curry$ function is a function that takes two separate inputs $a$ and $b$ and outputs $c$. Consequently, $curry(add') = add$, and thus for all $a, b$, $add(a, b) = add(a)(b) = a + b$.

Returning now to the characterization of the phonetics-phonology interface, the three questions from above can be restated. Using the type analysis, suppose that phonological representations have type $PR$ and the phonetic signal has type $PS$. The production process can now be characterized as an output function $O$ with type $O :: PR \rightarrow PS$ and the perception process can now be characterized as an input function $I$ with type $I :: PS \rightarrow PR$. The three questions required to characterize the phonetics-phonology interface can now be formally stated as:

1. What is the characterization of the *Phonological Representations* $PR$?
2. What is the characterization of the *Output* function: $O :: PR \rightarrow PS$? [Production]
3. What is the characterization of the *Input* function: $I :: PS \rightarrow PR$? [Perception]

While questions 2 & 3 are illustrated using typed functions, the Marrian computation level analysis of the interface requires more work. Not only are the types necessary, but one must also explain the actual computation being computed and the general strategy for doing said computation. Furthermore, both the *Output* and *Input* functions have internal structure that is of interest. While phonological *representations* are one aspect of the phonetics-phonology interface, there is also the phonological *grammar* that must be accounted for. The hypothesis underlying generative phonology is that the phonological grammar is a function that maps underlying phonological representations which are stored in the mind to surface phonological representations that are then further transformed into a phonetic representation (Heinz, 2018). Understanding how the phonological grammar, which is often thought to be a discrete symbol manipulator, interacts with other factors in the production and perception processes is key to properly characterizing the interface. In Chapter 3, I provide a production architecture, the blueprint model of production, which does exactly that.

### 1.3.2 Interpreting Phonetic Behavior in Relation to Phonological Theory

This section expands on the computational approach used throughout the dissertation and explains why an abstract approach is ultimately beneficial for understanding more concrete phenomena. Computational linguistics has a rich history in both the theoretical and applied domains, and early work in generative syntax helped spawn formal language theory (Chomsky, 1956). In this sense, there is nothing novel about using a computational approach for understanding natural language phenomena. But recently there has been an additional push outside of linguistics for using computational description and modeling to formalize verbal theories in the psychological sciences (van Rooij and Baggio, 2020; van Rooij and Blokpoel, 2020; van Rooij and Baggio, 2021; Guest and Martin, 2021). One of the main arguments given is that the computational approach helps eliminate ambiguities and forces any theoretical commitments to be stated clearly. A second argument is that formal clarity at the theoretical level leads to better predictions, explanations, and ideas at the empirical level. In Chapters 3 and 4, a recurring theme is that arguments against discrete phonological knowledge based on experimental results can stem from a lack of formal rigor when discussing the shape of the formal grammar and the structure of the interface.[4]

Understanding the relationship between phonetics and phonology crucially relies on experimental data that helps illustrate the way speech is produced and perceived in real time. These data are often used to decide between competing phonological theories. While formal experimentation is an important tool within the modern phonologist's toolbox, there are three arguments for why experimental results (like any data) should be interpreted with caution when it comes to phonological theory. First, for a causal explanation to hold, experimental results must follow from phonological theory and not vice versa (Guest and Martin, 2023). Second, any mismatch between experimental results and phonological theory can be caused by incorrect phonological theory, but it could equally be the result of some other auxiliary assumptions, such as implicit or even explicit claims about the phonetics-phonology interface (Quine, 1951; Duhem, 1954). This point is especially relevant for the analysis and discussion in Chapter 3. Third, one needs to distinguish between accounts of the data and explanations of the data (Cummins, 2000). Support for these arguments is drawn from formal logic below.

For falsifiability to hold when doing causal inference, the data must be a consequence of the theory and not vice versa. This idea for phonology can be formalized following recent discussions of artificial neural networks as models of general (Guest and Martin, 2023) and linguistic (Rawski and Baumont, 2023; Kodner et al., 2023) cognition, but abstracted here to *any* model of a cognitive process. Consider the following propositions: our model of phonology does what humans do ($P$) and our model is correlated with experimental data ($D$). Since the theory must imply the correlation with the data ($P \rightarrow D$), it would be infelicitous to affirm the consequent and say that because our model of phonology is correlated with experimental data then our model of phonology does what humans do. Instead, this logical formulation allows for inference following Modus Tollens: the model not being correlated with experimental data ($\neg D$) supports our model of phonol-

---

[4] For a similar idea applied to the concept of the phoneme in speech perception see Kazanina et al. (2018).

ogy not doing what humans do ($\neg P$). Taken together this means that experimental data cannot be used to support one phonological theory over another, but it can be used to falsify a specific theory. This must be the case since there are often multiple extensionally equivalent but different intensional descriptions in any domain of inquiry, including phonology (Chao, 1934).

Furthermore, experimental results crucially rely on the interaction of *competence* (phonological knowledge) and *performance* (phonetic manifestation, functional considerations, frequency, etc.); therefore, the above Modus Tollens inference gets us into trouble without an additional proposition: our model(s) of other performance factors that affect behavior do what humans do ($F$). The inference that can be made about phonology, these additional performance factors, and experimental data can be summarized in the following way.

(1) $(P \overset{\text{and}}{\wedge} F) \rightarrow D$ Modus Pollens: $(P \overset{\text{and}}{\wedge} F) \overset{\text{asserts}}{\vdash} D$ Modus Tollens: $\neg D \overset{\text{asserts}}{\vdash} (\neg P \overset{\text{or}}{\vee} \neg F)$

The Modus Pollens inference allows for prediction: given our theories of phonology and other factors are correct, then the data should look a certain way. The Modus Tollens inference tells us that if the model is not correlated with experimental data ($\neg D$) then it's unclear if this is due to our model of phonology $P$ being false or due to our model(s) of other factors influencing behavior $F$ being false. This relates to the Duhem-Quine thesis which states that when experimentally testing a scientific hypothesis, it is impossible to know if the reason the results don't match the theory is because the theory is wrong or if some unspoken auxiliary assumption is wrong (Duhem, 1954; Quine, 1951). Put another way, Lakatos (1970) claims that theory comparison is truly possible only within a framework with a common set of auxiliary assumptions. It is therefore crucial when interpreting experimental results to be explicit about both the phonological theory being tested, as well as all pertinent factors that may affect behavior while data is being captured. Minimally, this should include a theory of phonetic realization of phonological structure. In Chapter 3, this point is made explicitly in relation to phenomena like incomplete neutralization and variation in homophone duration.

Third, some researchers have proposed a distinction between explanations and non-explanatory accounts of psychological phenomena (Cummins, 1983, 2000). The latter are re-descriptions of the data, while the former are true understandings of the underlying causal structure. For example, Cummins (1983) discusses the pendulum law which tells us that the period of pendulum $T$ is equal to $2\pi$ times the square root of the pendulum's length $l$ divided by the constant of gravitation $g$. From this law, we can determine the period if we know the length, but we can also determine the length if we know the period. This highlights that the law itself simply describes the relationship but does not give a causal explanation. For this reason, it is an account but not an explanation.

Du and Durvasula (2024b) point out that crucial to the distinction between explanations and accounts is that a true explanation of a phenomenon $Q$ would be one where $\neg Q$ (that is the observation of an event that excludes $Q$) cannot be accounted for by the theory. Therefore, if the existence of both $Q$ and $\neg Q$ can be handled by the theory, then the theory does not provide an explanation for the phenomenon $Q$. In contrast, a non-explanatory account could accommodate both $Q$ and $\neg Q$. Furthermore, as has been known for a long time, a sufficiently elaborate model can always account for any data by essentially overfitting the data; therefore, any account that can accommodate the phenomenon $Q$ and $\neg Q$ are simply re-descriptions of the data. While the previous arguments focused on interpretations of the experimental data, this argument shows that it is also important to consider the implications of the theory before even doing any experimentation. As argued above, this can be done through mathematical formalization which is part of what van Rooij and Baggio (2021) call the "theoretical cycle" and is a way to ensure that the theory is *a priori* viable and doing the things the theorist believes it should do. Additionally, Guest and Martin (2021) point out that theories that are defined formally and explored computationally, are more robust to failures of replication at the experimental level.

## 1.4 Summary

This introductory chapter has done the following. First, it outlined the main question of interest: what is the formal relationship between phonological and phonetic knowledge? It then provided an overview of how this question would be pursued. The two main chapters that provide novel answers are Chapter 3 which focuses on *computational* knowledge and Chapter 4 which focuses on *representational* knowledge. Second, it provided useful mathematical preliminaries that provide a foundation for understanding the technical aspects of the dissertation. Third, it gave an overview of type theory and how it helps us understand the interface. Finally, it discusses the role of auxiliary assumptions in interpreting phonetic phenomena in relation to phonological theory.

The final two aspects from above provide a philosophical foundation for the abstract approach taken throughout the dissertation. As discussed in these sections, the computational approach is argued to provide a more robust understanding of the relationship between phonological and phonetic knowledge by providing a top-down approach to interpreting experimental data. Furthermore, the computational approach provides various types of support for not rejecting discrete phonological knowledge.

In Chapter 3, this is accomplished by showing that the specific computational structure of the production function determines how we interpret phonetic data in relation to phonological theory. I introduce the blueprint model of production which is a type-theoretic characterization of the interface and use simulations to show how adopting this architecture provides the ability to account for systematic continuous phonetic phenomena while maintaining discrete phonological knowledge.

In Chapter 4, this is accomplished by showing a type of computational equivalency between more phonetically-rich representations and symbolic strings. I use first-order logic and model-theoretic representations to define declarative programs which provide a mathematical foundation for this claim. Consequently, the types of distinctions that can be made are identical within first-order logic. This means that the phonetic information does not need to be stored directly in the representation but can instead be inferred. The analyses in Chapters 3 and 4 use computation and formalization to provide important insights for understanding the phonetics-phonology interface that are difficult to obtain with other types of methods.

# Chapter 2

# Phonological and Phonetic Knowledge

## 2.1 The Relationship between Phonetics and Phonology

While discussion of the relationship between phonetics and phonology predates *The Sound Pattern of English* (*SPE*; Chomsky and Halle, 1968), *SPE* is a natural starting point for the current discussion. In *SPE* it is assumed that the phonology contains rules that map binary features to a scalar value so that the surface representation (SR) of a lexical item is a temporally organized matrix of real numbers corresponding to phonetic features. The phonological grammar therefore contains rules that are both discrete and continuous. It is not explicitly stated whether or not both types of rules interact. Additionally, *SPE* assumes that there is a phonetics module that acts as a universal translator, turning the phonetic SR outputs into physical representations.

Keating (1985, 1988) discusses the *SPE* model of speech production further, pointing out that the assumption of a universal phonetics is likely to be incorrect. A main area of focus in her discussion is the tradeoff between enriching the phonological representation with phonetic detail versus having a less phonetically rich SR with language-specific phonetic implementation rules. Keating proposes that the grammar contains both phonological and phonetic rules. Kingston and Diehl (1994) argue that speakers use language-specific phonetic knowledge to alter their articulations to enhance phonological contrasts based on f0 depression around [+voice] segments. This knowledge is implemented outside of the phonological module. Keating (1990) similarly assumes that there are language-specific phonetic rules, but for her, there is phonetic information both inside and outside the phonological module.

It is also possible to consider whether or not we need two separate cognitive modules for phonology and phonetics. A strong argument against separating the two comes in the form of Port and Leary's (2005) paper titled *Against Formal Phonology*. They argue that a discrete formal symbolic system is unable to account for the variability in phonetic realization of identical symbols as well as certain temporal-based contrasts in behavioral data. Since these formal systems cannot simulate the natural language data on their own, Port and Leary (2005) argue against having a formal phonological grammar at all. Ohala (1990) takes a softer approach. He recognizes the different types of analysis being done within each domain, but argues that one cannot do phonology without phonetics and one cannot do phonetics without phonology. For him, the two are intertwined, and therefore viewing them as completely separate domains "is artificial and unnecessarily complicates the study of speech" (p. 156).

Two formal proposals that dissolve the distinction between phonetics and phonology are Flemming's (2001) unified model of phonetics and phonology and Browman and Goldstein's theory of Articulatory Phonology (1992, *et seq.*). Flemming (2001) develops a weighted constraint grammar that operates over scalar phonetic representations. He argues that phonological assimilation and phonetic coarticulation are essentially the same type of phenomena only with different grain sizes. What is considered to be phonetic coarticulation is just a fine-grained version of the more coarse-grained phonological assimilation (and vice versa). The representations in Flemming's model is therefore rich with physical phonetic structure such as formant values (Hz) and duration (ms).

Articulatory Phonology (AP; Browman and Goldstein, 1992) operates under the assumption that phonetics and phonology are just low and high level descriptions of the same dynamical system. At the high level of description, the basic phonological units in AP are gestures. Gestures are task-specific goals and therefore defined as the creation of a certain sized constriction in the vocal tract. For example, the word [ta] would be described as a tongue tip gesture that touches the alveolar ridge, a glottal spreading gesture (the default state of the glottis in AP is such that voicing occurs), and a wide tongue body gesture. The tongue tip and glottal gestures would occur in time with one another while the tongue body gesture would be timed to occur after the other two gestures.

At the low level of description, each gesture is represented as a second-order dynamical equation and implemented in the task-dynamic model of Saltzman and Munhall (1989). In the task dynamic model, each gesture competes for control of certain articulators while the gesture is active. Since the goal of a gesture is only to create a certain constriction type, the path the articulators take to create a specific constriction is largely dependent on the other gestures simultaneously activated within the dynamical system. For example, the back of the tongue will have to move differently for a velar stop produced after a front vowel than for one produced after a back vowel since the former will pull the back of the tongue slightly forward. This slightly more front position means the actual path to make the velar closure will be different than the path taken in the back vowel scenario. From an AP perspective, both phonological and phonetic processes are the lawful consequence of interacting gestures within a dynamical system. While AP is a specific theory that uses dynamical systems, their use more broadly has been successful in describing various interface phenomena (Tuller et al., 1994; Gafos, 2006; Gafos and Benus, 2006; Gafos et al., 2014; Roon and Gafos, 2016; Łukaszewicz, 2021, among others). Tuller et al. (1994) use dynamical systems to explain behavior in various speech categorization tasks while Roon and Gafos (2016) use them to show how perception and production are linked in response planning which can explain behavior in a response-distractor task.

If we reject the previously discussed accounts and instead favor distinct phonological and phonetic modules, then we are left with deciding where the demarcation point between the two lies. In other words, what exactly is a phonological process and what exactly is a phonetic process? The development of generative phonology coincided with a time when theories of cognition largely involved the manipulation of discrete, symbolic representations (*e.g.,* Newell and Simon, 1958). Despite *SPE*'s transformation of features into scalar values, it has largely been assumed that phonological processes are discrete since the representations are discrete and that gradience is the result of phonetic processes. This point of view is expressed throughout the literature. For example, Kingston (2019) points to various experimental studies that provide diagnostics for deciding whether a process is phonological or phonetic, all of which involve determining whether or not the process is gradient (Cohn, 1993, 2007; Myers, 2000; Solé, 1992, 1995, 2007).

If gradience is to be the dividing line between phonetics and phonology, there should be a consensus on what type of gradience counts. Gradience has been used in multiple ways when talking about phonology. One way it has been used is in relation to the productivity of phonological generalizations (Albright and Hayes, 2006; Ernestus, 2011). A second way regards grammatical acceptability judgments (Coleman and Pierrehumbert, 1997; Coetzee and Pater, 2008). A third way is in relation to representations (Smolensky and Goldrick, 2016; Lionnet, 2017).

Beyond deciding which type of phonological gradience applies to the phonetics-phonology interface, Pierrehumbert (1990, p. 379) points out a logical conundrum for this approach which is that "any continuous variation can be approximated with arbitrary precision by a sufficiently large set of discrete elements." Consequently, gradience on its own cannot determine whether or not a process is phonetic or phonological.

Gradience notwithstanding, some researchers are more content with interleaving phonetics and phonology. This point of view is represented in the collection *Phonetically Based Phonology* (Hayes et al., 2004). The chapters in this book present constraint-based phonological grammars that are either directly inspired by phonetic facts, or, in some cases, directly contain phonetic information. As an example of the latter, Zhang (2004) defines a set of constraints in service of providing a grammatical explanation for why contour tones are typically restricted to long vowels or stressed syllables. This includes a series of constraints which he calls *DUR($\tau_i$) that are defined such that for all segments in the rhyme, their cumulative duration in excess of the minimum duration in the prosodic environment in question cannot be $\tau_i$ or more. He further stipulates that if $\tau_i > \tau_j$, then *DUR($\tau_i$) $\gg$ *DUR($\tau_j$). The representations therefore must be structured in a way that includes real durational values and not just categorical approximations such as "long" or "short."

In a separate chapter, Gordon (2004) discusses the influence of phonetic properties on phonological syllable weight. Rather than encoding phonetic information directly into the grammar, Gordon showed how phonetic properties of a language could predict weight criteria for tones and syllabic templates. Unlike Zhang's analysis, Gordon retains categorical phonological representations. These examples show a wide range of views are available when discussing a phonetically based phonology. At one end there is phonetics *in* phonology while at the other end, there is something like phonetics *influencing* phonology. Due to this diversity, and unlike Flemming (2001), the essays in this collection are less explicit about the architecture of the grammar, but by using representations and constraints that are phonetic in nature, the lines between where phonology ends and phonetics begins are blurred.

In sharp contrast, the substance free phonology framework (Hale and Reiss, 2000, 2008; Reiss, 2018) demarcates a firm boundary between phonology and phonetics. A core tenet of this framework is that phonological computations should not be based on notions such as phonetic naturalness, typological frequency, and markedness. Instead, phonology should be viewed as a symbol manipulator that has one simple goal: to transform the phonological representation according to the rules of the language. For example, maintaining voicing at the end of a phrase has been shown to be difficult due to anatomical and aerodynamical reasons (Ohala, 1983; Westbury and Keating, 1986). A theory of phonology based on notions of markedness or phonetic naturalness would encode this directly into the grammar with a constraint against voiced obstruents in final position. Hale and Reiss (pp. 154–156; 2008) argue that this becomes especially problematic if the constraint set is universal and propose the following thought experiment: imagine in the future, the vocal tract of humans evolves in a way such that it is no longer difficult to maintain voicing at the end of phrases but instead is difficult to not maintain voicing at the end of phrases. It would then be phonetically natural to have a process of final voicing, but the grammar already has a universal constraint against final voiced segments because at a previous time, they were difficult.

If phonology is completely divorced from such substantive concerns, then one may wonder what connection it has to speech at all. A series of recent papers have clarified that it is only the phonological computations that are devoid of any substantive influence, but the phonological representations still have phonetic correlates (Volenec and Reiss, 2017; Reiss and Volenec, 2020). Volenec and Reiss (2017) adopt the fairly standard view that phonological representations are made up of binary feature bundles but highlight the fact that since they are viewing phonology as an encapsulated cognitive module (Fodor, 1983), its input and output should be made up of the same type of representations. Therefore, for them, the underlying and surface representations must both be binary phonological feature bundles. It is only through a separate *transduction* that any type of phonetic representation can be established. They posit a transducer which they refer to as "Cognitive Phonetics" which translates the output of phonology (an SR) into a phonetic representation (PR). The PR is "a complex array of neural commands that activate muscles involved in speech production" (p. 270), and feeds the sensorimotor system directly. Furthermore, the Cognitive Phonetics transducer is said to be universal which recalls *SPE*'s universal translator.

Backgrounded in the previous discussion is the (often tacit) assumption about the flow of information between phonetics and phonology. The standard view on the architecture of the grammar in generative phonology is what Pierrehumbert (2002) refers to as a "modular feed-forward" architecture.[1] This type of model is understood as a kind of abstract assembly line: a lexical item is chosen and then is modified through a series of specialized stations until it reaches the endpoint as a phonetic object that can be pronounced. Since assembly lines are successive, each station is essentially blind to the history of the objects it receives.

To make this metaphor more concrete, we can imagine that the Lexicon places Underlying Representations (URs) on a conveyor belt which takes them to the Phonology station to be worked on. At the Phonology station, URs are transformed into Surface Representations (SRs) and SRs are placed back on the conveyor belt to be taken down the line to the Phonetics station. The Phonetics station receives each SR with no knowledge of its previous history. The role of Phonetics in this instance is to transform each SR into a corresponding phonetic form (e.g., a gradient representation containing acoustic/articulatory instructions). In this example, Phonology acts as an intermediary between the Lexicon and Phonetics. Consequently, when two identical SRs are derived from distinct URs, the Phonetics station must treat those SRs exactly the same way. The modular feed-forward model leads to various issues, but as will be shown in later chapters, this structure of the interface is a separate assumption from the assumption about a discrete phonological grammar operating over discrete phonological representations (*cf.* discussion of Modus Tollens in Section 1.3).

---

[1] See also Bermúdez-Otero (2007) and Kenstowicz (2010).

## 2.2 Phenomena of Interest

The next three sections discuss empirical phenomena that have been argued to be problematic for theories of language production based on discrete and categorical generative models of phonology: incomplete neutralization (Port et al., 1981; Port and O'Dell, 1985), homophone durational variation (Gahl, 2008) and optionality/variability (Alderete and Finley, 2023). Port and Leary (2005) use incomplete neutralization as an example of why a discrete computational system (i.e., phonological grammar) is untenable. Gahl (2008) uses variation in homophone duration as an example of why discretized representations of lexical items fail to account for subtle phonetic phenomena. Optionality has also required rethinking the discreteness of the phonological grammar, but this is a consequence of how one thinks about the phonetics-phonology interface and the role of phonological knowledge in language use. I argue that these phenomena are not counterarguments to discrete phonological knowledge. This is due to the fact that the structure of the interface is itself an analytical assumption that needs to be carefully weighed when discussing the interaction of grammatical and extra-grammatical information in language production. In Chapter 3, I formally show how it is the modular feed-forward view of the interface, and not the discreteness of phonological knowledge, that makes phenomena like incomplete neutralization and variation in homophone a problem.

As discussed in Chapter 1, many philosophers of science have pointed out that refutation of a given scientific theory is dependent on auxiliary assumptions and shared background knowledge (Quine, 1951; Duhem, 1954; Popper, 1959; Feyerabend, 1965; Lakatos, 1970). Therefore, phonetic evidence alone does not bear on the nature of phonological knowledge but rather must be evaluated in tandem with a theory of how phonological knowledge is physically manifested. In other words, phenomena like incomplete neutralization and variation in homophone duration falsify discrete phonological knowledge provided we assume that the modular feed-forward structure of the interface is a shared assumption (or shared "interpretative theory" in terms of Lakatos (1970)). In this way, I aim to show that arguments for gradient phonological knowledge depend on a certain structure of the production function, but there are alternative ways to structure this function that do not require gradient phonological knowledge to account for the same phonetic facts.

This approach also results in further consequences for the understanding of how segmental information is represented in long-term memory. In many ways, this can be a difficult problem since the representational primitives influence what the production should look like, but in turn, what the production function looks like influences the encoding of the representational primitives. This mutually informing process highlights why McCarthy's (1988) quote, "if the representations are right, then the rules will follow," is unfortunately overly optimistic. While his quote was in reference to strictly the phonological grammar, it can be lifted to discussions of the interface of phonological representations and phonetic implementation. This section therefore ends, with a discussion of different ways that generative phonologists have argued for encoding segmental information in long-term memory as this is a key component for understanding the relationship between phonological and phonetic knowledge.

### 2.2.1 Incomplete Neutralization

Final devoicing is probably the most well studied example of a phonological neutralization process. This is a phenomenon where, at the end of some domain (often syllable or word), an obstruent loses its voicing feature and surfaces as a voiceless segment.[2] It has been attested in a variety of languages including, but not limited to, German (Bloomfield, 1933), Polish (Gussmann, 2007), Catalan (Wheeler, 2005), Russian (Coats and Harshenin, 1971), and Turkish (Kopkalli, 1994). The data in (2) provide an example from German (Dinnsen and Garcia-Zamor, 1971).

---

[2] In this section I assume a binary [voice] feature but recognize that more specific laryngeal representations have been proposed (Halle and Stevens, 1971; Iverson and Salmons, 1995; Avery and Idsardi, 2001).

(2) a. /bad + en/ → [baden] ‘to bathe’
b. /bad/ → [bat] ‘bath’
c. bat + en/ → [baten] ‘asked’
d. /bat/ → [bat] ‘ask’

In the 1980s, it was discovered that German speakers could discriminate between an underlying voiceless segment and a derived voiceless segment at a rate of 60-70%; further acoustic studies showed that these two types of segments systematically varied along certain acoustic dimensions (Port et al., 1981; Port and O’Dell, 1985). Acoustically, it was found that the preceding vowel was shorter for underlying voiceless segments, the duration of aspiration noise was longer for underlying voiceless segments, and the amount of voicing into stop closure was longer for underlying voiced segments. These properties make it appear as if the surface form maintained some of the properties of the underlying form. Because the values for the derived voiceless segments were intermediate between a surface voiceless segment derived from an underlying voiceless segment and a surface voiced segment in non-coda position, this phenomenon was termed “incomplete neutralization”.

Since these pioneering studies, final devoicing has been extensively studied in the lab, and found to be incomplete in additional languages such as Catalan (Dinnsen and Charles-Luce, 1984), Dutch (Warner et al., 2004), Polish (Slowiaczek and Dinnsen, 1985), Russian (Dmitrieva et al., 2010; Shrager, 2012), and Afrikaans (van Rooy et al., 2003). Additionally, the empirical coverage of incomplete neutralization has expanded to include the following processes: tapping/flapping in American English (Fox and Terbeek, 1977; Herd et al., 2010; Braver, 2014), intrusive stops in English (Fourakis and Port, 1986), coda aspiration in Andalusian Spanish (Gerfen, 2002; Bishop, 2007), coda liquids in Puerto Rican Spanish (Simonet et al., 2008; Beaton, 2016), French schwa deletion (Fougeron and Steriade, 1997), Russian voicing assimilation (Burton and Robblee, 1997), epenthesis in Lebanese Arabic (Gouskova and Hall, 2009), tone sandhi in Cantonese (Yu, 2007), Mandarin (Peng, 2000) and Fuzhou (Li, 2016), Japanese monomoraic lengthening (Braver and Kawahara, 2016), and word-final lenition in Chilean Spanish (Bolyanatz, 2020). Strycharczuk (2019) provides a recent review of findings and discusses various hypotheses for the sources of incompleteness.

Returning to final devoicing, Port and Crawford (1989) find that listeners appear to have control over the level of incompleteness of the neutralization based on communicative context and the salience of a given contrast. In their experiment, they used five different contexts (based on 4 sentence conditions) to evaluate how the level of neutralization changed depending on speakers' awareness of the task. Condition 1A/B were disguised sentences where the target word was embedded within a sentence. The 1A task involved participants reading a sentence from a written example. The 1B task used the same sentences, but this time participants were read the sentence and asked to repeat it back out loud to the experimenter. Condition 2 used contrastive sentences where both target words were in the same sentence, but clarifying information was included to differentiate the words. Condition 3 also used contrastive sentences, but removed the clarifying information. Condition 4 was the words in isolation.

They found incomplete neutralization in every condition when analyzing aggregated speaker data. No difference in the amount of incomplete neutralization was detected between conditions 1A and 1B in contrast to previous experiments (Jassem and Richter, 1989). In all other cases reported in Port and Crawford (1989), the level of incompleteness increased when the task highlighted the contrast between the two target words. Condition 2 was more incomplete than Conditions 1A/B and Condition 3 was even more incomplete than Condition 2. This makes sense because Condition 2 highlights the contrast, but includes extra material that can aid in distinguishing between the two words. Therefore, speakers may attempt to highlight the contrast with the amount of "voicing". Condition 3 meanwhile highlights the contrast, but provides no additional information. In this condition, speakers must use the amount of "voicing" to make sure the contrast is salient. Condition 4 also showed a greater amount of incompleteness than Condition 1A/B and was slightly lower than Condition 2.

These data support the idea that speakers have some level of control over how neutralized a segment is depending on the contrastive condition. The pragmatic conditions therefore influence a speaker's intent on maintaining an underlying contrast. In their nonlinear dynamic approach to production, Gafos and Benus (2006) include a variable called *intent* to account for this fact. For the remainder of this dissertation, the term *intent* will be used as a coverall term indicating pragmatic condition/desire to maintain an underlying contrast.

Incomplete neutralization is a problem for the modular feed-forward model of the interface because it predicts that two items with the same phonological surface form should have the same phonetic properties. One option is to say that the two phonological surface forms are distinct which is what van Oostendorp (2008) does using turbid representations (Goldrick, 2000). Braver (2019) points out a problem with this account, but see Nelson and Heinz (in press) for more discussion about how turbid representations smuggle underlying representation information into the surface form. In Chapter 3, I show that data like incomplete neutralization are not inherently problems for discrete/categorical phonological grammars, but only become problematic when assuming the modular feed-forward model of production.

### 2.2.2 Variation in Homophone Duration

Incomplete neutralization is a scenario where various lexical items have identical surface forms but phonologically distinct underlying forms. In these cases, the variation between underlying and surface forms allows for interpolation between the two. I now turn my attention towards a different scenario: homophones. It has been reported that many homophonic pairs have subtle phonetic differences, most notably along the temporal dimension (Walsh and Parker, 1983; Losiewicz, 1995; Gahl, 2008; Lohmann, 2018a,b). Like neutralized pairs, homophones share the same surface phonological form, but unlike neutralized pairs, there is no guarantee that they have diverging underlying forms.

Frequency has long been known to play a role in the phonetic realization of phonological units (Fosler-Lussier and Morgan, 1998; Bybee, 2001; Jurafsky et al., 2001; Bell et al., 2009). Leslau (1969) reports that the Arab Grammarians noted that more frequent words become “weaker”. Another dimension that can play a role in this phenomenon is part of speech. For example, words like “road (n)” and “rode (v)” have been found to vary in their pronunciation (Bell et al., 2009). Gahl (2008) looked at non-function word homophone pairs such as “time (n)” and “thyme (n)” and found that there was a difference in duration that correlated with the frequency of the lemma. This implicates lexical frequencies in production. Based on these findings, Gahl (2008) rejects discrete, symbolic lexical representations and instead argues for an exemplar-based organization of the grammar.

As was the case for incomplete neutralization, variation in homophone duration is a problem for the modular feed-forward model of the interface because phonologically identical surface forms, especially those with the same grammatical category, should have identical phonetic properties. These data will also be discussed in Chapter 3 as they relate to grammatical phonological knowledge and the structure of the interface. Again, the problem will be shown to be the structure of the interface and not the discreteness of the phonological knowledge.

### 2.2.3 Optionality

Optionality and variability are largely synonymous terms used when phonological knowledge does not apply under all conditions. Within this space, there are two types of phonological optionality: *free variation* occurs when a single phonological input has multiple outputs and *lexical variation* occurs when a phonological process or constraint applies to some, but not all, lexical items. The primary focus here will be free variation, but the ideas presented can be extended to lexical variation. For general overviews, as well as alternative accounts of these phenomena, see Anttila (2007) and Zuraw (2010).

Free variation is further divided into processes that optionally apply globally versus processes that optionally apply locally. The distinction between global and local optionality requires there to be more than one locus point for a given phonological process to occur. Suppose there is a process that can be described by the phonological rule $a \rightarrow b/c\_d$. Figure 2.1 shows the distinction between the two types of optionality. Globally optional processes require the process to occur at all possible points of application or none while locally optional processes allow processes to optionally occur at each possible point. Therefore, the set of possible outputs for a globally optional process will always be of cardinality $2$ while the set of possible outputs for a locally optional process will always be $2^n$ where $n$ is the number of possible points of application.

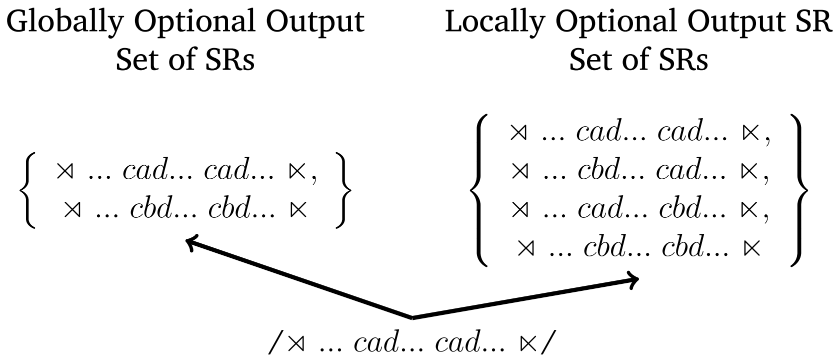


Figure 2.1: Local and Global mappings of input string $/\rtimes\ ...\ cad...\ cad...\ \ltimes/$ given rule $a \rightarrow b/c\_d$. $\rtimes$ and $\ltimes$ indicate left and right word boundaries.

Global optionality in Warao is reported by Osborn (1966). Labial voicing is a context free process which voices labial stops. Examples (3–4) provide specific surface forms where there is free variation between the all-or-nothing cases (a) but the cases where there is a mismatch in application are not observed (b).

(3) *weak*

a. [paroparera] ∼ [**b**aro**b**arera]

b. * [paro**b**arera] ∼ * [**b**aroparera]

(4) *he will put them*

a. [apaupute] ∼ [a**b**au**b**ute]

b. * [apau**b**ute] ∼ * [a**b**aupute]

Local optionality in English is reported by Vaux (2008). Tapping occurs in between two vowels when the second vowel is unstressed. Vaux reports it as being locally optional in his dialect and uses the word *marketability*, which has two points of application, as a working example. The examples in (5) show the four different ways in which the word surfaces.[3]

(5) a. [mɑɹkətəbɪləti] c. [mɑɹkətəbɪləɾi]
b. [mɑɹkəɾəbɪləti] d. [mɑɹkəɾəbɪləɾi]

A full review of the literature on grammatical accounts of optionality is beyond the scope of this dissertation. I instead focus on two broad approaches: diacritic rule-based accounts and probabilistic constraint-based accounts. Vaux (2008) discusses a diacritic approach to optionality with rule-based grammars. Each rule in the grammar is marked as [±`optional`] and [±`iterative`]. This generates four types of rules. Non-optional rules are marked [−`optional`] and can either be iterative or non-iterative. Optional rules can similarly be iterative or non-iterative where the difference corresponds to local vs. global optionality. When a given rule is marked [+`optional`] and [−`iterative`] it simply means that the optionality itself is non-iterative: if the rule applies to the first locus of application it must then apply to all following loci (or vice-versa). This generates a globally optional pattern. On the other hand, when a given rule is marked [+`optional`] and [+`iterative`], the optional application of the rule is determined at each potential point of application. This generates a locally optional pattern.

One advantage of this approach is its descriptive power from a small set of well-motivated assumptions. Since optional and iterative phonological processes are well documented, any rule-based grammar must parameterize individual rules along both dimensions, and the interaction of these two parameters follows naturally. What's missing from this account of optionality is an underlying causal mechanism. Variable rules, as

---

[3] These transcriptions deviate slightly from Vaux's (2008) transcription. He includes aspiration on the voiceless alveolar stops which I remove here since these unstressed positions usually do not contain aspiration and do not alter the larger point being made.

developed by variationist sociolinguists, are one way in which the underlying cause of optionality has been formally addressed (Cedergren and Sankoff, 1974; Sankoff and Labov, 1979). Here, the likelihood of a given rule occurring is tied to the presence or absence of certain social and grammatical factors. As Sankoff and Labov (1979, §4) discuss, there are probabilistic extensions of formal grammars, and therefore generative grammar could easily be extended along this dimension. Despite this formal relationship, generative phonologists of the rule-based persuasion have largely been hesitant to include probabilities in the grammar for various epistemological commitments (e.g., Reiss, 2017).

On the other hand, constraint-based approaches to phonology have largely transitioned towards probabilistic grammars. Hayes (2016, 2022) provides a general overview of three common approaches. Maximum entropy grammars (Hayes et al., 2009) create a probability distribution over the candidate set based on the calculation of harmony scores derived from weighted constraints. Stochastic Optimality Theory (Boersma, 1998) creates a probability distribution by giving each constraint a numerical value and allowing the constraints themselves to be adjusted by a slight bit of noise each time they are evaluated.[4] This leads to different rankings when two constraint weights are sufficiently close to one another. Noisy Harmonic Grammar (Boersma and Pater, 2016) comes in many varieties based on where the noise is added: constraint weights, violations, tableau cells, and harmony scores can all be made noisy. In all three cases, optionality emerges from the quantitative constraint model. Furthermore, sociolinguistic insights have been directly added to these types of models (Coetzee, 2016), resulting in a blurring of what type of information can be encoded directly into the phonological grammar.

The need to use the phonological grammar to explain optionality stems from viewing the output of phonology as wholly determining the phonetic form. Again, the modular feed-forward model more or less forces this view. The alternative view of production that will be developed in Chapter 3 offers a different approach to optionality. Under this

---

[4] Noise is modeled as a random variable drawn from a normal distribution with a mean of zero.

approach phonological knowledge itself is never optional: it exists as one type of information that is consulted during language use. Optionality therefore shifts from being a property of the grammar to choosing to apply phonological knowledge when computing the phonetic exponent of a given lexical item. While the distinction is subtle it accomplishes two things: first, it aligns with the view of grammatical knowledge as one of many factors that can be consulted during language use, and second, it simplifies the phonological grammar by eliminating extra mechanisms such as diacritics and probabilities. As Chomsky (1965) writes, "...it is perhaps worthwhile to reiterate that a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language by a speaker-hearer." The base is knowledge of a phonological process while use determines whether or not the knowledge is applied.

The ideas on optionality expressed in this dissertation are complementary to the claim by Tilsen (2023, 267) that "we should disprefer models of phonological cognition that explicitly resort to cognitive representations of probability." He contrasts what he refers to as *shallow stochasticity*, which occurs when probabilities are claimed to be directly represented and used in a cognitive system, with *deep stochasticity*, which occurs when probabilities are used to account for "our ignorance of the detailed dynamics" governing the system. Encoding probabilities directly into a phonological grammar would therefore count as an example of shallow stochasticity. What I will show is that, given a specific architecture of the phonetics-phonology interface, encoding these probabilities directly into the grammar is not necessary.

### 2.2.4 Segmental Representation

The primitive elements in generative phonology are typically not segments, but instead sub-segmental properties that group together in various ways. Typically, these sub-segmental properties are called *features* though the terms *particle* and *element* are also used by some phonologists (e.g., Anderson and Jones, 1974; Schane, 1984; Kaye, 1985; Harris and Lindsey, 2014). In all cases, segments emerge through the combination of specific substructures. Three broad approaches have been taken for ordering these subproperties within generative phonology. These are discussed in brief in the following paragraphs.

Chomsky and Halle (1968), following the precedent set by Jakobson et al. (1951), represent phonological segments as fully specified binary feature bundles. Every rule in their formalism therefore operates over feature values, though occasionally segments are used as a shorthand. Over time, generative phonologists have proposed alternative representations. Some of those representations maintain the feature bundle structure, but do not require binarity or for every feature to be specified (e.g., Steriade (1987); Archangeli (1988); Dresher (2009)).

There have also been attempts to provide further structure to subsegmental properties in order to account for long distance phenomena as well as dependencies between different features. These come in the form of autosegmental graphs (Goldsmith, 1976) and feature geometries (Clements, 1985; Sagey, 1986; McCarthy, 1988; Browman and Goldstein, 1990; Halle, 1995; Avery and Idsardi, 2001). Autosegmental graphs assume a timing tier where each element is represented generically as “X” or as a consonant “C” or vowel “V”. Features and other subsegmental properties are then attached to each timing element. This creates a tier for each property and allows for what appear to be long distance processes on the timing tier to appear local on a subsegmental property tier.

Feature geometries also assume a timing tier, but rather than having each feature connect to the root, they are structured to contain sub-dependencies. For example, the root node may split into a laryngeal node and a place node, each with potentially further fine-grained structure. One primary motivator for feature geometric representations was to explain different types of spreading processes (see McCarthy (1988) for an overview). Figure 2.2 shows a segmental, feature bundle, autosegmental, and feature geometric representation of a voiceless alveolar stop.

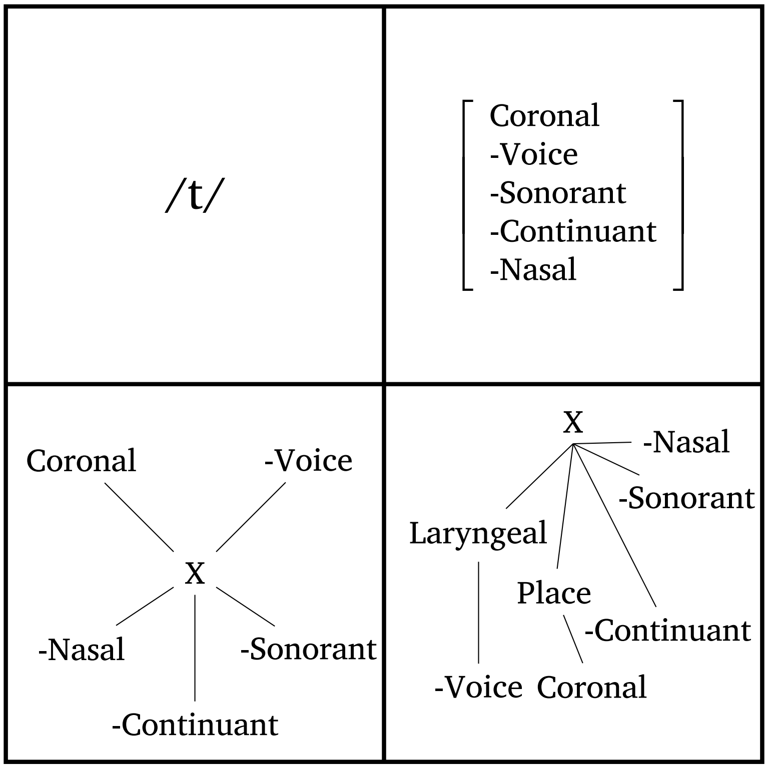


Figure 2.2: Different discrete representations of a voiceless alveolar stop used in generative phonology

In all of these representations, phonetic substance largely enters indirectly: the features might be tied to a phonetic dimension, but it requires further interpretation to go from a valued phonological feature to a specific phonetic value. If, instead, phonetic substance was thought to be directly encoded in long-term memory representations, there are various approaches that can and have been taken. One option would be to replace course-grained phonological features with scalar or continuous fine-grained phonetic features (Flemming, 2001). Relatedly, early exemplar models involve storing detailed acoustic/auditory representations of each token of each lexical item a person hears (Johnson et al., 1997; Bybee, 2001; Pierrehumbert, 2002). This approach radically rejects the discretized feature-based approach, though see Pierrehumbert (2016) for a proposal on how to unify the two. One other option is to take the Articulatory Phonology approach and encode the articulatory plan directly in the lexicon (Browman and Goldstein, 1992, 1995) by making the primitive units gestures. Again, here there are no features and no discrete representations per se, but the coupling graph representation being used in modern implementations of Articulatory Phonology (Nam, 2007a) essentially act as a de facto discrete representation. This will be discussed further in Section 4.1.[5]

The Articulatory Phonology approach does not require a transformative phonetics module that operates over the output of phonology since the articulatory plan is encoded directly into the long-term representation. Additionally, traditional segments do not exist in the theory. What look like segments are instead specific structural relations between different gestures.

## 2.3 Summary

This chapter discussed the relationship between phonological and phonetic knowledge within the generative tradition. The modular feed-forward model was set up as the standard view that is often tied directly to generative theories of phonology. Various phenomena were discussed in relation to the modular feed-forward model and how they have been used to make varying claims about phonological knowledge and its relationship to phonetic interpretation. For example, the phenomenon of incomplete neutralization, whereby phonologically neutralized segments have slightly different phonetic properties, poses a problem for discrete computational systems under the modular feed-forward view. This is because the output of phonology in both cases is identical, predicting identical phonetic forms. Similarly, variation in homophone duration once again highlights the mismatch between the prediction of the modular feed-forward model (two identical phonological forms should have identical phonetic properties), and the phonetic reality. Unlike incomplete neutralization, there is no morphophonological alternation to possibly explain variation and therefore this phenomenon poses a problem for discrete phono-

---

[5] Figure 4.2 provides an example of a coupling graph for the word *mash*.

logical representations. Optionality and variability were also discussed as they pose additional problems for discrete phonological computation: if the phonological grammar must directly provide multiple output forms for a single input form, then this cannot be done with a single output deterministic model such as the modular feed-forward model.

In the next chapter, I introduce the BLUEPRINT MODEL OF PRODUCTION, which is an alternative to the modular feed-forward model. By reconceptualizing the way different types of information are integrated at the phonetics-phonology interface, many of the problems highlighted in this chapter disappear.

In Chapter 4, I define declarative programs using first-order logic to show the lawful relationship between string representations of phonological structure and gesture-based representations used in Articulatory Phonology. The consequence for the interface here is that detailed articulatory plans can be inferred directly from discrete representations using a fairly restricted computational process. This further allows for direct theory comparison and it will be argued that some of the other tenets of the gestural approach can be explained by the production model discussed in Chapter 3.

Articulatory Phonology is also typically described as providing a unified explanation of both high-level planning in the form of the static equation parameters and low-level implementation in how the system of equations define dynamic movement over time. The programs defined in Chapter 4 provide an even higher level of representation which are traditional segmental level representations. It has previously been suggested that there may be this higher level of representation even within the framework of Articulatory Phonology. Fowler (2015) uses the "duality of patterning" (Hockett and Hockett, 1960), the "particulate principal" (Abler, 1989), and the success of alphabetic writing systems as evidence for why she believes combinations of gestures "are not far from conventional segments." The logical formulas I define in Chapter 4 show how this is true mathematically and see Kazanina et al. (2018) for additional arguments in favor of a segmental level representation in speech perception.

# Chapter 3

# The Blueprint Model of Production

This chapter discusses an alternative approach to the production part of the phonetics-phonology interface, the BLUEPRINT MODEL OF PRODUCTION (BMP). First, I show how the BMP reconceptualizes the relationship between phonology and phonetics. This is done using a type-functional analysis. One outcome of this is that the BMP can account for gradient phenomena without resorting to a gradient phonology. Consequently, arguments for replacing or removing the phonological module because of systematic phonetic details are not sufficient to displace discrete, symbolic phonology. After the formal introduction, simulations are used to support the claim that the BMP can explain these phonetic phenomena without resorting to gradient phonological knowledge.

There are two essential points to understanding the BMP. First, it concretely models the production process with multiple, simultaneous factors, of which phonology is just one. Second, the whole phonological module is a factor in production, not just the representations it outputs. Like Gafos and Benus (2006), this approach "does away with the problematic metaphor of implementation or precedence between phonology and phonetics without losing sight of the essential distinction between the two (qualitative, discrete vs. quantitative, continuous)." And while Gafos and Benus (2006) provide a specific way of accomplishing this using dynamical systems, it is not the only way possible. This chapter provides an alternative explanation grounded in type theory and the theory of computation.

In Chapter 2, the modular feed-forward model was likened to an assembly line. Instead, imagine that the phonology was not a station that a lexical item had to pass through during the production process, but rather a target design that the phonetics module was given alongside a lexical item. In this metaphor, the lexical item is a set of materials, the phonology is a blueprint for what the assembled form should look like, and the Phonetic Station is the module that is doing the assembling. The phonology still operates in the same way as in modular feed-forward models: given a UR as an input, it returns an SR as its output. Only now this process does not strictly precede phonetic implementation (cf. Gafos and Benus, 2006). This characterization of the production process situates the phonology in a way that allows it to maintain its primary role of determining the surface form of an underlying representation. It also allows the Phonetics Station simultaneous access to both the underlying representation and the phonological instructions on how to modify it. As will be explained in more detail later, by "phonological instructions" I simply mean a map from URs to SRs. No other history of a phonological derivation or evaluation is visible to the Phonetics Station.[1] The main point here, however, is that under this architecture, the lexical form is not invisible to the Phonetics Station.

The simultaneous, or parallel, view presented here may evoke connectionist models of cognition (Hinton and Anderson, 1981; Feldman and Ballard, 1982; Rumelhart et al., 1988).[2] The current use of simultaneity varies from the connectionist view since I am talking about it in terms of composing many smaller functions into a larger function. The computation of this larger function does not need to happen in parallel or require

---

[1] The only way around this would be to encode such information into the SR itself. For example, in OT-CC (McCarthy, 2007) chains of successive modifications to a form are evaluated. The history could be encoded in the output of the phonology if the whole chain were output instead of the just the last representation in the chain. There is also previous work in which the UR is encoded in some way within the SR (van Oostendorp, 2008, and others).

[2] For further discussion on the relationship between symbolic and neural theories of cognition see Smolensky (1988).

a neural architecture. The functions being proposed can be instantiated in any number of ways, including ones that follow connectionist/neural principles and ones that do not. That being said, the BMP is presented as essentially a "bag of things" model where information from multiple sources combines together to compute the phonetic exponent. These types of models have precedence in the sentence processing[3] (i.e., interactive models; Marslen-Wilson, 1975; Altmann and Steedman, 1988; Tabor and Tanenhaus, 1999; Grodner et al., 2005, among many others), word production (Harley, 1984; Stemberger, 1985; Dell, 1986; Rapp and Goldrick, 2000, among many others), and speech perception (McClelland et al., 2006; McClelland and Elman, 1986; Norris, 1994; Baese-Berk and Goldrick, 2009, among many others) literatures. This approach is also taken in many statistical approaches to language modeling (i.e., *large language models* or LLMs). One downside to the LLM approach is that we are largely ignorant of the internal workings of what is going on. Here, I take a structured and *interpretable* approach to try to understand how the different things in the "bag" specifically interact.

The formalization below expands on the view expressed in the previous chapter that each module can be thought of as a function (Roark and Sproat, 2007; Heinz, 2018). In the modular feed-forward model, the phonology module is a function that maps a $UR$ to an $SR$ and the phonetics module is a function that maps an $SR$ to a $PR$ (Phonetic Representation). The BMP continues to view the phonology module as a function that maps a $UR$ to an $SR$ but views the phonetics module as a higher order function that *takes the phonology module function as an input*. In addition, to generalize over all lexical items, the entire lexicon is considered to be an input to the phonetics module instead of a single $UR$.[4] The phonetics module is therefore a function with at least two inputs: the lexicon

---

[3] I consider my proposal to be most philosophically aligned with Ferreira and Nye (2017). They discuss how to unify ideas about modularity and interaction in the domain of sentence processing, primarily emphasizing the shallowness aspect of modularity over encapsulation. They point to a question like, "what sources of information are in fact part of the parsing module?" and claim that parallel computation is fully compatible with modularity.

[4] Treating the lexicon as a unitary object is common in computational treatments of morpho-phonology, where the lexicon is represented with a single finite-state transducer (Roark and Sproat, 2007; Gorman and Sproat, 2021).

and the phonology module; and one output: a set of phonetic representations $\{PR\}$. The two contrasting models are shown in Figure 3.1 below. The next section provides a formal definition of the BMP.

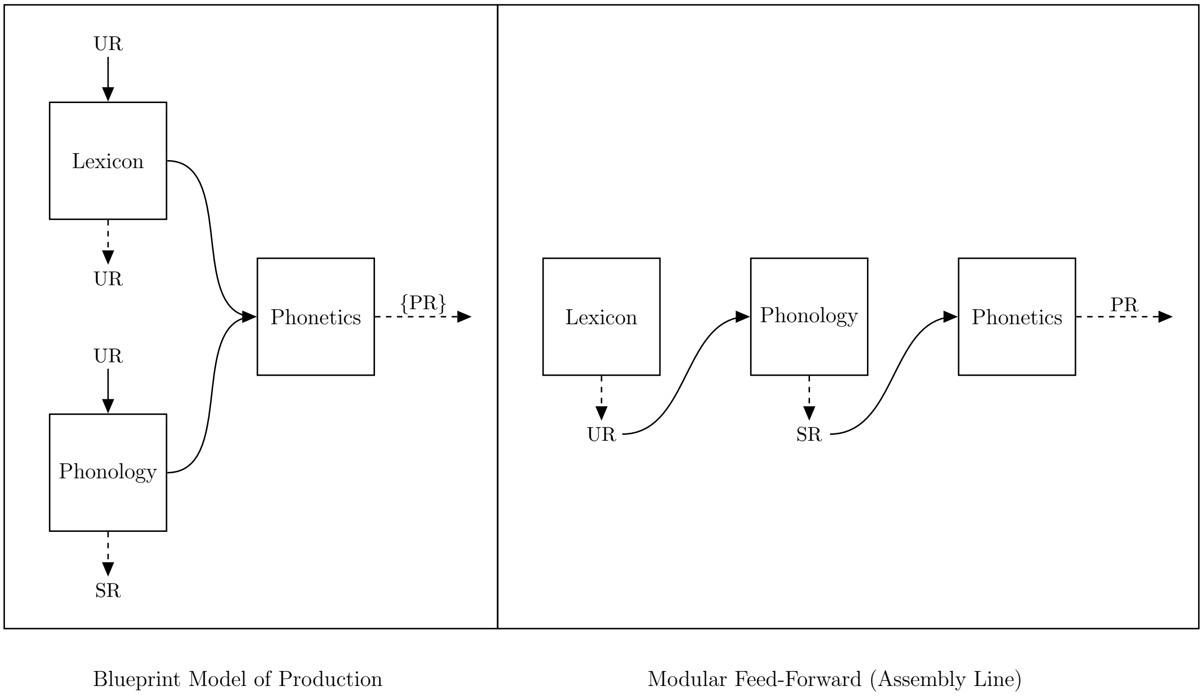


Figure 3.1: Visual comparison of the architecture for modular feed-forward models and the BMP. Each box represents a function/module. Solid lines represent the inputs to each function while dashed lines represent the outputs of each function.

## 3.1 From assembly line to blueprint: function (de)application

While giving the phonetics module direct access to the lexicon and phonology may seem like a large departure from the feed-forward model, the BMP can be related directly to the feed-forward model via function application. The analyses rely on the function type each module computes.

I now apply these ideas to architectures of language production. Throughout the remainder of this section, the following abbreviations are used: $L$, $P$, and $A$ as functions representing the **L**exicon, **P**honology, and and Phonetics (**A**rticulation or **A**coustics); $UR$, $SR$, and $PR$ to represent **U**nderlying **R**epresentations, **S**urface **R**epresentations, and **P**honetic **R**epresentations. The proposed types are listed in the table in Table 3.1.[5]

| Name | Meaning | Type |
|---|---|---|
| $L$ | Lexicon | $UR \rightarrow UR$ |
| $P$ | Phonology | $UR \rightarrow SR$ |
| $A_{\text{MFF}}$ | Phonetics$_{\text{MFF}}$ | $SR \rightarrow PR$ |
| $A_{\text{BP}}$ | Phonetics$_{\text{BP}}$ | $L \rightarrow P \rightarrow \{PR\}$ |
| $UR$ | Underlying Representation | $UR$ |
| $SR$ | Surface Representation | $SR$ |
| $PR$ | Phonetic Representation | $PR$ |

Table 3.1: Types

This paragraph describes the steps that turn the modular feed-forward model into the BMP. To start, the phonetics module in the modular feed-forward model has the following type.

(6) $A_{\text{MFF}} :: SR \rightarrow PR$

This idealizes the phonetics module as a map from surface representations to phonetic representations. Given a UR, the phonology P, and the definition of function application from above, one can decompose $SR$ into $UR \rightarrow (UR \rightarrow SR)$.

(7) $A :: UR \rightarrow (UR \rightarrow SR) \rightarrow PR$

Next, $(UR \rightarrow SR)$ is just another way of representing the phonology module.

(8) $A :: UR \rightarrow P \rightarrow PR$

To complete this reconceptualization we change $UR$ to $L$ in order to generalize over the entire lexicon. By doing so, the output is now a set of phonetic representations rather than a single specific representation. This gives us a new type for the phonetics function.

---

[5] In Figure 3.1 the Lexicon has type $UR \rightarrow UR$. In this case, it can be thought of as the identity function. This is an abstraction to facilitate the analysis.

(9) $A_{\text{BP}} :: L \rightarrow P \rightarrow \{PR\}$

The phonetics module is therefore a higher-order function with two arguments: the lexicon and the entire phonology module (a function). As is the case in the modular feed-forward model, the phonology still maps an underlying form to a surface form. Additionally, in both the BMP and the modular feed-forward model an underlying form is ultimately transformed into a phonetic representation. The main difference is phonology is no longer an intermediary between the lexical form and the phonetics module. Instead, the phonology and the lexical form are both input to the phonetics module.

If it is not clear yet as to why this is being called the BMP, consider this. For every $n$-ary function there is an equivalent $(n+1)$-ary relation. Since phonology is a unary function (i.e., it has one input which is a $UR$) it can also be envisioned as a binary relation consisting of $UR$ and $SR$ pairs $\langle UR, SR\rangle$. This latter perspective highlights the fact that we can view phonology not as a module that deterministically shapes the phonetic output, but instead as a set of instructions that informs the phonetics module as to how a given lexical item should be pronounced. In other words, in the same way one would query a blueprint, the phonetics module queries the phonology as to how a $UR$ should be pronounced.

The derivation shown above does not exhaustively represent all the factors that determine production. It simply shows how the BMP relates to the feed-forward model of production. Many other factors have been argued to influence speech production. For example, in the case of incomplete neutralization, it has been argued that the phonetic output is not only a blend of the phonological output ($SR$) and the lexical input ($UR$), but also that this blend can be scaled by extra linguistic factors relating to contrastive intent (Port and Crawford, 1989; Ernestus and Baayen, 2003; Gafos and Benus, 2006). This is an additional factor necessary to adequately account for production. As will be discussed in more detail in §3.2.1, this is accomplished with the BMP by adding the intent ($I$) as one of the arguments to production: $A :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$.

Additionally, the BMP highlights the relative importance of certain types of information during the production process. While each factor plays a role in determining the phonetic output, the long-term memory representation of the pronunciation of a lexical item is arguably the most important factor since the entire goal of the production process is to externalize it in some way. Phonology is also important since it is largely viewed as an automatic process that systematically adjusts category level aspects of the pronunciation in a context-dependent way.[6] On the other hand, while pragmatic intent and lexical frequency systematically influence the phonetic output, they do so by scaling the targets that are determined by the lexicon and phonology.

These additional factors are also related to the blueprint metaphor. Imagine there is a blueprint for building a picnic table. In one scenario a person uses this blueprint to build a table for an indoor area. In a second scenario, a different person uses the same blueprint to build a table to be used in an outdoor area. They both use the same materials and the same set of tools and end up with two tables that are practically identical. The person in scenario two then adds a clear coat of waterproofing since the table will be kept outside. To the naked eye there are still two identical tables, but closer inspection shows there is a fine-grained difference between the two. The blueprint is not explicit about how the table is used and therefore does not supply any further information beyond how to assemble the table. Despite this, sometimes there are factors beyond its construction that affect its final form.

The role of phonetics in the BMP is to take a set of materials (the lexicon) and a blueprint (the phonology) and construct the correct forms. Depending on the use of these forms, they are further altered by situational need (pragmatic context, frequency counts) to provide the final set of instructions to the motor system. In this sense, the BMP provides a *phonologically based phonetics* (c.f. Hayes et al., 2004). The phonetic form is dependent on the phonological output, but there is plenty of room for systematic influence from other

---

[6] I recognize that certain processes are optional and/or gradient, but argue that phonological accounts of them still automatically takes place. As it relates to previous analyses, the optionality and gradience is determined by the automatic application of the phonology function. Nonetheless, I expand on an alternative perspective in Sections 3.2.1 and 3.2.6 that was initially discussed in Section 2.2.3.

factors. In fact, the BMP in many ways is a formalized version of what Du and Durvasula (2024a) call the “classic generative phonology” view, which explicitly situates phonology as only one source of information in the production process. A clear description of this view comes from Mohanan (1986, p. 183; emphasis original):

> Practitioners of phonology often distinguish between *internal* evidence, which consists of data from distribution and alternation, and *external* evidence, which consists of data from language production, language comprehension, language acquisition, psycholinguistic experiments of various kinds, sound patterning in versification, language games, etc. [...] The terms “internal” and “external” evidence indicate a bias under which most phonological research is being pursued, namely, the belief that the behavior of speakers in making acceptability judgments is somehow a more direct reflection of their linguistic knowledge than their behavior in producing language, understanding language, etc. This bias appears to be related to the fact that linguistic knowledge is only *one* of the inputs to language production, language comprehension, and other forms of language performance. What accounts for the facts of performance is a *conjunct* of a theory of linguistic knowledge (”What is the nature of the representation of linguistic knowledge?”) and a theory of language performance (”How is this knowledge put to use?”).

## 3.2 Simulations

Importantly, there are several ways the BMP can be instantiated. This fact means there are distinct levels of analysis, which it is important to be clear about. I reserve the word ‘model’ for higher-level architectures of the phonetics-phonology interface (such as the BMP) and ‘simulation’ for a specific instantiation of a model (such as the ones we present in later sections). This is not standard usage and many scholars use the word ‘model’ to refer to what I am calling a 'simulation'. One reason to draw a firm distinction between the two is that it is quite easy to mistake a simulation with the higher level model, but as McCloskey (1991, p. 390) warns, "any simulation includes theory-irrelevant as well as theory-relevant details; hence, the details of a simulation cannot be identified straightforwardly with the details of the corresponding theory." Cooper and Guest (2014) provide a similar warning. Unlike McCloskey (1991), I use the word 'model' in place of 'theory.' This is because theories can exist at different levels (Marr, 1982), and across levels, which means a theory may consist of a specific implementation alongside a higher-level architecture.

The primary contribution of the BMP is at the level of the model and not the simulation. Nonetheless, the simulations are important because they illustrate concretely how the model can be implemented. They should not be confused, however, with the model itself. Simulations can be used to test aspects of the model, but simulations come with their own set of assumptions, some of which may be ancillary to the model itself. For example, a simulation may require a parameter whose exact value cannot be derived from the model and instead is estimated from data deemed relevant. Consequently, critiques of a simulation are not necessarily critiques of the model. It depends on the particulars: a critique of how a parameter's value in a simulation is estimated is not necessarily a critique of the model architecture.

As with the modular feed-forward model, the BMP itself, as an abstract characterization of the phonetics-phonology interface, has little to say about specific instantiations. Due to this *computational level* description of the BMP (Marr, 1982), there are infinitely many possible instantiations. Thus in this sense, the formal model overgenerates. But this is by design: our goal is to describe *capacities* (Cummins, 1983; van Rooij and Baggio, 2021), not specific implementations. This type of abstract analysis runs into the problem of *multiple realizability* (Putnam, 1967; Fodor, 1974; Guest and Martin, 2023). For example, having the capacity to sort a list of items does not say anything about which of the nearly 50 proposed sorting algorithms[7] is being used. In this same spirit, I am making a claim that language users have a capacity that involves combining lexical information, phonological information, and extra-grammatical information when producing speech. While this claim may seem modest, it stands in contrast to the feed-forward model, which prohibits the combination of lexical and phonological information that the BMP provides.

With the computational level description of the BMP in mind, the simulations I present are not without their own specific assumptions. For example, the simulations model variation in the productions of single speakers, not populations of speakers. In addition, they are mostly deterministic because the addition of stochastic variables does not change the overall findings in regard to the specific claim about the model's capacity to account for gradient phenomena with a discrete phonology. In other words, the goal in the simulations is not to find the best quantitative fit of all the variation reported in the literature but to capture important qualitative attributes sufficient for the argument. This "proof of concept" is step one in what I envision as a larger research program. In future work, it will be necessary to not only provide a qualitative fit, but a quantitative one as well. Part of this will involve restricting the space of functions that are used when defining specific implementations of the BMP. One area that I think will be especially useful in this regard is computational complexity theory as this has already been proposed as a way to restrict general cognition (Frixione, 2001; van Rooij, 2008) as well as phonological cognition (Chandlee, 2014; Heinz, 2018; Lambert et al., 2021).

As discussed above, the approach of Gafos and Benus (2006) has been rather influential on the formulation of the BMP. When accounting for final devoicing, they describe a constraint grammar based on nonlinear dynamics that contains separate equations for a markedness constraint (pulling the system towards a voiceless surface form) and a faithfulness constraint (pulling the system towards a voiced underlying form). The two ap-

---

[7] `https://en.wikipedia.org/wiki/Sorting_algorithm`

proaches share many aspects: the lexicon and grammar are expressed in terms of functions, extra-grammatical information can enter the computation, and there is no direct precedence of phonology over phonetics. Fundamentally, though, these ideas are expressed in two different mathematical frameworks. I use the language of functions and function types as used in programming language theory and other areas of theoretical computer science and discrete math. Gafos and Benus (2006) use the language of non-linear dynamics which allows them to simultaneously express discrete and continuous aspects of a complex system.

These two approaches make very different philosophical claims about cognition in terms of the symbolic nature of cognitive knowledge. One large advantage to the dynamical systems approach when it comes to phonetics and phonology is the fact that there is no extra translation mechanism needed to turn symbolic phonological knowledge into continuous phonetic substance. Nonetheless, I believe it is instructive to imagine an instantiation of the BMP which draws directly from the dynamics of Gafos and Benus (2006).

Consider an instantiation of the BMP to cases where the type realizations for the underlying representations, the surface representations, and the phonetics representations, are the same (i.e - $UR = SR = PR$). In particular, these representations reference specific phonetic cues which are given by differential equations of the form $\dot{x} = f(x) = -dV(x)/dx$ that describes a time-invariant first-order dynamical system in control of a cue, and $f(x)$ is a force function acting upon the state of the system and $V(x)$ is the related potential. For concreteness, consider the force function $\dot{x} = F(x) = x^{REQ} - x$ with corresponding potential $V(x) = x^2/2 - x^{REQ}x$ where $x^{REQ}$ is a set of target values $\{-x_0, x_0\}$ which are fixed based on the positive and negative values of some binary feature. If used as the functions for $UR$ and $SR$ in the BMP they would represent the underlying and surface values of the relevant phonetic cue. The phonetics module in the BMP could then combine them in the way as described in equation (10) above to get the final $PR$ form. Ultimately, Gafos and Benus (2006) take a different approach to their dynamics. They use a tilted anharmonic oscillator to formalize a markedness force function: $\dot{x} = M(x) = -k + x - x^3$ with corresponding potential $V_M(x) = kx - x^2/2 + x^4/4$. In addition, they use a $\theta$ parameter to express contrastive intent within their faithfullness force function as a way to influence the "underlying" form: $\dot{x} = F(x) = \theta(x^{REQ} - x)$; and corresponding potential $V_F(x) = \theta x^2/2 - \theta x^{REQ}x$. They then add the two forces together: $\dot{x} = M(x) + F(x)$.

The point of this exercise is to emphasize clear parallels between the BMP and the specific approach of Gafos and Benus (2006). Where the BMP associates a cue value with a UR, they have a force equation that places a fixed point at the corresponding lexical/underlying value for voicing (faithfulness). Where the BMP associates a cue value with a SR, they have a force equation that pulls the system towards a point corresponding to the surface value for voicing (markedness). In both cases, these values/equations are summed, but in the case of dynamical systems, the scaling controlled by the contrastive intent happens within these equations themselves and not with an external parameter as is the case for the BMP.

What I continue to stress in this dissertation is that language production involves the interaction of lexical, phonological, and extra-grammatical factors which the modular feedforward model fails to capture. Since this idea can be expressed using different types of mathematical formalisms, I believe that this idea is not a property of the specific mathematical implementation, but rather a property of the high-level architecture (a "model" in the terms outlined above). The simulations below stress this fact by showing that a non-dynamical implementation involving a discrete phonological grammar can also account for the qualitative behavior of individual language users.

Since the simulations do not use dynamical systems, they provide an alternative approach to the interface. However, I am not proposing these simulations *in opposition to* the dynamical approach. While I believe that the success of the simulations *sufficiently* captures important qualitative aspects of production, it does not *necessarily* negate the dynamical systems approach to the interface. The simulations are introduced to establish that the BMP is a framework with many possible instantiations. This further helps clarify which level of analysis provides the source of explanation for speaker behavior in the phenomena we study. In my opinion, it is at the level of the model and not the level of the simulation.

### 3.2.1 Expanding the formal definition of the BMP

**Intent**

Section 3.1 provided a formal characterization of the production process. The focus in that section was to show how the BMP conceptualizes the phonetics module as taking lexical forms directly alongside the phonology. This means that both UR and SR information are available, something that will be useful in accounting for incomplete neutralization. That being said, its formulation so far lacks explicit parameters for controlling extra-grammatical factors such as the speaker's intent to maintain an underlying contrast. The BMP can be updated to include an intent variable in the input, which will scale the production in some way between the UR and SR targets. I use $I$ for the intent variable, updating the function to be: $A_{\text{BP}} :: L \to P \to I \to \{PR\}$.

In other words, the inputs to the phonetics module reflect multiple factors in production: the lexical form, the phonological instructions, and the pragmatic context. This is a high-level description, and in principle, one can see how the phonetics module can account for incomplete neutralization with this kind of architecture. Nonetheless, it is beneficial to provide one concrete instantiation to show how the BMP can simulate the gradient incomplete neutralization data while maintaining a discrete phonology. The one outlined below is used in the simulations in the remainder of the chapter.

Recall that the acoustic cues in incompletely neutralized segments are usually in the direction of what might be expected for a phonetic token of the underlying segmental quality. For example, Warner et al. (2004) found that Dutch words containing an underlying voiced stop that was devoiced in word-final position were pronounced with a longer preceding vowel than similar Dutch words containing underlying voiceless stops in the same position. Directionality of incompleteness is therefore essential to any account of incomplete neutralization. Additionally, as discussed in Section 2.2.1, Port and Crawford (1989) showed that the level of incompleteness seems to be scaled according to pragmatic context: there was more incompleteness when the phrase participants were asked to record did not contain information that helped recover the lexical meaning. Finally, it is a subset of cues that are found to be incomplete when taking acoustic measurements. Deciding which cues show up as being incomplete and why it is only a subset of cues lies beyond the scope of this paper. In the subsequent discussion, I talk about a single abstract cue along a one-dimensional space for an individual speaker for expositional simplicity and not epistemic commitment.

Returning to the German final devoicing example in (2), consider a one-dimensional space for some cue $c$ in the set of all cues $C$ that signify the voicing contrast for an individual speaker. Imagine dividing the space in a way such that there is a point where every value equal to or less than that point signifies a [+voice] sound while everything greater than that point signifies a [-voice] sound. Within the [+voice] sub-section there may even be different cue values depending on the position of the voiced sound. For example, an intervocalic voiced obstruent may be further away from a specific cue's boundary than a word-final voiced obstruent. It is also the case that the [-voice] sub-section can be full of different realizations. In the case of final devoicing, a faithful [-voice] sound may have value $n$ in the cue space. Likewise, a [+voice] obstruent in final position may have value $m$ in the cue space.[8]

---

[8] I am assuming here that the phonetics module is able to map a [+voice] sound at the end of a word onto some phonetic representation. Since the translation is feature-based this should not be a problem. The reason that a speaker of a language with final devoicing may never produce a [+voice] sound in this position is due to the phonology and not the phonetics. Anecdotally, speakers of languages with final devoicing can produce a word-final obstruent as voiced if absolutely forced to do so.

Since the BMP has access to both UR and SR information, the phonetic form is a blend of the phonetic form given the UR and the phonetic form given the SR. This means that the two points $m$ and $n$ provide a theoretical bound on the cue value for the devoiced obstruents in final position. If we assume that the intent variable introduced above controls how much influence the UR or SR has, then the cue $c$ can in theory surface as any value between $m$ and $n$. Of course, this also depends on the specific implementation of the intent value and scaling process. The next paragraph discusses one way in which the scaling procedure may be implemented. Figure 3.2 provides a visual conceptualization of the cue space for the words in (2). Arrows point to possible realizations. Notice that it is only the alternating case where multiple options exist for a given form.

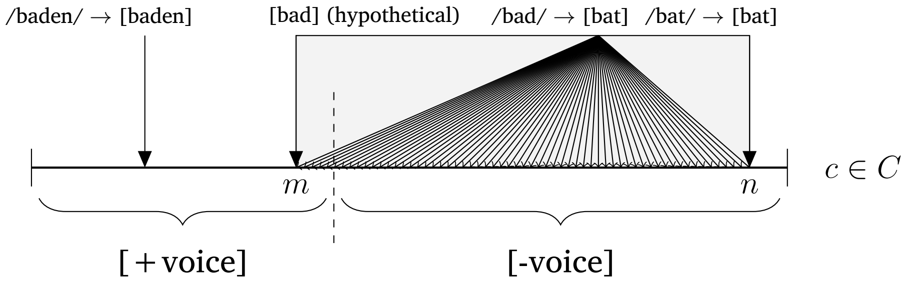


Figure 3.2: Hypothetical Cue space

The main idea sketched above is that the phonetic form is some combination of UR and SR influence.[9] This is the $I$ in the $A_{\text{BP}}$ function shown at the beginning of this section. Since intent can be thought of as a percentage corresponding with the level of desire a speaker has to maintain the underlying form, one way to formalize this notion is as a value in the unit interval $[0, 1]$. Here, the lower bound 0 represents a speaker with 0% intent to maintain the underlying contrast while the upper bound 1 represents a speaker who

---

[9] A similar idea has been used to account for phonological variation in the form of scaled faithfulness constraints in Harmonic Grammars (Coetzee and Kawahara, 2013; Coetzee, 2016). The scaling factors have been tied to concepts like register, speech rate, and lexical inhibition. The influence of each of these variables is controlled by a scaling factor which is added as an offset to each faithfulness constraints contribution to the harmony score. The non-grammatical factors therefore influence how much desire there is to maintain the fully faithful form.

wants to 100% maintain the underlying contrast. The exact value for cue $c$ is computed by simply taking a weighted sum of $c_{UR}$ and $c_{SR}$. In Figure 3.2, $c_{UR} = m$ and $c_{SR} = n$.

One simple way to combine the two values is to use the intent value directly as a weight. This suggests that the scaling process is linear. Another option is to allow for an exponential scaling process. Since incomplete neutralization typically results in subtle phonetic differences, a linear weighting might indicate that we would expect to see more intermediate cue values when measuring phonetic forms. Exponential scaling still allows for the UR value to have an effect on the phonetic form, but only in circumstances where there is a high intent value will it result in anything that is indistinguishable from noise. The following formula provides an exact formulation of exponential scaling where $\alpha > 0$.

(10) $c = c_{UR} \times I^{\alpha} + c_{SR} \times (1 - I^{\alpha})$

This formula has desirable properties. First, when $I = 0$, there is no effect of the UR on the output, and when $I = 1$ there is no effect of SR on the output. While this may seem trivial, it does match the informal explanation of intent. Second, since the scaling weights sum to 1, it is impossible for $c$ to fall outside the bounds set by $c_{UR}$ and $c_{SR}$. If we assume $c_{UR} < c_{SR}$, then for any arbitrary values of $I$, the only way for $c > c_{SR}$ is to have $c_{UR} \times (1 - I^{\alpha}) > c_{SR} \times (1 - I^{\alpha})$. But this reduces to $c_{UR} > c_{SR}$ which is a contradiction. This proof works the same way to show how it would not be possible to get a value lower than $c_{UR}$ in this same scenario. Third, because the $\alpha$ parameter is tied to a specific cue, it provides a potential explanation for *how* only certain cues can be incomplete. Again, I choose not to speculate on *why* certain cues show up as incomplete while others do not, but do provide this mechanism as a way to include the variation.

When $\alpha = 1$ there is a linear effect of the UR on the final output. In this case, the percent influence of the UR is equal to the intent value. As alpha increases, the influence of the UR becomes less and less for lower intent values. For high values of alpha, it is only high values of intent that will allow for the UR to have any influence on the output form. This exponential scaling potentially explains why the effects of incomplete neutralization are subtle, and also that, under extreme circumstances, speakers can produce something very UR-like (see fn. 8).

It is also worth briefly considering how this account handles *complete neutralization*. Various neutralizing properties such as manner of articulation in Korean word-final coronal obstruents (Kim and Jongman, 1996), and vowel length in Dutch Lahiri and Hankamer (1988), Egyptian Arabic (Broselow et al., 1997), and Palestinian Arabic (Hall, 2017) have been shown to result in complete neutralization. As mentioned previously in this section, low intent values in a phonologically neutralized form would still result in a different phonetic target than the faithful form. Teasing apart this type of subtle variation from noise would require phonetic experiments with much higher statistical power and even then would likely involve rethinking the types of experiments that would best bring out these types of results. The approach taken in this dissertation relies on already obtained phonetic data and focuses on incomplete neutralization since it has been used as a phenomenon that does not support discrete phonological knowledge. That being said, as discussed in Chapter 1 and later in Section 3.4, a well formalized theory can guide experimental design and may even be more robust to failures of replication.[10]

Under the currently sketched account, complete neutralization occurs when the intent value is sufficiently close to 0, and incomplete neutralization occurs when the intent value is sufficiently close to 1. While intent is tied to pragmatics here due to the experimental design of Port and Crawford (1989), many factors could influence the desire to maintain some influence of the underlying form. Ernestus and Baayen (2003, 2006) take a slightly different approach where the phonetic influence is not the underlying form directly, but rather results in simultaneous activation of related word forms in a connectionist network. But underlying forms in a generative/symbolic approach also serve as a way to relate morphologically distinct word forms that share the same base UR. Therefore, the intent value in the simulations throughout this dissertation could be affected by knowledge of

---

[10] See Guest and Martin (2021) for deeper discussion on this last point.

morphologically related words.

Furthermore, since the lexicon itself is an argument to the phonetics function, lexical statistics can straightforwardly be integrated when considering the incomplete neutralization behavior of speakers producing nonce words. Ernestus and Baayen (2003) showed that speakers used lexical statistics to determine a possible UR in Dutch pseudo-words like [tif] and [dɑup]. Because of the final devoicing process that is active in Dutch phonology, the UR is underdetermined based solely on these forms. As part of their experiment, they asked listeners to listen to these types of words and then write down a "past tense form" with either *-te* indicating a final voiceless sound or *-de* indicating a final voiced stop. The choices in responses were ultimately best predicted by phonological similarity to morphemes in the Dutch lexicon.

Since the responses captured in this experiment do not include phonetic information it is unknown if speakers incompletely neutralize the pseudo-word forms in these types of cases. Nonetheless, it is possible to speculate how they might behave and unify the general spreading activation account with their pseudo-word behavior. Since nonce words do not have underlying forms in the lexicon directly, one can imagine the following process for how speakers may produce nonce word forms: perceive the form as ending in a *phonologically* final voiceless segment, use the entire word form against the lexicon to determine a possible UR, use the regular morphological and phonological knowledge to determine the set of related forms, weigh how much influence of the UR the pragmatics of the situation dictate there should be, and finally produce the phonetic form. The specific implementation would need to be worked out since it involves levels beyond what is discussed here (speech perception, morphology), but the general idea can be neatly integrated into the larger idea of the BMP.

### Frequency

In the same way that Intent is an input to the Phonetics function in 3.2.1, frequency information can be added to help explain frequency induced variation such as the different in homophone durations discussed previously. Frequency is represented as a function $F$ and the phonetics function is updated accordingly: $A_{\text{BP}} :: L \to P \to I \to F \to \{PR\}$. In other words, the phonetic implementation is a function that takes in the lexicon, the phonology, an intent variable, and a frequency function. The frequency function we envision has the type $F :: L \to \mathbb{R}$. Since the Lexicon is a set, the frequency function maps each item in the lexicon to a number that corresponds to its frequency. Again, the inclusion of the input form of lexical items *vis-à-vis* the lexicon is what allows us to account for the phonetic variation. Furthermore, it is important that the same phonological form does not entail the same lexical item since they are distinguished by syntactic and semantic information in the lexicon.

Another way to think about this is through the analogy of a computer's memory system. Each lexical item would be represented in memory as a unique bit string. The memory system does not care about the content of what it is storing, it just has different values stored at different bit addresses. The lexicon can be thought of in this same way. Under this type of architecture, the frequency information for a given lexical item is determined by a function rather than stored directly in the lexical entry. This can be seen as a way to encode the difference between knowledge *of* language and knowledge *about* language. The former refers to grammatical knowledge while the latter refers to language use. Based on the studies discussed in the previous section, it is clear that both are necessary for the production process.

Before continuing further, I introduce a function $\pi :: (UR \mid SR) \to PR$ that converts objects of the type $UR$ or $SR$ into a phonetic representation. Here, we assume this is a tuple of ordered cue parameter vectors. These may be articulatory or acoustic cues as long as they contain both spatial and temporal information. Given $\pi$, formula (10) discussed in the previous section for the implementation of the intent scaling would now be (11).

(11) $\tau = \pi(L) \times I^{\alpha} + \pi(P(L)) \times (1 - I^{\alpha})$.

Recall that $L$ contains URs and $P(L)$ returns SRs. So this is just the intent scaling over all cues for all phonemes of a given lexical item that is being produced. Here $\tau$ can be thought of as determining the overall target value with type $\tau :: L \rightarrow P \rightarrow I \rightarrow \{PR\}$. It therefore provides a foundation that other factors can slightly alter. With that idea in mind, consider a duration scaling factor $\delta :: \mathbb{R} \rightarrow [0,1]^n$. Specifically, $\delta$ maps frequencies to the unit interval. These functions $\pi$, $\tau$ and $\delta$ can be considered sub-programs within the larger phonetics function $A_{\text{BP}}$.

In order to complete the description of this process, it is necessary to explain how the various input elements interact. In this simulation, I propose that the target value output by the $\tau$ function is multiplied by the output of the $\delta$ function to provide a frequency scaled phonetic output. Following the assumption that the phonetic representation is a vector of parameters, the $\delta$ function outputs a vector rather than a scalar. In this way, frequency effects can occur under the architecture of the BMP without needing to place them directly in the lexicon or the phonological grammar. Instead, they are just one more factor alongside the lexicon, phonological grammar, and pragmatic intent that influences production.

This particular implementation is inspired by Pierrehumbert's (2002) simulation of leniting bias. She defines the production of a given token $x$ as $x = x_{target} + \varepsilon + \lambda$, where $x_{target}$ is the specific phonetic target that has been computed based on an exemplar model, $\varepsilon$ is some random error, and $\lambda$ is the leniting bias. This is motivated because leniting bias is closely related to duration (Priva and Gleason, 2020) and duration is related to frequency. For this implementation, the equivalent of $x_{target}$ is the output of $\tau(L, P, I)$, the equivalent of $\lambda$ is the output of $\delta(F(L))$, and instead of adding the bias term to the target, this implementation multiplies them.

While the data being modeled only involve temporal cues, the implementation would equally apply to spectral cues as well. This raises the question as to whether or not frequency information can also influence non-temporal cues. The answer appears to be yes. In a recent review of phonetic reduction, Clopper and Turnbull (2018) discuss ways in which various factors such as frequency affect both spectral and temporal cues. The primary spectral cue that has been investigated in relation to frequency is the F1 $\times$F2 vowel space which has been shown to be more contracted for more frequent words (Munson and Solomon, 2004). Crucially, Munson and Solomon (2004) found vowels in low-frequency words to be longer than vowels in high-frequency words but found no statistically significant interaction between duration and vowel-space expansion. Therefore, a simulation that accounts for both spectral and temporal cues would necessarily have to tease apart the influence of duration from the influence of frequency. This, however, has no impact on the architecture of the BMP since it already claims that both types of information are available during the production process.

### Optionality

Optionality/variability can be thought of as a parameter $v \in \{0, 1\}$ that chooses whether or not to use the surface form (phonological knowledge) in the computation. When $v = 1$ no phonological knowledge is applied and when $v = 0$ phonological knowledge is applied. We can update the scaling formula shown in (10) to include the $v$ parameter.

(12) $c = c_{UR} \times \text{max}(v, I^{\alpha}) + c_{SR} \times (1 - \text{max}(v, I^{\alpha}))$

Now the scaling is controlled based on the max value of $i$ and $v$. When $v = 0$, the equation becomes equivalent to the form in (10) because $i$ will always be greater than or equal to $v$. When $v = 1$, the opposite occurs because $i$ can only equal $v$ if it is equal to 1. Therefore, the surface form (phonological knowledge) is never used in this case.

### 3.2.2 Final Devoicing in German

The intent argument was added to the BMP to account for Port and Crawford's (1989) results from German that show that the level of incompleteness can vary based on pragmatic factors. This section shows how the intent argument and the $\alpha$ parameter can interact to simulate their findings. The simulation below focuses on burst duration which was the main cue Port and Crawford (1989) found to be incomplete, and closure duration which they found to be complete. The exact cues found to incompletely neutralize have varied across studies. For example, there have been conflicting results about whether or not preceding vowel duration is an incomplete cue in German final devoicing. Nicenboim et al. (2018) ran a statistical meta-analysis using a Bayesian random-effects regression model and found a main effect that supported vowel duration as a significant cue of incomplete neutralization. Port and Crawford (1989) on the other hand reported preceding vowel duration as being complete in their findings.

My assumption, based on the results of Port and Crawford (1989), is that the level of incomplete neutralization can be dynamically controlled based on pragmatic context. With this in mind, I provide a simulation of their results to highlight the distinction between cues that are complete and cues that are incomplete, while accounting for the pragmatic scaling based on intent. Since the conclusions between Port and Crawford (1989) and the meta-analysis conflict concerning vowel duration, I avoid this cue altogether. Ultimately, the simulation results do not depend on the specific cues that neutralize incompletely or not, but rather on the working assumption that cues can vary in this way at all.

Mean values for both closure duration and burst duration for each neutralized final stop pair and each condition are shown in Table 3.2. These data are taken directly from Port and Crawford (1989, Table 1; p. 265).[11] The ratio columns were added by dividing the final /d/ values from the final /t/ values within each condition. Since only the voice-

---

[11] This is data aggregated across multiple speakers. Our simulation treats this as one speaker. I discuss how to lift the simulation to populations at the end of this section.

| Condition | | Closure Duration (Mean) | Ratio | Burst Duration (Mean) | Ratio |
|---|---|---|---|---|---|
| 1A | /d/ | 54.72 | 0.91 | 20.08 | 0.78 |
| | /t/ | 59.84 | | 25.59 | |
| 1B | /d/ | 50 | 0.91 | 16.54 | 0.58 |
| | /t/ | 54.72 | | 28.35 | |
| 2 | /d/ | 68.89 | 1.02 | 32.87 | 0.83 |
| | /t/ | 67.52 | | 39.37 | |
| 3 | /d/ | 86.22 | 1.03 | 25.20 | 0.29 |
| | /t/ | 83.46 | | 85.63 | |
| 4 | /d/ | 88.98 | 0.99 | 59.06 | 0.89 |
| | /t/ | 89.93 | | 66.51 | |

Table 3.2: Data from Port and Crawford (1989) for neutralized final stops by condition. Ratio indicates the mean value of /d/ divided by the mean value of /t/.

less target is recoverable from the phonetic data in final position, I rely on the ratio to relate surface final /d/ to some hidden underlying target.

The results are simulated by assuming a single intent value for each pragmatic context, but a different alpha value for each cue in the scaling function. Abstracting away from specific values, I assume for all cues that a value of 1 is equal to the voiceless target and a value of 0 is equal to the voiced target. Since the focus is on accounting for the different levels of incompleteness, using ratios abstracts away from the condition specific variation. Therefore, the ratios reported in Table 3.2 can be used to estimate intent values.

Each subfigure within 3.3 shows the estimated cue values for both burst duration and closure duration. In general, the ratio of closure duration for underlying /d/ segments to underlying /t/ segments was 0.91 or higher for each condition. Since burst duration (represented as $+$) was found to significantly vary between derived and faithful surface /t/ segments in the pooled data, but closure duration (represented as $\times$) was not, the $\alpha$ parameter was set to 1 for burst duration and 20 for closure duration.

The ratios for burst duration varied from 0.29 for condition 3 to 0.88 for condition 4. Intent values were determined by subtracting the burst duration ratios from 1. The resulting plot shows that even with largely varying Intent values, the alpha parameter can make it so only a single cue shows up as being incomplete.[12]

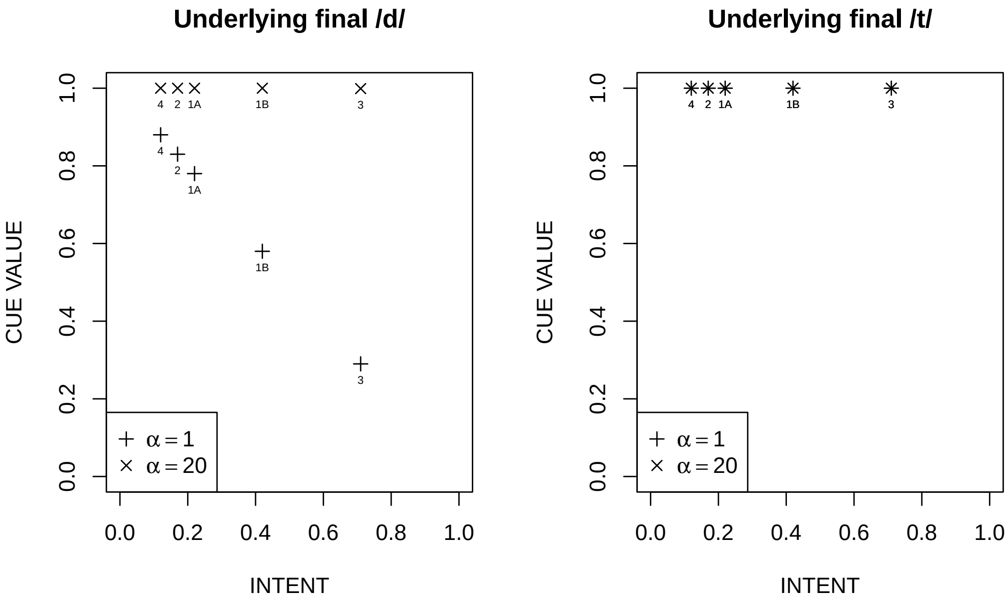


Figure 3.3: Simulated cue values for Port and Crawford (1989) results. The left plot shows values for /d/-final words and the right plot shows values for /t/-final words. Symbols + and × represent burst and closure duration, respectively.

From the figures, it is possible to compare both within plots and between plots, resulting in four comparisons. Based on the Port and Crawford (1989) data, we expect variation between the two cues for final /d/ and no variation between the two cues for final /t/. We should also expect to see variation between final /d/ and final /t/ for burst duration, but no variation between final /d/ and final /t/ for closure duration. Within the left plot, the cue values are shown to vary between burst duration (+) and closure duration (×), as expected. Cue values close to 1 indicate that the final /d/ that has been neutralized has acoustic properties that are similar to the faithful final /t/ segments. The closure duration cue values for /d/ are close to 1 as they are for /t/. For /t/, the burst duration is close to 1 as well. Again, this is expected given the data. While it may seem

---

[12] In their discriminant analysis, Port and Crawford (1989) found that condition 2 was more easily recognized as underlying /d/ than conditions 1. This goes against the acoustic data presented in the paper that shows that conditions 1a and 1b were more incomplete based on what the ratios suggest. This is likely due to glottal pulsing not being included as a cue in the discriminant analysis.

trivial that all of the underlying final /t/ values are right at 1 given that they were the denominator for determining ratios, these values were derived with the same formula that derived the final /d/ values. That is, the same $\alpha$ and the same intent values were used, but the formula ensures that final /t/ values are unaltered.

This simulation shows that the interaction of the intent and the alpha parameters captures the aggregate behaviors observed by Port and Crawford (1989), where burst duration was incompletely neutralized and varied according to pragmatic context, while closure duration did not. Closer inspection of individual behavior in Port and Crawford (1989) shows that there was variation across individuals in the manifestation of cues in relation to incompleteness as well as interpretation of pragmatic context. The simulation could be modified to simulate this kind of population level behavior in different ways. One way would include a probability distribution over the $I$ and $\alpha$ parameters from equation (10). While this may better capture group behaviors, this would not provide any further insight into what I find important: the capacity of the individual speakers. The simulations therefore are deterministic under the assumption that a given speaker, with a specific intent value, and specific alpha values would act in a certain way. Likewise, an error term could be added to account for noise in the system not captured by the simulation, but this again would not change the interpretation of the qualitative behavior the simulation exhibits.

The overall structure of the BMP allows for lexical influence on phonetic form. It also accounts for incomplete neutralization while maintaining a singular phonological devoicing rule, contra Port and Crawford (1989) who claim that their data refutes such a possibility. They write, “One can apparently only write accurate rules for German devoicing by making them speaker-dependent and by employing a very large set of articulatory features to capture the detailed dynamic differences between the speakers’ implementation of the contrast (p. 280)”. This interpretation follows from conceptualizing the phonetics-phonology interface in terms of the modular feed-forward model, but it does not follow from conceptualizing it in terms of the BMP. This is because the BMP can capture "dynamic differences between the speakers' implementation of the contrast" by recognizing multiple *simultaneous* factors influencing phonetic production. One factor is the lexical form and another can be a discrete phonology with a singular devoicing process. Port and Crawford (1989) show that pragmatic context is a necessary ingredient, which is formalized in the BMP as intent. Individual speakers' implementation of contrast does not need to be encoded in the phonological grammar, because with the BMP speakers have access to the contrast outside of the phonological module. This highlights the role that both competence and performance play in the production process (cf. Chomsky, 1965). In both cases, there is knowledge that is being used during implementation: lexical knowledge, discrete phonological knowledge, and a continuous representation of contrastive intent. It follows that under the BMP a continuous phonetic output does not require a continuous phonological grammar.

### 3.2.3 Tonal Near Merger in Cantonese

The similarity between incomplete neutralization and near merger has been well documented (Ramer, 1996; Nycz et al., 2005; Winter and Röettger, 2011; Yu, 2011; Braver, 2019). While the term incomplete neutralization emerged from the phonetics and phonological literature, the term near-merger originated within the sociolinguistic literature. Near-merger can be traced back to Labov et al. (1972) and their work on New York City English. Words like *source* and *sauce* were reported to be identical by speakers but then consistently produced with slightly different phonetic forms. Near merger is therefore usually used when two classes of sounds are perceived as being of the same category, but produced with subtle variation.

One aspect of Port et al.'s (1981) argument for incomplete neutralization was that listeners could correctly guess the specific word at an above chance level, highlighting the perceptibility of the contrast. This suggests that the primary difference between incomplete neutralization and near-merger is whether or not the difference is perceptible. There is also the synchronic versus diachronic distinction. Near-merger has been used by sociolinguists to explain sound change while incomplete neutralization is often related to the active production process.

Alternations also help distinguish the two. In the *source* vs. *sauce* example, there is no alternation driving the neutralization, but incomplete neutralization is dependent on there being an alternation. Regardless of whether or not these two phenomena are one and the same, certain cases of near merger can be explained with the same mechanisms that have developed for incomplete neutralization using the BMP.

Tonal near merger in Cantonese as discussed by Yu (2007) is one such case. Unlike the *source* vs. *sauce* example, it involves morphological alternations called *pinjam*. These alternations involve a non-high level tone turning into a mid-rising tone.

(13) a. sou33 ‘to sweep’ → sou35 ‘a broom’
b. pɔŋ22 ‘to weigh’ → pɔŋ35 ‘a scale’
c. tsʰɵɥ11 ‘to hammer’ → tsʰɵɥ35 ‘a hammer’

The derived mid-rising tones of these *pinjam* words were compared with lexical mid-rising tones in lexical near-minimal pairs. The f0 value at the onset of the tone, the inflection point, and peak of the rise were all found to be higher for the *pinjam* words. Furthermore, a follow up study on this phenomenon showed that listeners were unable to tell the two types of mid-rising tones apart, thus giving it its near-merger status.

At first glance, this seems to make the opposite prediction of what might be expected given the UR/SR scaling account that has been developed so far. The derived *pinjam* 35 tones should be lower than the lexical mid-rising tones since they (potentially) correspond with a non-high level tone. A closer look shows that the phonological analysis involves an underlying floating high tone: pɔŋ22(55) → pɔŋ35 ‘a scale’ where parentheses indicate a floating tone. In this case, it may be interpreted that the reason that the *pinjam* mid-rising tone has higher f0 values than the lexically specified mid-rising tones is due to the inclusion of an underlying high tone.

Yu (2007) explains the data using an exemplar model with further support coming from contracted syllables (sandhi). The morphemes /tsɔ/ and /tɐk˺/ both surface with a mid-rising tone in contracted syllables:

(14) a. paŋ22 tsɔ35 → pɔ35 ‘to weigh (PERF)’
b. pɔŋ22 tɐk˺55 → pɔ35 ‘to weigh (POTENTIAL)

What makes it interesting is that /tsɔ/ has an underlying mid-rising tone while /tɐk˺/ has an underlying high tone. The f0 value at all of the three points was found to be higher for the mid-rising tone derived from the underlying high tone than for the mid-rising tone that was underlying mid-rising. In the BMP, this is exactly what would be expected. That is, a surface mid-rising tone that was derived from an underlying high tone should have its f0 values raised, given a non-zero intent value. Despite the exemplar interpretation, Yu (2007, p. 207) recognizes this fact and writes, “Thus, the extra-high f0 of the [derived mid-rising tone] can be interpreted as the retention of the tonal profile of an underlying [high] tone.”

Figure 3.4 shows simulated data for the sandhi process. The goal of this simulation is to show the qualitative “inbetween-ness” of the derived mid-rising tone at all three measured points. I take the same approach as in §3.2.1 where we abstract to a [0,1] cue space. In this example, 1 corresponds to a high tone (5) and 0 corresponds to a low tone (1). Using an $\alpha$ value of 2 and an intent value of 0.4, the values for three types of mappings are shown. A faithful mapping of the high tone (/55/->[55]), a faithful mapping of a mid-rising tone (/35/->[35]), and an alternation where an underlying high tone turns into a mid-rising tone (/55/->[35]). Squares indicate surface tone: squares are mid-rising and circles are high. The color indicates underlying tone: white is mid-rising and black is high. The derived mid-rising is therefore the black square.

In the simulation, the faithful mappings are unaffected by the $\alpha$ and intent values, and the values for the alternation mapping is an interpolation between these two extremes. This shows once again that this instantiation of the BMP captures important qualitative aspects of this tonal phenomenon.

One downside is that this simulation fails to capture the size of the difference at different points, and is therefore a shortcoming of the specific implementational choices. Nonetheless, the goal was to simulate the inbetween-ness and not the exact magnitude. One potential fix would be to vary the cue value for [55] at each point. Currently, the size of the difference is based on the size of the difference between the [35] target and the [55] target. It seems reasonable to say that the [55] peak is the true "1" value on the cue dimension and the onset and inflection point values are lower. Therefore, if the [55] peak is made relatively high enough, the difference at the peak will always be greatest. Since there is no data provided by Yu (2007) on the phonetic properties of the [55] tone, I leave this for future work. I stress once again that this specific implementational choice does not impact any claims about the model structure (i.e., the type of information available and the way it combines).

Yu (2007) also found that the mid-rising tone derived from an underlying high tone in the contracted syllables had higher f0 values than the *pinjam* mid-rising tone (also derived from an underlying high tone). The exemplar model explains this data with an averaging effect. An alternative explanation is that the act of syllable contraction may highlight the underlying form more directly than *pinjam* and therefore speakers are more likely to have a higher intent value, thus pulling the final phonetic form towards the underlying high tone values.

This section shows that while near merger and incomplete neutralization have been described as two separate phenomena, they can, in certain cases, emerge from the same basic system. The BMP only relies on a correspondence between underlying and surface forms which is anticipated through the phonological mapping. Any phonological change, whether it be morphologically driven or otherwise, will predict the same type of pho-

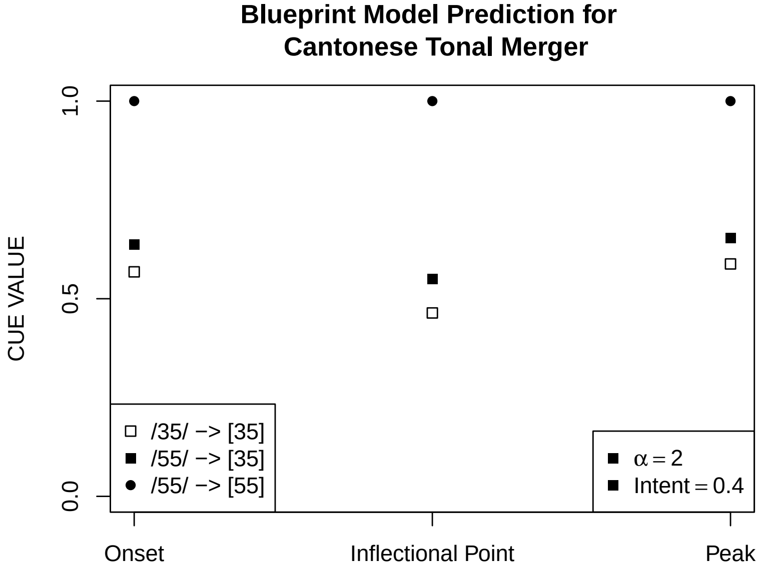


Figure 3.4: Simulated cue values for Yu (2007) tone-sandhi data

netic effects in this model. The phonetic distribution of any segment should therefore be bounded between what we would expect given the underlying form and the actual surface form.

### 3.2.4 Epenthesis in Arabic

Lebanese Arabic speakers epenthesize an [i] vowel to break up word final CC clusters. Gouskova and Hall (2009) performed an acoustic study that had speakers pronounce words with underlying forms /CVCC/ and /CVCiC/. Words of the first form are pronounced the same as the second form due to the epenthesis process. In both cases, the final vowel is an [i]. Measurements of the acoustic properties of these vowels found that the epenthetic [i] showed statistically significant differences in duration and occasionally F2 frequency when compared to [i] tokens that were present in the underlying form. Notably, the authors write, "...epenthesis introduces something *less than an [i]*: the vowel is backer and shorter, all properties that would make this vowel closer to [ɨ] or [ə] – and, arguably, to zero" (emphasis original). While they use Optimality Theory with Candidate Chains to explain these findings (OT-CC; McCarthy, 2007), the fact that the acoustic properties of the epenthetic vowel are more similar to zero is expected given the BMP.

Since the BMP relies on a segmental correspondence between underlying and surface forms, the correspondent of an epenthesized segment is arguably zero. The spatial cues for a zero segment may be the neutral articulatory values for the speaker/language, but the durational cue would be zero. This means that phonologically epenthetic vowels would range from 0ms when *intent* was 1 to the target duration for an [i] vowel when *intent* was 0. If the level of *intent* is in between 0 and 1 then the duration of the epenthetic vowel will always be closer to zero, which is exactly what Gouskova and Hall (2009) find.

Hall (2013) follows up on Gouskova and Hall's (2009) work with a larger number of speakers. In the original study, it was found that the level of incompleteness varied from person to person and this finding was strengthened in the follow up study, most notably in relation to formant values. No difference in duration was found between the lexical (61ms) and epenthetic (60ms) vowels at the group level.[13] Hall (2013) hypothesizes that this may be a result of the faster speech rates used in data collection for this study than those used in data collection in Gouskova and Hall (2009). For this reason, the simulation focuses on the formant values.

When comparing the mean value of epenthetic versus lexical [i], Hall (2013) groups speakers into two categories: dramatic difference and non-dramatic difference. She further claims that the non-dramatic difference ranges from speakers with a small difference to those with no difference at all. I therefore use three groups in the simulation: DRAMATIC DIFFERENCE, SMALL DIFFERENCE, and NO DIFFERENCE. Notably, the DRAMATIC DIFFERENCE speakers all have a higher/fronter lexical [i] compared to the other speakers. This can be taken into account in a simulation by having the dramatic speaker have a different surface [i] target than the other two types of speakers. Figure 3.5 shows the

---

[13] Individual differences were not reported.

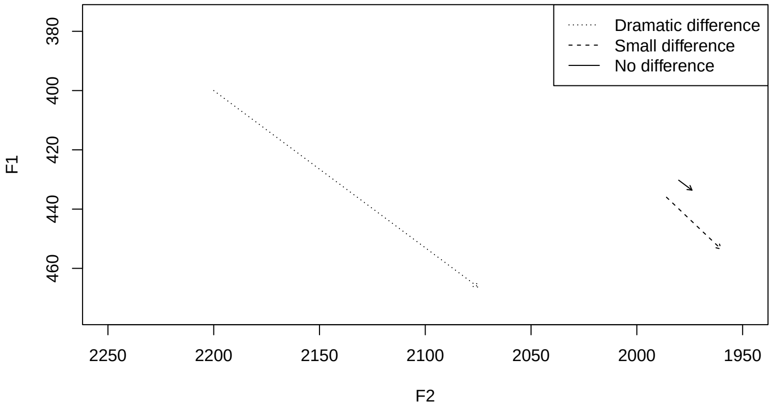


Figure 3.5: Simulated formant values for lexical and epenthetic [i] vowels based on Hall (2013) for a dramatic difference speaker, a small difference speaker, and a no difference speaker

simulated F1 and F2 values for each type of speaker. The starting point of the arrow is the lexical [i] values and the end point of the arrow is the epenthetic [i] values.

This paragraph lists the parameters used to determine the values in the scaling simulation. For the DRAMATIC DIFFERENCE speaker, lexical [i] was assigned the F1× F2 vector (400,2200) and the other two speakers were each assigned the vector (450,2000) to indicate a more central vowel. Some noise was added to the second two speakers' vectors to provide visual separation in the plot. This is because otherwise the lines on which each arrow sat would be overlapping. Since there was more movement along the F2 dimension in the Hall (2013) data, the F2 cue was determined with an $\alpha$ value of 2 while F1 was determined with an $\alpha$ value of 2.4 (since a higher $\alpha$ leads to less incompleteness). Finally, Intent levels were set to 0.5, 0.3, and 0.15 for the DRAMATIC DIFFERENCE, SMALL DIFFERENCE, and NO DIFFERENCE speakers. This is not the only way to simulate the different types of speakers. For example, it is possible to have a single intent value and instead have the $\alpha$ levels for different cues vary across speakers. There is not enough empirical data to choose between simulation strategies here. Therefore, I again emphasize that this simulation is only one way to instantiate the BMP.

Another dimension that can affect the simulation results is the spatial parameters of the underlying zero form. This also varies drastically based on what choices are made in regard to phonetic representations. If phonetic representations are acoustic targets, then a zero morpheme would have to have some type of acoustic target even if its duration was also 0. One plausible set of values is those corresponding to the default/neutral segment within the language (Archangeli, 1984; Broselow, 1984; Pulleyblank, 1988; McCarthy and Prince, 1994). In the simulation above, we chose a neutral vowel (schwa) as the F1 and F2 targets, but this is ultimately an implementation choice rather than an architectural choice. The main point continues to be about the latter, but by being explicit we can investigate the consequences of the former. Ultimately, it may make more sense to think about zero morphemes in terms of articulation. A durationless target may still have spatial targets, but they can be thought of as the neutral position of the articulators – which would also lead to the vowel being more central.

In the original study, Gouskova and Hall (2009) claim that the phenomenon at hand is a case of incomplete neutralization, but Hall (2013) suggests that what is going on is more likely to be near merger. Regardless of what it should be called, there is some type of intermediary effect between an underlying form and a surface form and this is what the BMP predicts by having access to the lexicon, the phonological grammar, and the pragmatic context in which utterances are being made. The BMP is agnostic to perception and therefore the perceptibility of a given token plays no role in the synchronic phonetic realization. This is what allows for a unified explanation of the German final devoicing, Cantonese tonal merger, and Lebanese Arabic epenthesis.

### 3.2.5 Homophone Duration Variation in English

In this section, I present a simulation that shows how the functions described in the previous section may be implemented using frequency data from Celex (Baayen et al., 1996) and duration data from the Switchboard corpus (Godfrey et al., 1992). These

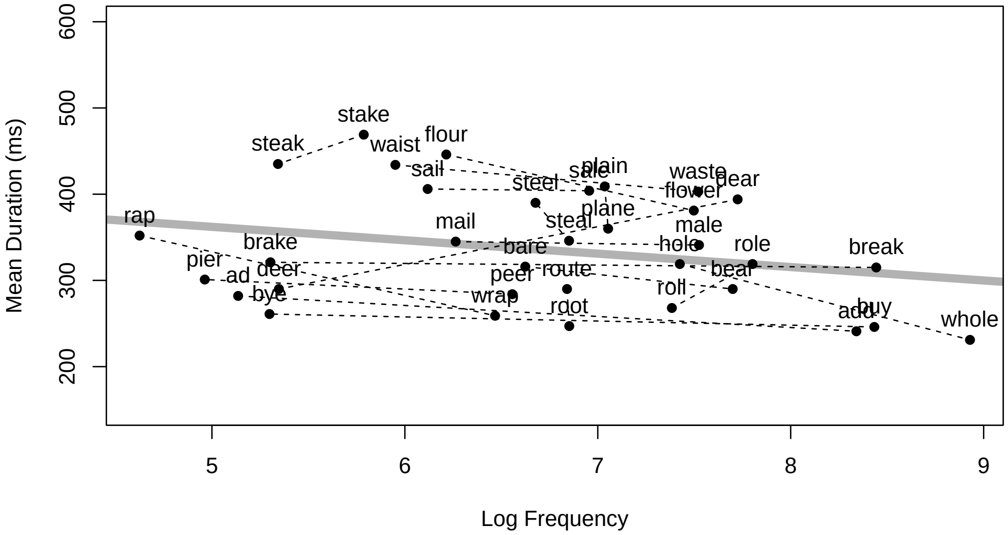


Figure 3.6: Average duration and log frequency for 17 words and their homophone twin. These data come from the Switchboard corpus (Godfrey et al., 1992). Dashed thin lines connect all homophonous pairs. The thick gray line is the output of a linear model ($\mathcal{I}$) of these points showing a general negative correlation.

data were gathered following the methodology presented by Gahl (2008, pp. 479–480) including using the time-aligned orthographic transcript originally created by Deshmukh et al. (1998). Figure 3.6 shows the mean duration and log frequency of 17 homophonous pairs. Each point represents a word in the corpus and is connected to its homophonous pair by a dashed line. While there is an overall negative correlation between duration and log frequency in the plotted pairs, it is not the case that every individual pair showed a negative relationship.[14]

This simulation uses a linear model to predict the effect of frequency on duration. In the remainder of this section, it will be referred to as $\mathcal{S}$ to reinforce the difference between the abstract model and this specific implementation. $\mathcal{S}$'s outcome variable ($y$) is duration

---

[14] Three of the most pronounced positive relationships all contain words where the same spelling results in different lexemes. For example, *deer* and *dear* have a large variation in frequency and a positive duration relationship. Following Gahl (2008) I collapsed words with the same spelling due to the difficulty of teasing apart meaning from orthography alone. Orthographic *dear* can stand for the noun or adjective. A closer analysis may show that splitting these forms apart may show duration and frequency values that do follow the general trend. This is beyond the current scope of the dissertation.

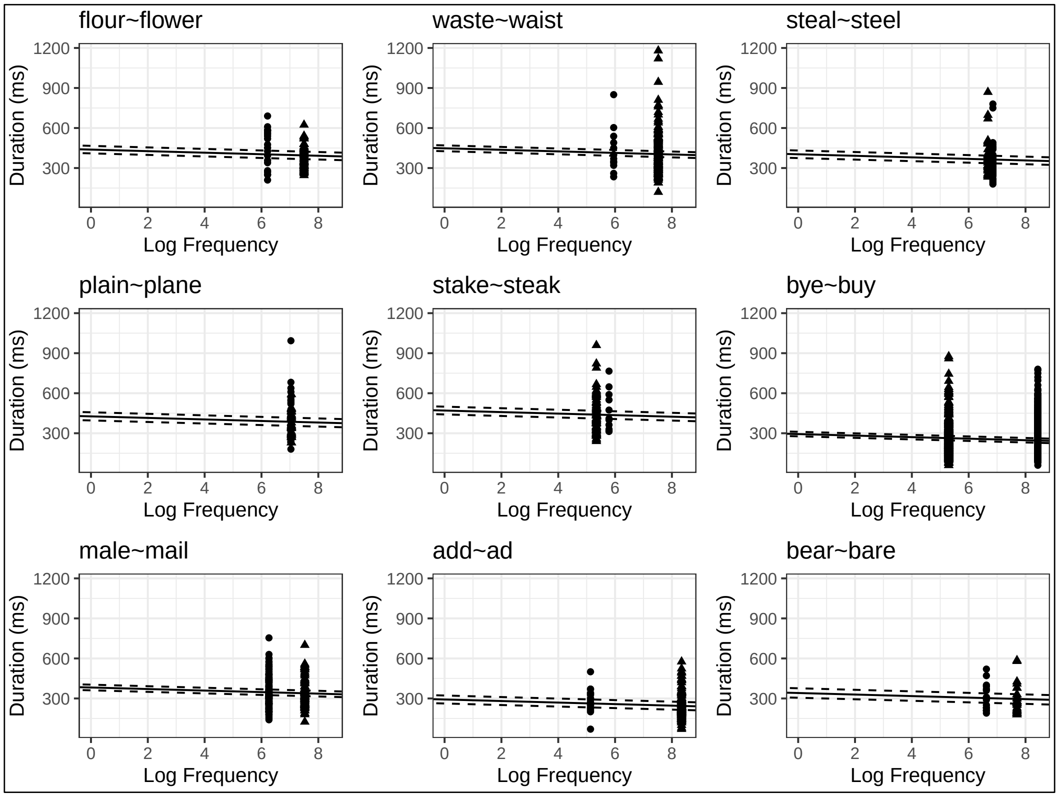


Figure 3.7: Frequency and Duration information for individual tokens of 9 randomly selected homophonous pairs. Each plot represents a single pair. The solid black lines are the predicted linear relationship for that phonological form. Dashed lines indicate 95% confidence intervals.

(ms) and has two predictor variables: log frequency and phonological form. This results in a single slope based on log frequency and varying intercepts based on phonological form and can be directly related to the functions for determining duration-influenced phonetic output. (15) shows the structure of $\mathcal{S}$ in full.

(15) $y = \beta_0 + \beta_1 \times LogFreq(x) + \sum_{i=1}^{|L|} \beta_i \times [l_i = x] + \varepsilon$

These parameters can be broken down to show how they relate to the functions above. Under the operating assumption that duration scales linearly with frequency, the underlying target value, which corresponds to the function $\tau(L, P, I)$ will be equal to the equation in (15) with $\beta_1 \times LogFreq(x)$ removed. In other words, the intercept for each phonological form is the hypothesized target value.[15]

To relate $\mathcal{S}$ to the duration scaling function $\delta(F(L))$ above, it is necessary to do some rearranging of terms. In its current form, $\mathcal{S}$ is similar to the Pierrehumbert (2002) approach. For the sake of exposition, replace $\beta_0 + \sum_{i=1}^{|L|} \beta_i \times [l_i = x]$ with a constant $k$ and remove the error term. The formula then becomes $y = \beta_1 \times LogFreq(x) + k$. Basic algebra derives an equivalent form: $y = k \times (1 + \frac{\beta_1 \times LogFreq(x)}{k})$. Since $\beta_1$, the slope coefficient, is negative and $LogFreq(x)$ is guaranteed to be non-negative, the value of $(1 + \frac{\beta_1 \times LogFreq(x)}{k})$ is guaranteed to be less than 1. As long as $\beta_1 \times LogFreq(x)$ is less than or equal to $k$, the value of $(1 + \frac{\beta_1 \times LogFreq(x)}{k})$ is also guaranteed to be greater than or equal to 0. Under these conditions, this works exactly as a scaling factor in the way necessary to implement the effect of frequency with the functions described above. The function $\delta(F(L))$ above is therefore instantiated as (16).

---

[15] Since the $\beta$ weight on frequency is a free parameter that is fit to the data, then there is nothing restricting the directionality of the effect. Previous work has related the direction of the effect to exemplar storage (Gahl, 2008) or motor practice (Bybee, 2001). Another possibility is ordering the lexicon by frequency and implementing access as a linear search (cf. Yang, 2016). In this case, more frequent words are shorter because they are accessed more quickly. This would also account for the direction of the effect. As discussed previously, the BMP, as a computational level description, has nothing to say about this issue and freely overgenerates. Other kinds of evidence or principals will be necessary to constrain it.

(16)

$$1 + \frac{\beta_1 \times LogFreq(x)}{\beta_0 + \sum_{i=1}^{|L|} \beta_i \times [l_i = x]}$$

Figure 3.7 shows the individual duration values for nine randomly selected homophonous pairs as well as the output of $\mathcal{S}$ for each phonological form. $\mathcal{S}$ has a significant effect for frequency ($\beta$ = -5.761, t=-3.561, p < 0.001). To illustrate how this works, consider the pair bye∼buy. $\mathcal{S}$ predicts an intercept of 293.852 for this phonological form and therefore provides the equation $\hat{y} = 293.852 - 5.761 \times LogFreq(x)$. This can now be translated into the form $PR_{dur} = \tau(L, P, I) \cdot \delta(F(L))$. The term $\tau(L, P, I)$ equals 293.852. For the form *bye*, the LogFreq is equal to 5.30, making $\delta(F(L))$ equal to $(1 + \frac{-5.761 \cdot 5.3}{293.852}) = (1 + \frac{-30.5333}{293.852}) = (1 - 0.1039071) = 0.896$. Using the same method, $\delta(F(L))$ for *buy* is $0.835$. These values therefore predict that the frequency influenced duration value for *bye* should be $293.852 \cdot 0.896 \approx 263$. The mean duration for all tokens of *bye* in the data set is 261 ms. The frequency influenced duration value for *buy* is $293.852 \cdot 0.835 \approx 245$. The mean duration for all tokens of *buy* in the data set is 246 ms.

Success on an individual pair does not tell the entire story. To begin with, word frequency is not the only factor that affects duration. Second, the previous paragraph pairs the predicted value with the mean value for a given lexical item. Visual inspection of Figure 3.7 clearly shows that the data for each lexical item is quite spread. This suggests that the error term in $\mathcal{S}$ can be directly thought of as the aspects of production other than frequency that influence duration for a given production. Therefore, specific results of $\mathcal{S}$ presented here should be interpreted conservatively.

Rather than focus on perfect prediction, the goal here was to show how the architecture of the BMP can be used to simulate this type of frequency and duration data. The assumptions being made in this simulation are: 1) the phonology maps discrete inputs to discrete outputs; 2) there are multiple inputs to the phonetic module: the target lexical item, the phonological map, the intent value, and frequency information; 3) the lexical item, phonological map, and intent are used to produce a phonetic representation; 4) this representation is further scaled based on frequency information for individual lexical items. Consequently, adopting an exemplar model or gradient phonology is not necessary to account for the types of duration effects that Gahl (2008) and others have documented.

### 3.2.6 Optionality in Warao and English

Returning to the data in (3-4), we can illustrate how the formula in (3.2.1) accounts for variation in instances of global optionality. To begin, let us assume that the underlying form of the morphemes contains the $[-\texttt{voice}]$ form and the grammar contains a process described with the rule $[-\texttt{voice}] \rightarrow [+\texttt{voice}]$. Furthermore, we assume that $i = 0$, indicating no intent to maintain an underlying contrast, and $\alpha = 1$. This is done to remove the influence of these parameters since they do not affect the larger point being made. Finally, for simplicity, assume that $[-\texttt{voice}]$ maps to $0$ in the cue space and $[+\texttt{voice}]$ maps to $1$. To account for the case where surface voicing is observed due to phonological knowledge being applied, $v$ parameter is set to $0$. Since the phonetic form $c = 1$ is equal to the value for $[+\texttt{voice}]$, we expect the forms [barobarera] and [abaubute]. Changing the $v$ parameter to $1$ results in the opposite. In this instance, the phonetic form $c = 0$ is equal to the value for $[-\texttt{voice}]$, and the forms [paroparera] and [apaupute] are expected.

When applying the scaling formula instantiation of the BMP over words, global optionality emerges based on how the $v$ parameter is set. As discussed in previous sections, a mechanism for optionality should also explain the rates at which each form occurs. To answer this question, $v$ can be thought of as the sample space for a probability distribution $V$ which itself is governed by a set of factors. Now, we turn to a different problem: the grain size of application.

Word-size chunks have been used to explain the various phenomena looked at so far, but the underlying math translates directly into different grain sizes. Rather than planning based on word-sized chunks (or even larger syntactic units), planning could occur on a morpheme by morpheme basis. This would result in the blending procedure occurring over smaller windows.

The example given for English flapping was the multi-morphemic word *marketability*. Under the assumption that production occurs linearly, there are four possible groupings of morphemes for this word. The first would involve all three morphemes being grouped together. This, on its own, would result in the global optionality pattern with either the form where there is flapping at both possible loci or no flapping at all. The second would involve the first two morphemes being grouped together and the second morpheme being grouped on its own. This would result in the complete local optionality pattern shown in (5). The first chunk would provide the correct structural description for flapping to occur. Therefore it would produce a flap when $v = 0$ and no flap when $v = 1$. The second chunk, the morpheme /əti/, will also variably produce a flap. Combining the two chunks therefore provides a way to generate all four possible forms. The final two groupings would involve /maɹket/ being chunked on its own, therefore never providing the necessary structural description for flapping to occur. One of these groupings would further group /əbɪl/ and /əti/ together, which provides no new point of application, and the other separates each morpheme into its own chunk. Table 3.3 summarizes these predictions.

| Planning Chunks | Possible Outputs | | | |
|---|---|---|---|---|
| (/mɑrkətəbɪləti/), () | [maɹkətʰəbɪlətʰi] | | | [maɹkəɾəbɪləɾi] |
| (/mɑrketəbɪl/), (əti) | [maɹkətʰəbɪlətʰi] | [maɹkəɾəbɪlətʰi] | [maɹkətʰəbɪləɾi] | [maɹkəɾəbɪləɾi] |
| (/mɑrket/), (/əbɪləti) | [maɹkətʰəbɪlətʰi] | | [maɹkətʰəbɪləɾi] | |
| (/mɑrket/), (/əbɪl/), (/əti) | [maɹkətʰəbɪlətʰi] | | [maɹkətʰəbɪləɾi] | |

Table 3.3: Possible planning groupings for the multimorphemic word *marketability* as well as the possible phonetic outputs that each grouping would predict.

Planning cannot occur on a strictly morpheme-by-morpheme basis as this would prevent any morpho-phonological processes from ever occurring. This is evident based on the last row of Table 3.3 which prevents the first /t/ from ever flapping since its point of application is only created when /mɑrket/ and /əbɪl/ are combined. The second row then appears to be the only chunking strategy that, when combined with the scaling formula, can predict the entire suite of possible forms. This suggests the possibility of a buffer that can vary in the number of morphemes being processed at any given time.

Planning has previously been proposed as a way to account for optionality, though typically in cases of external sandhi/post-lexical processes (Wagner, 2012; Tanner et al., 2017; Kilbourn-Ceron and Goldrick, 2021; Du and Durvasula, 2024a). In these cases, it is typically framed in terms of lexical access, but a lexical access (or, more specifically to the case at hand, morpheme access) account alone cannot account for the flapping facts. This is because the triggering conditions for the morpheme /əti/ are within the morpheme itself and don't depend on its concatenation with any other morphemes. Variability in flapping here therefore cannot simply be due to processing.

## 3.3 General Discussion

The final type given to the phonetic production function is $A_{\text{BP}} :: L \rightarrow P \rightarrow I \rightarrow F \rightarrow \{PR\}$. This is a curried function. What this means is that the lexicon, phonology, intent, and frequency are all inputs to the function, and each argument can be given one at at a time. A function of arity $n$ is said to be saturated if it has received $n$ arguments. This perspective allows for the description of a chain of partially saturated production functions:

(17) a. $A_{\text{BP}} :: L \rightarrow P \rightarrow I \rightarrow F \rightarrow \{PR\}$

b. $A^{l}_{\text{BP}} :: P \rightarrow I \rightarrow F \rightarrow \{PR\}$

c. $A^{l,p}_{\text{BP}} :: I \rightarrow F \rightarrow \{PR\}$

d. $A^{l,p,i}_{\text{BP}} :: F \rightarrow \{PR\}$

These functions can be interpreted such that (17b) is the production function given a specific lexicon $l$ in the set of all possible lexicons $L$, (17c) is the production function given a specific lexicon and a specific phonology function $p$ in the set of all possible phonology functions $P$, and (17d) is the production function given a specific lexicon and phonology, as well as a specific intent value $i$ in the set of all possible intent values $I$.

Consider another possible type, $A'_{\text{BP}} :: (L, P) \to (I, F) \to \{PR\}$. Here, the inputs are split into two tuples, one containing the lexicon and phonology and one containing the intent and frequency. This essentially can be viewed as the split between knowledge of language and knowledge about language. Since the act of production involves many factors beyond what has been discussed in this chapter, it is possible to switch $(I, F)$ to a cover type $E$ which stands in for all of the information other than the lexicon and phonology that go into the production process. With this in mind, it is possible to have a partially saturated function with type $A'^{l,p}_{BP} :: E \to \{PR\}$. Ignoring $E$ completely here would result in a set of phonetic outputs influenced only by the lexicon and phonology.

Why does this matter? While it may appear that the phonetics module has been complicated by adding extra material to its input (the lexicon, intent, frequency), I argue instead that it has been simplified. Typed functions allow for the larger production process to be broken down into its smaller pieces. What looks like a complicated system is instead the interaction of many different simple systems. In this way, type analysis is a new tool by which one can better understand the modular relationship between phonetics and phonology.

One consequence of this simplicity is that the BMP may appear too flexible, allowing all kinds of interactions that are not manifest in the phonetics-phonology interface. In general, models of the phonetics-phonology interface will have the same flexibility due to the level of analysis at which they are couched. For example, the feed-forward model itself is similarly “too flexible.” It can similarly be instantiated in many possible ways (e.g., discrete or continuous representations and computations). Nonetheless, this level of analysis still allows one to contrast the capacities and properties of different models. For example, as we have shown, the BMP alleviates problems inherent to the feed-forward model.

Any particular theory of the interface will necessarily constrain the possibilities in some significant way. One may ask what kind of criteria would be used to constrain the BMP. The answer is evidence from any scientific investigation that can be brought to bear upon this question. For instance, I have reviewed careful phonetic experimentation which has yielded evidence for the importance of extra-grammatical factors such as intent to maintain an underlying contrast and frequency on production. Additionally, other experimental work has shown the importance of maintaining categorical phonological knowledge (Du and Durvasula, 2024a; Mai et al., 2022). Considering van Rooij and Baggio's (2021) characterization of experimental and theoretical cycles in scientific research, this proposal can be thought of as a response to an experimental cycle dominated by the feed-forward model of the interface. The proposal in this chapter takes a step towards a new theoretical cycle, which can then lead to a new experimental cycle conducted within the perspective offered by the BMP.

## 3.4 Summary

What has been shown in this chapter is that phenomena like incomplete neutralization and systematic variation in homophone durations do not necessarily require gradient phonological knowledge. What has not been shown (or argued for) is that phonological knowledge must necessarily be discrete. Based on the discussion of auxiliary assumptions in Chapter 1, I showed that it is not enough to make the argument that certain phonetic production data implies non-categorical phonology. Instead, either of the two inferences are possible: phonetic production data + the modular feed-forward model of the interface implies non-categorical phonology or phonetic production data + the blueprint model of production does *not* imply non-categorical phonology. In both cases, it is the structure of the interface that changes what we can infer, thus highlighting its importance in understanding the relationship between phonological and phonetic knowledge.

The primary goal in this chapter has not been to assert that gradient phonological phenomena do not exist, but rather to highlight the fact that gradient measurements do not automatically imply gradient knowledge since there may be alternate ways to account for this gradience (such as with the structure of the interface). There may be other phenomena that call for more gradient representations, but it should be noted that even eliminativist exemplar models that completely abandon symbolic representations run into issues accounting for systematic generalizations. Goldrick and Cole (2023) suggest a hybrid exemplar approach that takes influence from symbolic generative phonological theory just as much as it takes influence from eliminativist exemplar theories. The exact structure of the lexicon under the architecture of the BMP remains underdetermined and can therefore support a hybrid exemplar approach, especially one that maintains a symbolic level of phonology as is the case for the example hybrid model shown in Goldrick and Cole (2023, figure 4).

The type analysis of the BMP provided in this dissertation, along with simulations in the case studies, provide multiple entry points for further investigation of the BMP on its own terms or in comparison to other models of the interface. I began some comparison with the research using dynamical systems because of its significant influence on formulating the BMP, but Jurafsky et al. (2002, figure 3) and Shaw and Tang (2023) are other possible examples of research that could be instantiations of the BMP. The BMP also makes predictions about the phonetic realizations of other kinds of phenomena including deletion, the realization of absolutely neutralized segments, morphological boundary effects, and potentially many more. For example, Danis (2020) finds that there is a small but significant difference in vowel length between faithful short vowels and short vowels derived from hiatus resolution in Yoruba. The short vowel derived from deleting a neighboring identical vowel is on average 12ms longer than the faithful short vowel. Through the lens of the BMP, this appears to be retention of the UR target duration for the deleted segment being retained in the output.

# Chapter 4

# The Content of Phonological Representation

This chapter focuses on phonological representations and what type of information needs to be directly stored in long-term memory. The primary case of interest is whether dynamic articulatory plans need to be stored directly, or if they can be reasonably inferred from segmental representations of lexical items. While basically every theory related to generative phonology assumes some type of sub-segmental representation, segments are used since they are most easily represented and any feature-based interpretation can be computed directly from the symbolic representation using only logical disjunction (discussed more in-depth below). Articulatory Phonology is a theory of phonology based entirely around gestural representations and is implemented with nonlinear dynamical equations in opposition to classic symbolic computing. Therefore, this chapter begins with a review of Articulatory Phonology. It then introduces model-theoretic phonology as a way to approach the computational properties of phonological generalizations as well as a tool for doing cross-theory comparison. Model theory and logic are used to provide an equivalency between string representations and the dynamic coupling graph representations used in Articulatory Phonology. Given that the computational power required to state this equivalency is relatively weak, it is argued that phonological representations need not encode this type of information directly as it can be inferred from static symbolic representations.

## 4.1 Articulatory Phonology

Articulatory Phonology can largely be seen as a response to the failure to find invariant, context-independent, static properties in the speech signal. To square these findings with the fact that language acts as a discrete, combinatoric system, Browman and Goldstein (1989, 1992) proposed that the invariant properties of speech were gestural representations that are governed by interconnected dynamical systems. A gesture can be idealized as the formation or release of a constriction in the vocal tract. Formally, this is modeled as a dynamic equation in the task dynamical model of Saltzman and Munhall (1989). Gestures are represented using a critically damped mass-spring system and second-order dynamical equations of the form $m\ddot{x}+k(x-x_0)=0$, where $m$ and $k$ correspond to the mass of the articulator (always modeled as 1 in Articulatory Phonology) and $k$ corresponds to the stiffness of the spring (how quickly it reaches its target equilibrium), $x_0$ is the position where equilibrium is reached, and $x$ is the articulators current position. While these types of equations describe *continuous* motion, they are *static* in the sense that the parameters for mass, stiffness, and target do not change over time. It is only the position of the articulator $x$ that varies over time. Therefore, invariance appears in the higher-level properties of the dynamics rather than the lower-level outcome of the dynamics.

Since the ultimate goal of a gesture is to create a constriction of a certain size at a certain location, the equations directly govern the qualities of the constriction and not the movement of the articulatory organs. This separates how a goal is achieved from the specific achievement of the goal. For example, one gesture involved in producing the bilabial stop [p] is a lip closure gesture. In the words *leap* [lip] and *lap* [læp], the lip closure gesture will show some variation in how it is achieved due to the different jaw heights used while producing the high vowel [i] and the low vowel [æ]. In *lap,* the jaw is in a lower position and requires extra movement to achieve the goal of bringing the lips fully together. In both cases, the end result is the same, but the path to achieve it is different. A gestural representation based on task dynamics can therefore explain both contextually influenced variation (coarticulation) as well as context invariant outcomes (based on the parameter settings for the tasks).

Gestures can further be defined based on which articulatory organs are involved in making the constriction. It is worth clarifying that the articulators involved in the constriction are only a subset of those involved in the overall gesture. Returning to the example from the previous paragraph, the articulators involved in making the constriction are the lips, but to do this, both the lips and the jaw need to move. In other words, the goal is to move the lips into a certain position, but more articulators than just the lips are involved in reaching that state. The articulatory organs that are thought to be involved in the making of constrictions include the lips, the tongue tip (TT), the tongue body (TB), the velum (VEL), and the glottis (GLO). The constriction degree (CD) and constriction location (CL) are represented qualitatively as tract variables. Since the task dynamic model is a two-dimensional representation of the vocal tract, it can be useful to think of constriction degree as the y-dimension parameter and constriction location as the x-dimension parameter. It is possible to define both tract variables for each organ that is involved in a constriction, but in practice, this is not the case. Constrictions made with the lips and tongue are defined on both dimensions while the constrictions made with the glottis and velum are restricted to one dimension (aperture).

In dynamic equations, tract variables are represented as continuous numerical values. Without any further explanation, this would predict an infinite number of contrasts. It has been hypothesized that speakers attempt to attune their speech gestures to the larger speech community based on the acoustic signal (Goldstein and Fowler, 2003). Since there is a nonlinear relationship between constrictions and acoustic outcome (Stevens, 1989), the tract variable space can be discretized based on stable regions where differences in tract variable values do not have large effects on the acoustic outcome. These discrete zones are where speakers are most likely to settle during the attunement process. For constriction degree, there are five categories for tract variables: closed, critical, narrow, mid, and wide. Constriction location tract variables are named based on the familiar areas used in International Phonetic Alphabet descriptions (e.g., labial, velar, alveolar, etc…).

The description of gestural representations so far is enough to explain the spatial properties of gestures. We can describe an [s] using the tract variables TTCL: alveolar and TTCD: critical, but we are missing any temporal information. The task dynamic model explains how articulators move in both space and time. Temporal information is therefore represented as an activation period, or more informally, the points in time in which the tasks of a given gesture are controlling the dynamical system. The simultaneous activation of different gestures is responsible for the patterns of co-articulation observed since articulators will act in different ways depending on what gestures are active at any given time. Articulatory Phonology is therefore able to explain high-level contrast effects based on which tract variables are chosen, while simultaneously explaining low-level phonetic effects based on the output of the dynamical system. Unlike other theories of phonology, the theory of Articulatory Phonology specifically rejects a modular approach to phonetics and phonology.

It is possible to represent both the spatial and temporal information of a given phonological form with a gestural score. Gestural scores use boxes to show the tract variable values and activation periods for each gesture. The horizontal position of the boxes shows when each gesture is active, while the information inside of each box indicates what the current tract variables are. For example, in the word *mash* [mæʃ] the initial [m] is made up of a lip closure gesture and a wide velum gesture. This overlaps with the vocalic [æ]

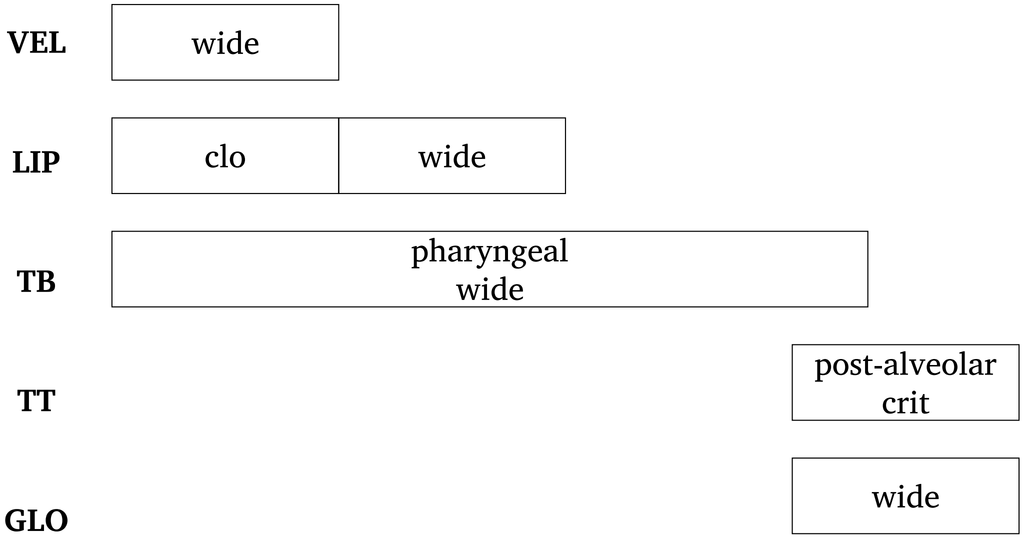


Figure 4.1: Gestural score for *mash*.

gesture which is made by indicating a wide tongue body gesture at the pharyngeal position. The final [ʃ] sound is made with a critical tongue tip gesture at the post-alveolar position as well as a simultaneous wide glottal gesture. In Articulatory Phonology, it is assumed that the natural state of the glottis for speech is close together to induce voicing. It is therefore only devoicing gestures that are specifically encoded into the gestural representations. Figure 4.1 shows the complete gestural score for *mash*.

Since there is no segmental unit in this type of gestural representation, the multiple gestures that make up a sound like [m] can only be indirectly inferred due to simultaneous activation and/or overlapping activation. Browman and Goldstein (1990) hypothesized that there may be consonant and vowel tiers that each gesture is connected to, thus allowing for higher-level organization. This hypothesis led to further investigation into inter-gestural timing relationships. In the early days of Articulatory Phonology, individual gestures were coordinated based on symbolic rules to form inter-gestural CONSTELLATIONS which included phase relations between connected gestures (Browman and Goldstein, 1995). These symbolic rules betray the dynamic philosophy at the heart of the theory. Recently, Articulatory Phonology has replaced constellations with gestural

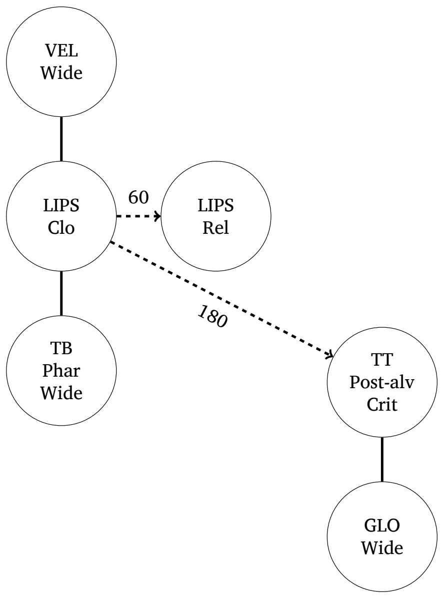


Figure 4.2: Coupling graph for *mash*.

coupling graphs. Coupling graphs serve a similar purpose, in the sense that they encode phase relations between connected gestures. The primary difference is that they are themselves dynamical oscillators that automatically generate gestural scores. Furthermore, the addition of coupling graphs resulted in theoretical clarity. As Nam (2007a, p. 38) writes, “In this model, planning oscillators associated with the set of gestures in a given utterance are coupled in a pairwise, bidirectional manner specified in a coupling graph (or structure) that is part of the lexical specification of a word.” Lexical entries in the theory are coupling graph structures where each node of the graph corresponds to specific parameters for the dynamical equations governing articulatory movement and each edge corresponds to the timing relation between gestures in terms of phasing.

Figure 4.2 shows the coupling graph for *mash*. Each gesture is represented as a circle. The lines connecting each gesture indicate a coupling relationship. Undirected lines are indicative of in-phase or simultaneous coordination. Directed lines indicate non-simultaneous coordination and are specified with a phasing degree.

Saltzman et al. (2008) proposed a dynamic model of inter-gestural timing by assigning each gesture to an internal timing oscillator. Each timing oscillator is further dynamically coupled to other oscillators within the system. The coupling relations are based on phasing. Research in coordination has shown that there are two phasing relationships that are easily performed without learning: in-phase (0°) and anti-phase (180°) (Turvey, 1990). Despite this fact, other types of non-simultaneous coordination can be learned and used. Other "eccentric" couplings can also be learned. In the graph for *mash* the Lips-CLO and Lips-REL gestures are coupled at 65° so that the release follows the closure. Previous research supports the splitting of stop (and possibly other consonant) gestures in this way (Browman, 1994). Goldstein (2011) also discusses eccentric phasing of 30° for coda consonant clusters. The timing model of Saltzman et al. (2008) hypothesizes that speakers exploit these modes of coordination when figuring out how to time gestures.

Furthermore, this model of inter-gestural timing has been used to argue for syllable structure within the Articulatory Phonology framework. While both in-phase and anti-phase coordination relationships are easy to learn, in-phase coordination is more stable. The oscillator clocks for onset consonants are hypothesized to be coupled in-phase with vowel gestures while the oscillators for coda consonants are hypothesized to be coupled anti-phase with the vowel gesture (Browman and Goldstein, 1989; Nam and Saltzman, 2003; Nam, 2007b). These coupling relations also help explain the C-center effect (Browman and Goldstein, 1988). The C-center effect has to do with the timing relationship between onset consonants and vowels. Despite the number of consonants in the onset of a syllable, the vowel gesture will always be timed so that it begins in the center (middle point) of the combined onset gestures. Nam and Saltzman (2003) and Nam (2007b) use a computer simulation to show that this is exactly what the coupled oscillator model predicts.

Returning to the graph for *mash*, the lip closure gesture, the vocalic tongue body gesture, and the velum widening gesture are all coupled in-phase with one another. Crucially, the velum gesture is coupled directly with the lip closure gesture. Since the coupling graphs can be considered part of the phonological representation, we now have a way to directly connect the two gestures involved in producing an [m] sound in a way that was not possible with the gestural score. The different coupling relations can therefore be used as another way to indicate contrast. If we changed the graph such that the velum gesture was coupled directly to the vowel gesture, we may end up with a percept like [bæ̃ʃ], where the nasalization is timed relative to the vowel as opposed to the stop gesture. Using coupling relations to determine contrast runs into a similar problem as with the tract variables of the gestural score. Even if we limit degrees to be discrete, there is still a large space of values that can be used to form contrasts, but is likely that this space can also be split up so that there are stable zones (in-phase and anti-phase being the two obvious starting points).

Unlike traditional phonological analysis, Articulatory Phonology provides an account of many allophonic processes without needing symbolic rules. For example, within Articulatory Phonology, anticipatory vowel nasalization in English is not due to a rule (or set of constraints) that ensures that vowels preceding nasal consonants are marked with a $[+\texttt{nasal}]$ feature, but instead due to the coupling graph structure of the lexical entries. Furthermore, Articulatory Phonology has been argued to provide a better account for various types of speech phenomena such as gestural hiding (Browman and Goldstein, 1990; Beckman, 1996), speech errors (Pouplier, 2003; Pouplier and Goldstein, 2005), and developmental data (Browman and Goldstein, 1992; Goldstein and Fowler, 2003). Contrast can also emerge from gestural representation in the following ways: presence or absence of a certain gesture, parameter differences among gestures, or timing differences among the organization of the same gestures.

With that in mind, categorical alternations are notably difficult for Articulatory Phonology. In her survey article on AP, Hall (2010) writes, “At present, there is no real consensus on how to deal with categorical alternations in AP,” and additionally references Zsiga’s (1997) work which blends categorical phonology with AP style phonology. One attempt at accounting for categorical alternations is Goldstein's (2011) account of the past tense alternation in English. Crucially, he relies on Gafos and Benus's (2006) dynamical implementation of an OT-like grammar to account for different coordination patterns. What I will go on to show in the remaining sections of this chapter is that the gestural representations used in Articulatory Phonology can be inferred from segmental string representations. This provides a way to compare the symbolic account of the past tense alternation with the dynamical account of the past tense alternation. In both cases, the "epenthetic" alternant emerges when there is a local substructure of two identical elements that leads to the output structure where epenthesis has occurred. Furthermore, any gestural representation can be translated back into a string representation without loss of any information. This suggests that a model of Articulatory Phonology with a grammar that selects between different coordination relations makes the same distinctions as a symbolic phonology that manipulates discrete representations.

## 4.2 Model-Theoretic Phonology

This section outlines the model-theoretic approach to studying phonological knowledge. In general, phonological structure can be broken down into three aspects: elements of the structure, order of the elements, and properties of the elements. Different theories of phonological representation make different claims about all these aspects, but no theory of representation can do without them. Finite Model Theory provides a way to formalize different aspects of phonological structure. Furthermore, this structure can then be evaluated using mathematical logic. These tools provide a way to clearly state phonological generalizations about both "rules" and "representations" (Rogers et al., 2010, 2013; Strother-Garcia, 2019; Oakden, 2020; Chandlee and Jardine, 2021; Jardine et al., 2021; Nelson, 2022; Heinz, Forthcoming). The language used to describe this is intentionally neutral as it can include segments, feature bundles, x-slots, morae, autosegments, particles, gestures, elements, and any other type of information one believes should be encoded in a phonological representation. This section discusses how this is done in detail and ends with an explanation of how model theory further can act as a shared meta-language for doing cross-theory comparison.

### 4.2.1 Phonological Representations

Suppose we wanted to encode a standard segmental theory of phonological representation (i.e. - "beads on a string"). This type of representation is more or less the standard string model used in formal language theory and other areas of computer science (Büchi, 1960). The model signature is given below:

$$\langle \lhd, \{R_\sigma \mid \sigma \in \Sigma\}\rangle \tag{4.1}$$

Here, $\lhd$ is a binary ORDERING relation. $x \lhd y$ means element $y$ is the successor of element $x$. $R_\sigma$ is a unary LABELING relation drawn from a set of possible labels $\Sigma$. For this example, suppose $\Sigma$ is the set of all possible IPA symbols. The relation $R_{\text{ʃ}}(x)$ assigns the property "ʃ" to domain element $x$. Suppose we wanted to describe a model for the English word *dish* given the model signature in (4.1). Figure 4.3 shows the mathematical form of the model (left) and graphical form of the model (right).

$$\begin{aligned}\langle D &= \{0,1,2\},\\ \lhd &= \{(0,1),(1,2)\},\\ \mathrm{d} &= \{0\},\\ \text{ɪ} &= \{1\},\\ \text{ʃ} &= \{2\},\\ x &= \{\}; x \in \Sigma \backslash \{\mathrm{d}, \text{ɪ}, \text{ʃ}\}\rangle\end{aligned}$$

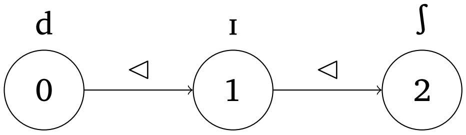


Figure 4.3: Model for the English word *dish* using the segment signature. On the left is the mathematical object itself while on the right is a visualized form of the model.

Suppose instead that we wanted to maintain linear ordering, but instead of segments being directly encoded in the representation we wanted the primitive properties to be *features* such that segments emerge from specific combinations of the features. The only change we need to make then is to say that the set of all possible labels ($\Sigma$) contains phonological features rather than IPA symbols as before. In contrast with the previous example, let us use $\mathcal{F}$ in place of $\Sigma$ to represent the set of all phonological features. Our model signature then becomes:

$$\langle \lhd, \{R_f \mid f \in \mathcal{F}\}\rangle \tag{4.2}$$

The only change from (4.1) is that our labeling relation now provides feature properties rather than segmental properties. $R_{\text{VOICE}}(x)$ therefore assigns the property of being [+voice]/[voice] to domain element $x$. Suppose we wanted to describe a model for the English word *dish* given the model signature in (4.2). Figure 4.4 shows the mathematical form of the model (left) and graphical form of the model (right). For this example, and without loss of generalization, the feature set $\mathcal{F}$ is limited to VOICE, SONORANT, CONSONANT, CONTINUANT, CORONAL, ANTERIOR, SIBILANT, HIGH.

$$\begin{aligned}
\langle D &= \{0,1,2\},\\
\lhd &= \{(0,1),(1,2)\},\\
\textsc{Voice} &= \{0,1\},\\
\textsc{Sonorant} &= \{1\},\\
\textsc{Consonant} &= \{0,2\},\\
\textsc{Continuant} &= \{2\},\\
\textsc{Coronal} &= \{0,2\},\\
\textsc{Anterior} &= \{0\},\\
\textsc{Sibilant} &= \{2\},\\
\textsc{High} &= \{1\},\rangle
\end{aligned}$$

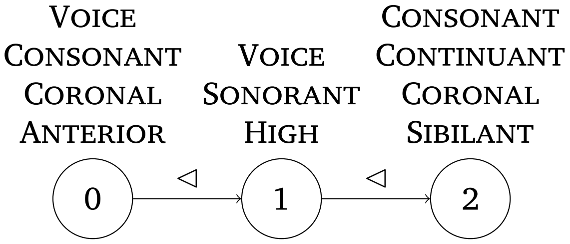


Figure 4.4: Model for the English word *dish* using the feature signature. On the left is the mathematical object itself while on the right is a visualized form of the model.

The graphical representations of the models only differ in the properties assigned to the individual elements, but the number of elements and ordering of the elements are the same across models. Given these different structures, we can also evaluate how to pick out substructures of the larger model. Substructures that are of interest to phonologists largely are groups of individual domain elements that have a shared property, i.e. - *natural classes*, or sequences of elements that could either act as the *structural description* of a phonological process or be marked as *phonotactially illicit*. The model-theoretic approach to phonology is not limited to building phonological structure, as discussed so far, but can also interpret phonological structure using mathematical logic.

## 4.2.2 Phonological Constraints

First-Order logic (FO) is used to describe the truth conditions of structures. Recall that a logical language is the combination of first-order logic with a specific model signature that provides the non-logical symbols. An $S$-structure $A$ under model signature $S$ can then be evaluated based on the properties that it holds. Suppose that $A$ is our featural model of the English word *dish* and we have the following $S$-sentences:

$$\phi_1 \overset{\text{def}}{=} \forall x[(\neg\textsc{Sonorant}(x) \wedge \neg\exists y[x \triangleleft y]) \rightarrow \neg\textsc{Voice}(x)] \tag{4.3}$$

$$\phi_2 \overset{\text{def}}{=} \forall x[\textsc{Consonant}(x) \rightarrow \exists y[\neg\textsc{Consonant}(y) \wedge x \triangleleft y]] \tag{4.4}$$

$$\phi_3 \overset{\text{def}}{=} \exists x[\neg\textsc{Consonant}(x)] \tag{4.5}$$

Each $S$-sentence above can be thought of as a phonotactic constraint. (4.3) is a constraint that says the final element must not be a voiced obstruent,[1] (4.4) is a constraint that says

---

[1] Actually, I am being a little sloppy here because $x$ doesn't necessarily need to be the final element. Suppose we had 3 domain elements {0,1,2} and the only specified ordering relation was $0 \triangleleft 1$. The constraint $\phi_1$ would then only be satisfied if both domain elements 1 and 2 did not have the property of being a voiced obstruent. But in this sense, domain element 2 is not "final" and its unclear if we should call domain element 1 "final" either. There's no default requirement that the structure be *connected* but with a further constraint like $\phi_4 \overset{\text{def}}{=} \forall x, y[\ \triangleleft^*\ (x, y)]$ we can eliminate non-connected structures. The predicate $\triangleleft^*$ here refers to the transitive closure of the successor relation. This approach to connected structures is based on Chandlee et al. (2019).

any consonant element must be immediately followed by a non-consonant element, and (4.5) is a constraint that says there must be at least one non-consonantal element in the structure.

The Theory of $A$ is all the $S$-sentences $\phi$ that $A$ satisfies. Therefore, when $A$ is the structure for English *dish* built using the feature model signature, its Theory is $\{\phi_1, \phi_3\}$. In phonological terms, the Theory of a structure can be thought of as the set of phonotactic constraints that it satisfies.[2]

In the opposite direction, the Spectrum of an $S$-sentence $\phi$ is all the $S$-structures $A$ that satisfy $\phi$. Given the $S$-sentence $\phi_1$ in (4.3), its Spectrum would include the structure for English *dish* that has been discussed. But it would also contain an infinite number of finite structures. The model signature provides a way to define finite structures but puts no limit on the number of structures so definable. Furthermore, the sentence $\phi_1$ puts no restriction on the size of the model it is interested in. All that $\phi_1$ says is that the final domain element should not have the property of being a voiced obstruent. Therefore, the set of structures it satisfies includes any structure that doesn't end with a voiced obstruent. This includes structures for real English words like *dish*, but also non-words like *dɪ*$^{1000}$*ʃ* where the superscript 1000 refers to 1000 repeats of the vowel ɪ. The phonological interpretation of Spectra then are structures that satisfy a phonotactic constraint. When considering individual constraints, phonologists often want to find structures that do not satisfy a constraint (say, to assign a weight or violation mark). This can be easily accomplished by negating a constraint.

Putting these ideas together, the Model of a given set of $S$-sentences $T$ is the set of $S$-structures that satisfy all the $S$-sentences in $T$. If $T$ is the set of sentences from above (including $\phi_4$ from footnote 1), i.e. - $\{\phi_1, \phi_2, \phi_3\phi_4\}$, then the MODEL for $T$ is all and only the structures which satisfy all those constraints. Thus, the structure for *dish* is not a model

---

[2] It could also very well be the set of morpheme structure constraints that the structure satisfies. The theoretical connection depends on whether or not the analyst wants the property to hold at the level of UR or SR and not the mathematical definition.

for $T$ because it violates $\phi_2$. Phonologically, $T$ can be thought of as all the phonotactic constraints in a language which makes the MODEL of $T$ all the possible phonological structures that satisfy all the phonotactic constraints. Note, the MODEL then does not include all and only the "real" words of a language but instead all and only the "possible" words of a language. It therefore provides a way to describe phonotactic knowledge.

### 4.2.3 Phonological Transformations

The discussion up to this point has been framed around describing static properties of structures. While this is useful for discussing phonotactic knowledge and morpheme structure constraints, it ignores a central aspect of phonological theory which is the transformation of an input structure into a corresponding output structure. The model-theoretic approach is well-equipped to discuss phonological transformations as well. There is a history of theory *interpretation* in model theory where one structure $A$ is defined in terms of a second structure $B$ (Hodges, 1993, 1997). This idea was extended by computer scientists interested in characterizing string-to-string functions (Courcelle, 1994; Engelfriet and Hoogeboom, 2001). The combination of these two approaches provides a way to talk about *structure to structure* functions where an output structure is returned based on properties of an input structure using monadic second-order logic. This provides flexibility for talking about phonological transformations over different types of representation.

As a general example, consider the phonological rule $a \rightarrow b\ /\ c\_d$. As a function, this maps any string containing the substring $cad$ to a string where everything remains the same other than the substring which is expressed as $cbd$. In other words, $\Sigma^*cad\Sigma^* \mapsto \Sigma^*cbd\Sigma^*$. The model-theoretic description of this mapping requires describing what the input properties of a string are that lead to specific output properties. Assume the segmental alphabet $\Sigma = \{a, b, c, d\}$ and the model signature from (4.1) describing both the input and output structures. In order to define the input-output function, one must define all of the relations in the output signature in terms of input properties. In other words, the phonological rule from above is defined by describing what determines changes to the ordering relation ($\lhd$) and to the labeling relations ($\{R_a, R_b, R_c, R_d\}$).

Mathematically an interpretation of structure $A$ in terms of structure $B$ is a function denoted by a set of $n$ formulas $\{\phi_i, \ldots, \phi_n\}$ where $n$ is equal to the number of functions, relations, and constants in $A$'s model signature, plus a domain formula, copy set, and licensing formula. A formula $\phi_P(x) \overset{\text{def}}{=} Q(x)$ denotes that domain element $x$ has property $P$ in the output structure only if it has property $Q$ in the input structure. The domain formula indicates which domain elements of the input structure are in the domain of the function.[3] The copy set indicates how many copies of each domain element there are in the output structure. When the copy set is equal to 1, the size of the input and output structures are identical. Phonologically, the copy set can be used in instances such as epenthesis or when adding suprasegmental structure such as stress marking or syllabification to a segmental input structure. The licensing formula indicates which elements in the output structure should remain. Among other things, this can be used to ensure that the output structure is smaller in size than the input structure. One clear phonological use for the licensing function is deletion (Heinz, Forthcoming).

In the current example, the size of the structure remains the same in the output and the transformation applies to all domain elements, so the domain formula, copy set, and licensing formula are not discussed. Focusing on the functions $\{\phi_1, \ldots, \phi_n\}$, the following observations are made. First, the successor function in both the input and output is identical. That is, the phonological process being described does not change the ordering of elements. The following formula is therefore included:

$$\phi_{\lhd}(x, y) \overset{\text{def}}{=} x \lhd y \tag{4.6}$$

This simply states that the successor relation on the output is *interpreted* the same as the successor relation on the input.

---

[3] As you can see "domain" is overloaded here.

Before thinking about the remaining formulas it is worthwhile to just think about the generalization that the phonological process being described captures. Given the small set of symbols in the alphabet, it is easy to enumerate the possibilities:

1. An input string $a$ is rewritten as $b$ if it is preceded by $c$ or followed by $d$, otherwise it is (re)written as $a$.
2. An input string $b$ is always (re)written as $b$.
3. An input string $c$ is always (re)written as $c$.
4. An input string $d$ is always (re)written as $d$.

An alternative perspective, more aligned with the model-theoretic approach, is to flip the generalizations and describe what conditions lead to specific output strings.

1. An output string $a$ is the result of an input string $a$ which is not preceded by a $c$ or followed by a $d$.
2. An output string $b$ is the result of an input string $b$ or an input string $a$ which is preceded by a $c$ and followed by a $d$.
3. An output string $c$ is the result of an input string $c$.
4. An output string $d$ is the result of an an input string $d$.

These generalizations can be directly translated into the logical formulas needed to denote a model-theoretic interpretation.

$$\phi_a(x) \overset{\text{def}}{=} a(x) \wedge \neg\exists y, z[y \lhd x \lhd z \wedge c(y) \wedge d(z)] \tag{4.7}$$

$$\phi_b(x) \overset{\text{def}}{=} b(x) \vee (a(x) \wedge \exists y, z[y \lhd x \lhd z \wedge c(y) \wedge d(z)]) \tag{4.8}$$

$$\phi_c(x) \overset{\text{def}}{=} c(x) \tag{4.9}$$

$$\phi_d(x) \overset{\text{def}}{=} d(x) \tag{4.10}$$

Equation (4.7) only evaluates to `TRUE` when a domain element $x$ in the input structure has the property of being an $a$ and is not preceded by a domain element $y$ which has the property of being a $c$ and followed by a domain element $z$ which has the property of being a $d$. In other words, an output domain element only has the property of being an $a$ when it does not satisfy the structural description of the phonological process. If a domain element does satisfy the structural description, equation (4.8) ensures that it is interpreted as a $b$ in the output structure.

Suppose we wanted to interpret the string $cadab$ based on this mapping. The model-theoretic structure has 5 domain elements, each labeled with the given segmental property and linearly ordered. Each of the $\phi$ formulas above is then evaluated for each domain element and fully determines the output structure. Figure 4.5 shows the input structure and how it is related to the output structure while Table 4.1 shows the truth values for each of the $\phi$ formulas. Domain elements 0, 2, 3, and 4 remain faithful to their input labels in the output because this is what the formulas tell us. In the case of domain element 3, it remains $a$ in the output because the right half of the conjunct ultimately evaluates to true (corresponding to it being an $a$ not matching the structural description of the process). Likewise, domain element 4 remains a $b$ because it satisfies the left half of the disjunct. On the other hand, domain element 1 is unfaithful to the input structure and changes to $b$ in the output because it fails to satisfy the righthand side of the conjunct in equation (4.7), instead satisfy the righthand side of the disjunct in equation (4.8). Table 4.1 contains the evaluation table, showing which formulas are satisfied by which domain elements.

|  | $x$ |  |  |  |  |
|---|---|---|---|---|---|
| Formula | 0 | 1 | 2 | 3 | 4 |
| $\phi_a(x)$ | $\bot$ | $\bot$ | $\bot$ | $\top$ | $\bot$ |
| $\phi_b(x)$ | $\bot$ | $\top$ | $\bot$ | $\bot$ | $\top$ |
| $\phi_c(x)$ | $\top$ | $\bot$ | $\bot$ | $\bot$ | $\bot$ |
| $\phi_d(x)$ | $\bot$ | $\bot$ | $\top$ | $\bot$ | $\bot$ |

Table 4.1: Truth values for the transformation of $cadab$ to $cbdab$.

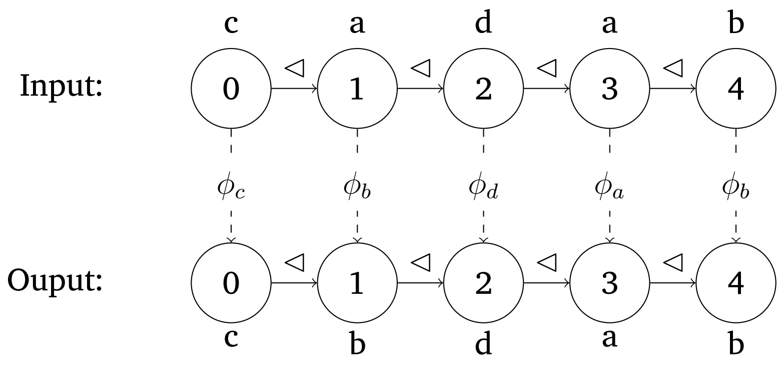


Figure 4.5: Mapping *cadab* to *cbdab*. Dashed arrows with overlayed formula indicate which formula describes the output property of a given domain element given its input properties.

### 4.2.4 Phonological Theory Comparison

The preceding subsections have shown that the model-theoretic approach is well-equipped to analyze and describe many different types of phonological knowledge: model signatures provide ways to build phonological structure (representations) and mathematical logic provides a way to analyze structure (constraints on representations) and transformations on structure (input-output mappings). This section ends with a discussion of how the model-theoretic approach can also be used for meta-theory comparison. While the model-theoretic approach is successful in formalizing phonological generalizations, its greatest strength lies in its ability to act as a shared meta-language between different theories of phonology. This further provides a way to pinpoint the exact way in which two theories vary, but more importantly, it provides a way to show similarities between two theories. Progress is best made when the wheel isn't constantly being reinvented. For this reason, formalization using a shared language allows for a better understanding of what various theories bring to the table. In turn, this allows for more pointed questions to be pursued.

Previous examples of this approach include Jardine et al. (2021) who compare standard autosegmental graphs to q-theory representations (Inkelas and Shih, 2016), and Oakden (2020) who compares two different tonal feature geometries. In both cases, it is found that the different representations being compared are best thought of as notational variants. Strother-Garcia (2019) provides multiple comparisons of different syllable structures using model theory and Nelson (2022) examines different phonological feature systems and the natural classes they produce from this model-theoretic perspective. The previous studies that have compared representations have done so using cases where there are at least some shared theoretical priors (generative phonology). It has yet to be fully explored what happens when the comparison is done between representations that have different theoretical priors. Later in this chapter, this is the exact approach that I take in comparing the sequential symbolic representations used in most generative theories of phonology with the dynamic coupling graphs used in Articulatory Phonology.

One reason that the model-theoretic approach works so well for phonological theory is due to mathematical logic acting as the bedrock for both describing the static constraints over structure as well as the interpretations of one structure in terms of another structure. As a working example, suppose we wanted to compare a theory of phonology that used only segments as the atomic elements and one that used only features as the atomic elements. Our model signature for the segmental theory will be the following.

$$\langle \lhd, R_t, R_d, R_n, R_a \rangle \tag{4.11}$$

This contains the successor ordering relation as well as labeling relations for $t$, $d$, $n$, $a$ which cover voiced and voiceless obstruents, nasals, and vowels. Our model signature for the featural theory will be the following.

$$\langle \lhd, R_{\texttt{voi}}, R_{\texttt{son}}, R_{\texttt{syl}} \rangle \tag{4.12}$$

This also contains the successor ordering relation and labeling relations corresponding to the features [`voice`], [`sonorant`], and [`syllabic`]. The relations can be viewed as encoding a positive valuation for these features when they evaluate to `TRUE` and a minus valuation when they evaluate to `FALSE`.

The process to interpret one structure in terms of another structure is the same as above, only now the model signature in the input and output structures vary. First, let us define the translation from a segmental structure to a feature structure. To do this, we need to define functions for each relation within the feature model signature using only relations from the segment model signature. These are shown below (successor is left out since it once again remains the same across the interpretation).

$$\phi_{\mathtt{voi}}(x) \overset{\text{def}}{=} d(x) \vee n(x) \vee a(x) \tag{4.13}$$

$$\phi_{\mathtt{son}}(x) \overset{\text{def}}{=} n(x) \vee a(x) \tag{4.14}$$

$$\phi_{\mathtt{syl}}(x) \overset{\text{def}}{=} a(x) \tag{4.15}$$

Feature properties are the disjunct of the symbols which carry the property. On the other hand, interpreting the segmental structure in terms of a feature structure shows that segmental properties are the conjunct (intersection) of the various feature properties.

$$\phi_t(x) \overset{\text{def}}{=} \neg\mathtt{voi}(x) \wedge \neg\mathtt{son}(x) \wedge \neg\mathtt{syl}(x) \tag{4.16}$$

$$\phi_d(x) \overset{\text{def}}{=} \mathtt{voi}(x) \wedge \neg\mathtt{son}(x) \wedge \neg\mathtt{syl}(x) \tag{4.17}$$

$$\phi_n(x) \overset{\text{def}}{=} \mathtt{voi}(x) \wedge \mathtt{son}(x) \wedge \neg\mathtt{syl}(x) \tag{4.18}$$

$$\phi_a(x) \overset{\text{def}}{=} \mathtt{voi}(x) \wedge \mathtt{son}(x) \wedge \mathtt{syl}(x) \tag{4.19}$$

While any finite structure can be interpreted in terms of another finite structure, there are certain aspects of the interpretation that may provide deeper insight into how similar the two structures are. For example, one dimension that provides for a level of equivalency is the power of the logic used to define the interpretation. There is precedence in relating the computational power of a function in terms of the type of logic required to do the computation (Büchi, 1960; Elgot, 1961; Trakhtenbrot, 1961; McNaughton and Papert, 1971; Thomas, 1982; Bhaskar et al., 2020, 2023). Since an interpretation is a function between structures, the power of the computation can be directly measured based on the minimal logic needed to describe the translation. For example, FO logic has been used to describe phonological constraints and transformations, but the full power of FO is not being used in our current example.

First of all, both the interpretation from features to segments and the interpretation from segments to features do not require quantification anywhere in their definitions. A logical formula containing quantification requires the computation to be sensitive to the global structure of the input. When interpreting segments as sets of features or sets of features as segments, the properties of surrounding elements have no bearing on the interpretation. This is why there are no quantifiers in the formula. Therefore, these interpretations are strictly less powerful than interpretations that require the full power of quantification provided by FO logic. FO logic without quantifiers is simply referred to as *quantifier-free* first-order logic (QF; Chandlee and Lindell, 2016).

QF logic was used by Strother-Garcia (2019, p. 39) in her study of different types of syllabic representation. In this work, she introduced the idea of two models being $L - interpretable$: a word model $\mathcal{M}^1$ is $L - interpretable$ in terms of another, $\mathcal{M}^2$, if one can write a graph transduction...from $\mathcal{M}^1$ to $\mathcal{M}^2$ using logic $L$...If $\mathcal{M}^1$ is $L - interpretable$ in terms of $\mathcal{M}^2$ and vice versa, then we say that the two are $L - bi - interpretable$. This expresses an idea of equivalence with respect to a given logic $L$ as the two models can be thought of as interchangeable. Given this characterization, being equivalent under a weaker logic suggests the differences between the two models are less significant. Strother-Garcia argues that the weak, local, QF provides a way to indicate notational equivalence.

Oakden (2020) expands on this idea of notational equivalence and bi-interpretability, but uses a slightly different definition provided by Friedman and Visser (2014): an interpretation $K : U \to V$ gives us a construction of an internal model $\widetilde{K}(\mathcal{M})$ of $U$ from a model $M$ of $V$. We find that $U$ and $V$ are bi-interpretable iff, there are interpretations $K : U \to V$ and $M : V \to U$ and formulas $F$ and $G$ such that, for all models $\mathcal{M}$ of $V$, the formula $F$ defines an isomorphism between $\mathcal{M}$ and $\widetilde{M}\widetilde{K}(\mathcal{M})$, and, for all models $\mathcal{N}$ of $U$, the formula $G$ defines an isomorphism between $\mathcal{N}$ and $\widetilde{K}\widetilde{M}(\mathcal{N})$. The notion of bi-interpretability will be used below when comparing feature and segmental models of phonological structure with gestural models

## 4.3 Translating Between Coupling Graphs and Segmental Strings

This section contains three parts. First, I show how to represent the coupling graphs used in Articulatory Phonology within the model-theoretic approach. Second, I show how to interpret a string as a coupling graph and then third I show how to interpret a coupling graph as a string. Both of these interpretations preserve the necessary structure between the two representations and therefore qualify as an isomorphism. Third, I discuss what the results mean for notational equivalence. The resulting interpretation is not quantifier-free, meaning it doesn't quite reach previous definitions of notational equivalence (Strother-Garcia, 2019; Oakden, 2020). Nonetheless, I suggest a possible phase function to replace the phase relations used in Section 4.3.1 that would result in a quantifier-free interpretation and therefore satisfy notational equivalence.

### 4.3.1 Representing Coupling Graphs

One problem that arises when representing coupling graphs results from the primitive ordering relations being defined in terms of phases. In theory, there are an infinite number of phase relationships that could hold between two elements. Under this view, it would be impossible to define a finite set of ordering relations. One possible solution would be to use an ordering function instead. $\Phi(x,y) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R} \in [0,360)$ would provide a real-numbered phase value between two domain elements. This move brings further problems as it would require expanding the logic to also include a $<$ function to distinguish between different types of phase relationships. Rather than assume all phase relationships are possible, I will instead use four binary relations that cover the main type of phase relationships used in Articulatory Phonology: in-phase (0 degrees), anti-phase (180 degrees), abutting (60 degrees), and eccentric (30 degrees). As discussed in section 4.1, the first two are the stable coordination patterns and the latter two are the learned coordination patterns for special situations such as the release of a consonant gesture or the coda-specific cluster pattern. Table 4.2 shows the symbols used to represent each of these relations.

| Relation | Label |
|---|---|
| $\Diamond(x,y)$ | In-phase |
| $\lhd_{180}(x,y)$ | Anti-phase |
| $\lhd_{60}(x,y)$ | Abutting |
| $\lhd_{30}(x,y)$ | Eccentric |

Table 4.2: Ordering relations for coupling graph model

Table 4.3 shows the labeling relations necessary to account for the gestures represented by each node of a coupling graph. These include the articulator, the constriction location and the constriction degree.

| Relation | Label | Relation | Label |
|---|---|---|---|
| `LIPS` | Labial Articulator | `rel` | Constriction Degree: release |
| `TT` | Tongue Tip Articulator | `pro` | Constriction Location: protruded |
| `TB` | Tongue Body Articulator | `dent` | Constriction Location: dental |
| `VEL` | Velum Articulator | `alv` | Constriction Location: alveolar |
| `GLO` | Glottis Articulator | `palv` | Constriction Location: postalveolar |
| `clo` | Constriction Degree: closed | `pal` | Constriction Location: palatal |
| `crit` | Constriction Degree: critical | `vel` | Constriction Location: velar |
| `nar` | Constriction Degree: narrow | `uvul` | Constriction Location: uvular |
| `V` | Constriction Degree: vowel | `phar` | Constriction Location: pharyngeal |
| `wide` | Constriction Degree: wide | | |

Table 4.3: Labeling relations for coupling graph model.

A coupling graph model for the English word *laughed* is shown in Figure 4.6. There are a few things to note here. First, onset /l/ has both a primary tongue tip gesture and a secondary tongue body gesture (Sproat and Fujimura, 1993). Second, all consonant gestures have both closure and release gestures for the primary articulation (Steriade, 1993; Browman, 1994; Nam, 2007b). Third, there is only a single devoicing gesture coupled in phase to the first consonant of the coda cluster (Goldstein, 2011). Each of these choices are made following relevant empirical data.

### 4.3.2 Translating from Strings to Coupling Graphs

The translation from strings to coupling graphs requires expanding the information that a given lexical representation directly encodes. In the TADA software package (Nam et al., 2022), there is a program called GEST that takes string representations of English lexical items (ARPABET) and directly translates them into coupling graphs. The other two main programs within TADA compute inter-gestural and inter-articulator coordination and are "meant to be part of a model of the human speech production process." On the other hand, GEST is claimed to be "a heuristic that is not meant to model how a speaker would go about construction (*sic*) a coupling graph for an arbitrary form." Therefore, the ability to compute a coupling graph from an arbitrary English phonetic transcription is not novel. However, GEST is written in Perl, an imperative programming language. Programs in imperative programming languages largely describe how a process takes place in contrast with programs in declarative programming languages which describe what the conditions are for a process to take place. The model-theoretic approach discussed above can be interpreted as defining declarative programs (cf. Declarative Phonology; Bird et al., 1992; Bird, 1995; Scobbie et al., 1996). Therefore, despite the existence of GEST, the translation defined below offers an alternative perspective on the computational relationship between strings and coupling graphs. The approach taken here can also provide an upper bound on the expressive power needed to produce a coupling graph from a string which

$$
\begin{aligned}
\mathcal{D} &:= \{1,2,3,4,5,6,7,8,9\} \\
\Diamond &:= \{(1,2),(2,4),(5,9)\} \\
\lhd_{180} &:= \{(4,5)\} \\
\lhd_{60} &:= \{(2,3),(5,6),(7,8)\} \\
\lhd_{30} &:= \{(5,7)\} \\
\texttt{LIPS} &:= \{5,6\} \\
\texttt{TT} &:= \{2,3,7,8\} \\
\texttt{TB} &:= \{1,4\} \\
\texttt{GLO} &:= \{9\}
\end{aligned}
$$

$$
\begin{aligned}
\texttt{dent} &:= \{5\} \\
\texttt{alv} &:= \{2,7\} \\
\texttt{uvul} &:= \{1\} \\
\texttt{phar} &:= \{4\} \\
\texttt{clo} &:= \{7\} \\
\texttt{crit} &:= \{5\} \\
\texttt{nar} &:= \{1,2\} \\
\texttt{wide} &:= \{9\} \\
\texttt{rel} &:= \{3,6,8\} \\
\texttt{V} &:= \{4\}
\end{aligned}
$$

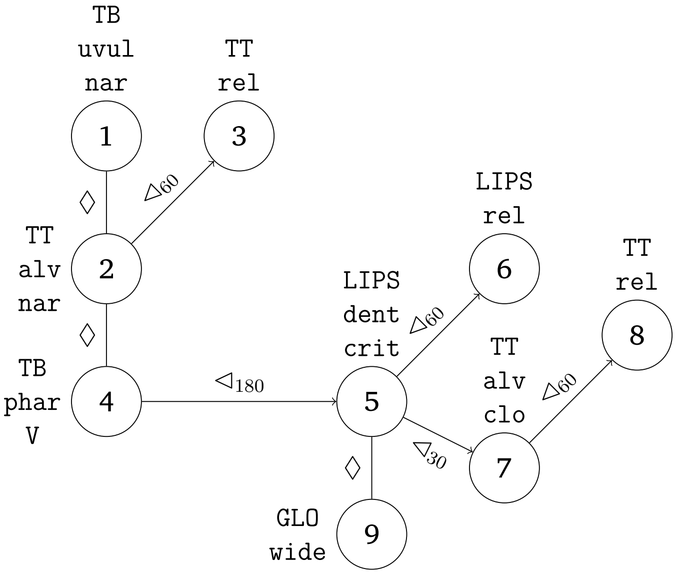


Figure 4.6: Coupling graph model for English word *laughed*.

$$\begin{aligned}
\langle \mathcal{D} &:= \{1,2,3,4\}\\
\lhd &:= \{(1,2),(2,3),(3,4)\}\\
\text{æ} &:= \{2\}\\
\text{f} &:= \{3\}\\
\text{l} &:= \{1\}\\
\text{t} &:= \{4\}\\
\sigma &:= \{\}; \sigma \in \Sigma \setminus \{\text{æ,f,l,t}\}\rangle
\end{aligned}$$

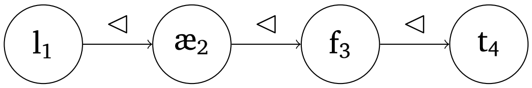


Figure 4.7: String model for English word *laughed* ([læft])

is less transparent when using the declarative approach.

The specific transduction from strings to coupling graphs given here will focus on English, but the general approach can be adjusted on a language-specific (or even dialect-specific) basis. We will start with the "beads on a string" model signature given in (4.1) above and repeated in (4.20).

$$\langle \lhd, \{R_\sigma \mid \sigma \in \Sigma\}\rangle \tag{4.20}$$

The alphabet, $\Sigma$, will just be the IPA symbols for each English phone. The output model signature will be the coupling graph model described in section 4.3.1.

Translation from a string to a coupling graph will be explained in parts, though the declarative aspect of the formalism means there is no proper order for how the transformation takes place.[4] The procedural explanation given simply lists the various properties that determine the structure of the output form. The string [læft] will be used as a guiding example with only the necessary equations shown. The Appendix contains additional equations for the English string-to-coupling graph translation. An input model is shown in Figure 4.7.

The first step will be explaining the copy set. As described previously, copy sets are the technical machinery required to have an output structure bigger than the input structure. Copies in this sense should not be thought of as *literal* copies, but rather a useful way to

---

[4] With that in mind, logical programming can be given a procedural interpretation (Kowalski, 1974; Lloyd, 2012).

provide a bound on the number of output elements. For a structure of size $n$, the number of elements in the output structure can be no greater than $n \times m$ where $m$ is the number of copies. Furthermore, copy sets can be used advantageously in the current instance. There are essentially four aspects of a coupling graph that are required: a primary gesture (all sounds), a release gesture (for consonants only), a secondary articulation (for a subset of sounds), and a nasal or (de)voicing gesture. Since voicing is not contrastive on nasals in English, the same copy set can be used to add a nasal or devoicing gesture. In total, four types of information potentially can be inferred from a single segment. Therefore, the copy set is of size four.

$$C \overset{\text{def}}{=} \{1,2,3,4\} \tag{4.21}$$

For expository clarity, I will include what each formula does to a “workspace”. The idea behind the workspace is that each formula in the transduction puts a constraint on the workspace. The final output is therefore the intersection of the sub-workspaces defined by each formula. The copy set is given first in Figure 4.8, as it restricts the size of the workspace. Each node is represented as a sequence $(x, y)$ where $x$ corresponds with the domain label of the input structure and $y$ corresponds to the copy. $(1, 3)$ therefore represents the third copy of the first domain element.

Each copy can be thought of as a different possible gesture that needs to be labeled in a specific way. These labeling relations from Table 4.3 need to therefore be defined in terms of the string model. Each relation must be defined for each copy set. The necessary formulas to define the transduction for [læft] are shown below. An output formula $\varphi^i_p(x)$ determines whether or not domain element $x$ has property $p$ on its $i$th copy. As will

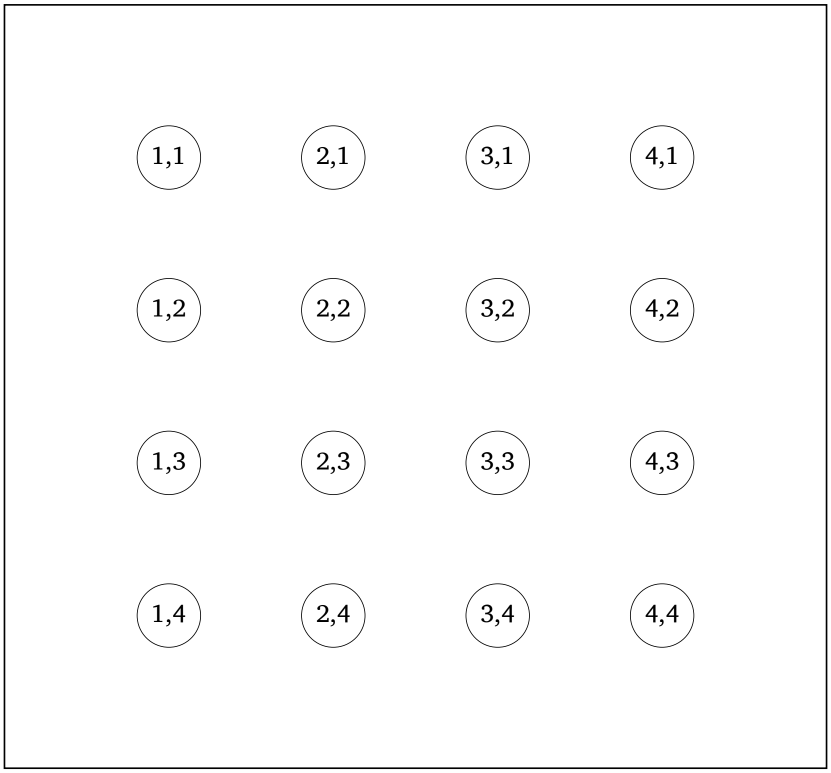


Figure 4.8: Workspace generated by the copy set formula (4.21) for the string to coupling graph transduction.

continue to be the case, the additional formula can be found in the Appendix.

$$\varphi^{1}_{\texttt{LIPS}}(x) \overset{\text{def}}{=} \text{b}(x) \lor \text{p}(x) \lor \text{m}(x) \lor \text{v}(x) \lor \text{f}(x) \tag{4.22}$$

$$\varphi^{2}_{\texttt{LIPS}}(x) \overset{\text{def}}{=} \varphi^{1}_{\texttt{LIPS}}(x) \tag{4.23}$$

$$\varphi^{1}_{\texttt{TT}}(x) \overset{\text{def}}{=} \text{d}(x) \lor \text{t}(x) \lor \text{n}(x) \lor \text{ð}(x) \lor \text{θ}(x) \lor \text{l}(x) \lor \ldots \tag{4.24}$$

$$\varphi^{2}_{\texttt{TT}}(x) \overset{\text{def}}{=} \varphi^{1}_{\texttt{TT}}(x) \tag{4.25}$$

$$\varphi^{1}_{\texttt{TB}}(x) \overset{\text{def}}{=} \text{g}(x) \lor \text{k}(x) \lor \text{ŋ}(x) \lor \text{j}(x) \lor \text{w}(x) \lor \text{æ}(x) \lor \ldots \tag{4.26}$$

$$\varphi^{3}_{\texttt{TB}}(x) \overset{\text{def}}{=} \text{z}(x) \lor \text{s}(x) \lor \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x) \lor \text{l}(x) \tag{4.27}$$

$$\varphi^{4}_{\texttt{GLO}}(x) \overset{\text{def}}{=} \text{p}(x) \lor \text{t}(x) \lor \text{k}(x) \lor \text{f}(x) \lor \text{θ}(x) \lor \text{s}(x) \lor \ldots \tag{4.28}$$

$$\varphi^{1}_{\texttt{clo}}(x) \overset{\text{def}}{=} \text{b}(x) \lor \text{p}(x) \lor \text{m}(x) \lor \text{d}(x) \lor \text{t}(x) \lor \ldots \tag{4.29}$$

$$\varphi^{1}_{\texttt{crit}}(x) \overset{\text{def}}{=} \text{v}(x) \lor \text{f}(x) \lor \text{ð}(x) \lor \text{t}(x) \lor \text{z}(x) \lor \ldots \tag{4.30}$$

$$\varphi^{1}_{\texttt{nar}}(x) \overset{\text{def}}{=} \text{j}(x) \lor \text{w}(x) \lor \text{ɹ}(x) \lor \text{l}(x) \tag{4.31}$$

$$\varphi^{3}_{\texttt{nar}}(x) \overset{\text{def}}{=} \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x) \lor \text{l}(x) \lor \text{w}(x) \lor \ldots \tag{4.32}$$

$$\varphi^{4}_{\texttt{wide}}(x) \overset{\text{def}}{=} \text{p}(x) \lor \text{t}(x) \lor \text{k}(x) \lor \text{f}(x) \lor \text{θ}(x) \lor \text{s}(x) \lor \ldots. \tag{4.33}$$

$$\varphi^{1}_{\texttt{V}}(x) \overset{\text{def}}{=} \text{i}(x) \lor \text{ɪ}(x) \lor \text{eɪ}(x) \lor \text{ɛ}(x) \lor \text{æ}(x) \lor \ldots \tag{4.34}$$

$$\varphi^{2}_{\texttt{rel}}(x) \overset{\text{def}}{=} \texttt{True} \tag{4.35}$$

$$\varphi^{1}_{\texttt{dent}}(x) \overset{\text{def}}{=} \text{v}(x) \lor \text{f}(x) \lor \text{ð}(x) \lor \text{θ}(x) \tag{4.36}$$

$$\varphi^{1}_{\texttt{alv}}(x) \overset{\text{def}}{=} \text{d}(x) \lor \text{t}(x) \lor \text{n}(x) \lor \text{z}(x) \lor \text{s}(x) \lor \text{l}(x) \tag{4.37}$$

$$\varphi^{1}_{\texttt{phar}}(x) \overset{\text{def}}{=} \text{æ}(x) \lor \text{ɑ}(x) \lor \text{ɔ}(x) \tag{4.38}$$

$$\varphi^{3}_{\texttt{uval}}(x) \overset{\text{def}}{=} \text{l}(x) \tag{4.39}$$

Given the input string model for [læft] defined above, each of these formulas can be evaluated for each domain element. This is shown in Table 4.4. These truth values can generate a workspace as well. One thing to note is that the formulas as they are currently defined overgenerate. For example, copy set 2 automatically gets labeled as a

|  | $x$ |  |  |  |
|---|---|---|---|---|
| Formula | 1 | 2 | 3 | 4 |
| $\varphi^{1}_{\texttt{LIPS}}(x)$ | $\bot$ | $\bot$ | $\top$ | $\bot$ |
| $\varphi^{2}_{\texttt{LIPS}}(x)$ | $\bot$ | $\bot$ | $\top$ | $\bot$ |
| $\varphi^{1}_{\texttt{TT}}(x)$ | $\top$ | $\bot$ | $\bot$ | $\top$ |
| $\varphi^{2}_{\texttt{TT}}(x)$ | $\top$ | $\bot$ | $\bot$ | $\top$ |
| $\varphi^{1}_{\texttt{TB}}(x)$ | $\bot$ | $\top$ | $\bot$ | $\bot$ |
| $\varphi^{3}_{\texttt{TB}}(x)$ | $\top$ | $\bot$ | $\bot$ | $\bot$ |
| $\varphi^{4}_{\texttt{GLO}}(x)$ | $\bot$ | $\bot$ | $\top$ | $\top$ |
| $\varphi^{1}_{\texttt{clo}}(x)$ | $\bot$ | $\bot$ | $\bot$ | $\top$ |
| $\varphi^{1}_{\texttt{crit}}(x)$ | $\bot$ | $\bot$ | $\top$ | $\bot$ |
| $\varphi^{1}_{\texttt{nar}}(x)$ | $\top$ | $\bot$ | $\bot$ | $\bot$ |
| $\varphi^{3}_{\texttt{nar}}(x)$ | $\top$ | $\bot$ | $\bot$ | $\bot$ |
| $\varphi^{4}_{\texttt{wide}}(x)$ | $\bot$ | $\bot$ | $\top$ | $\top$ |
| $\varphi^{1}_{\texttt{V}}(x)$ | $\bot$ | $\top$ | $\bot$ | $\bot$ |
| $\varphi^{2}_{\texttt{rel}}(x)$ | $\top$ | $\top$ | $\top$ | $\top$ |
| $\varphi^{1}_{\texttt{dent}}(x)$ | $\bot$ | $\bot$ | $\top$ | $\bot$ |
| $\varphi^{1}_{\texttt{alv}}(x)$ | $\top$ | $\bot$ | $\bot$ | $\top$ |
| $\varphi^{1}_{\texttt{phar}}(x)$ | $\bot$ | $\top$ | $\bot$ | $\bot$ |
| $\varphi^{3}_{\texttt{uval}}(x)$ | $\top$ | $\bot$ | $\bot$ | $\bot$ |

Table 4.4: Truth values for the gesture labels in the coupling graph interpretation of the string model for [læft].

release gesture, even though only consonants are ultimately given release gestures. This highlights the parallel process and how the final structure requires satisfying all formulas at the same time. Later, when the licensing function gets defined, only copies that were underlying consonants will remain in the output. The workspace generated from the labeling formulas can be seen in Figure 4.9.

Now that the labeling relations are taken care of, the ordering must be determined. The "heuristics" mentioned above almost entirely have to do with ordering and not the primary and secondary gestures themselves. In the following section, the focus will be on single syllable words, though this method can be expanded to multi-syllabic words without needing to expand the logical power (Strother-Garcia, 2019). The first heuristic is that a release gesture has an abutting phase relationship with its corresponding closure gesture. Therefore, it is possible to continue with the tactic of overgeneration and say

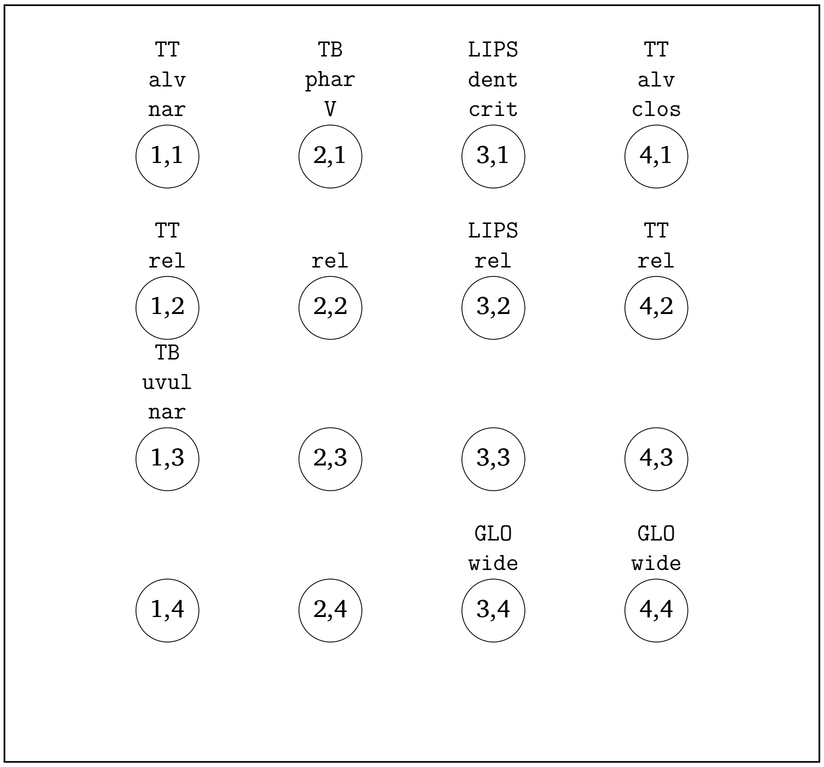


Figure 4.9: Workspace properties generated by the labeling relations for the string to coupling graph transduction of [læft].

that the abutting relation always occurs between copy sets one and two when the index is identical.

$$\varphi^{1,2}_{\lhd 60}(x,y) \overset{\text{def}}{=} (x = y) \tag{4.40}$$

Many of the remaining heuristics are defined in terms of whether or not something is a consonant or vowel gesture. To make the formula more readable, the helper predicates `C` and `V` are defined as follows.

$$\texttt{C}\ (x) \overset{\text{def}}{=} \mathbf{b}(x) \vee \mathbf{p}(x) \vee \mathbf{m}(x) \vee \mathbf{v}(x) \vee \mathbf{f}(x) \vee \mathbf{l}(x) \vee \mathbf{t}(x) \vee \ldots \tag{4.41}$$

$$\texttt{V}\ (x) \overset{\text{def}}{=} \text{i}(x) \vee \text{ɪ}(x) \vee \text{eɪ}(x) \vee \text{ɛ}(x) \vee \text{æ}(x) \vee \ldots \tag{4.42}$$

Helper predicates are strictly not part of the defined syntax of the logical languages being used, but provide a way to aggregate over many elements which share similar properties. Consequently, the formulas of interest become more readable. The next heuristic that can be defined for determining the phase relationships of the coupling graph has to do with the primary gestures. Any consonant is in-phase with a vowel that directly follows it. In fact, in a single syllable word, any consonant that precedes the vowel will be in the onset and therefore in-phase with the vowel. This ordering relation helps explain the widely documented "C-center" effect (Browman and Goldstein, 1988).[5] For now, I will assume that onsets and codas in English are limited to containing three consonants (*e.g. sprints*).

$$\varphi^{1,1}_{\Diamond}(x,y) \overset{\text{def}}{=} \texttt{C}\ (x) \wedge \texttt{V}\ (y) \wedge [x \lhd y \vee \exists z[\texttt{C}\ (z) \wedge x \lhd z \lhd y] \vee \exists w, z[\texttt{C}\ (w) \wedge \texttt{C}\ (z) \wedge x \lhd w \lhd z \lhd y]] \tag{4.43}$$

Equation (4.43) defines an in-phase relationship between any pre-vocalic consonant and vowel by checking whether or not the immediate, secondary, or tertiary successor of any consonant in the string is a vowel.

---

[5] Using general precedence rather than successor as the ordering relation in the string model would allow for an easier definition, but the definition itself would not hold up when discussing multi-syllabic words.

Determining coda consonants is slightly different. Consonants that immediately follow vowels are in an anti-phase relationship with the vowel.

$$\varphi^{1,1}_{\lhd_{180}}(x,y) \overset{\text{def}}{=} \texttt{V}\,(x) \wedge \texttt{C}\,(y) \wedge x \lhd y \tag{4.44}$$

The remaining coda consonants are eccentrically phased with the preceding consonant. Once again, the size of the coda will be limited to three consonants, therefore making the final ordering relation on the primary gesture copy set as follows.

$$\varphi^{1,1}_{\lhd_{30}}(x,y) \overset{\text{def}}{=} \texttt{C}\,(x) \wedge \texttt{C}\,(y) \wedge x \lhd y \wedge [\exists z[\texttt{V}\,(z) \wedge z \lhd x \lhd y] \vee \exists w,z[\texttt{V}\,(w) \wedge \texttt{C}\,(z) \wedge w \lhd z \lhd x \lhd y]] \tag{4.45}$$

Despite the heavy use of quantification, these formulas are looking for local substructures to determine the phase relations: If the string contains substructure $V_xC_y$ then $x \lhd_{180} y$; If the string contains substructure $V_zC_xC_y$ then $x \lhd_{30} y$; If the string contains substructure $V_wC_zC_xC_y$ then $x \lhd_{30} y$.

Two ordering relations remain to be defined. First, primary gestures and secondary gestures have an in-phase relationship. Second, devoicing and nasal gestures also have an in-phase relationship with the primary gesture.[6]

$$\varphi^{1,3}_{\Diamond}(x,y) \overset{\text{def}}{=} (x = y) \wedge [\mathbf{l}(x) \vee \mathbf{ɹ}(x) \vee \mathbf{j}(x) \vee \mathbf{w}(x) \ldots] \tag{4.46}$$

$$\varphi^{1,4}_{\Diamond}(x,y) \overset{\text{def}}{=} (x = y) \wedge [\mathbf{m}(x) \vee \mathbf{n}(x) \vee \mathbf{ŋ}(x) \vee \mathbf{p}(x) \vee \mathbf{t}(x) \vee \mathbf{k}(x) \vee \mathbf{f}(x) \ldots] \tag{4.47}$$

Figer 4.10 shows the workspace generated by these ordering formulas.

The final aspect needed to fully define the transduction is the licensing function. Not every copy will be retained in the final output structure. Each copy set has a specific

---

[6] In practice, the phasing of devoicing gestures for stops and fricatives is slightly different due to aspiration. In the GEST program, the devoicing gesture for stops is phased 20 degrees after the closure gesture. Accounting for this would involve adding a $\lhd_{20}$ relation to the model and an output predicate for this relation between copy set 1 and copy set 4. I leave that aside for now as it does not affect the larger point being made.

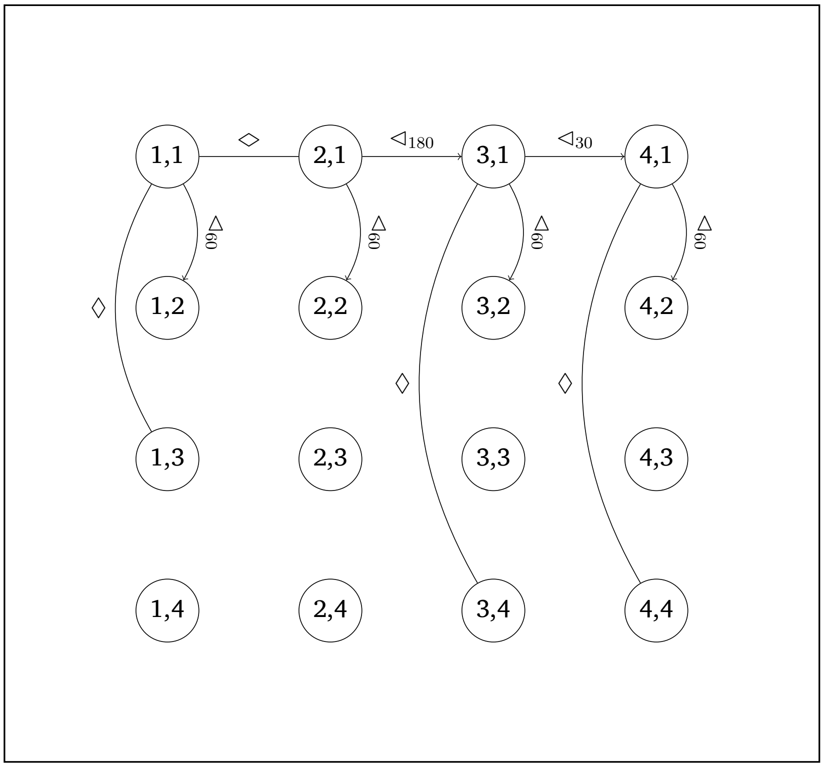


Figure 4.10: Workspace properties generated by the ordering relations for the string to coupling graph transduction of [læft].

licensing function. Based on the way the rest of the transduction was defined, every element in copy set 1 is licensed as this set contains the primary gestures. Only domain elements that correspond to consonants are licensed in copy set 2 while only domain elements which have secondary gestures are licensed in copy set 3. Elements in copy set 4 are only licensed if they correspond to a nasal segment or a voiceless segment that is the initial segment in a voiceless cluster. This is due to there only being a single devoicing gesture in these types of scenarios.

$$\varphi^{1}_{license} \overset{\text{def}}{=} \texttt{True} \tag{4.48}$$

$$\varphi^{2}_{license} \overset{\text{def}}{=} \texttt{C}\,(x) \tag{4.49}$$

$$\varphi^{3}_{license} \overset{\text{def}}{=} \text{ʒ}(x) \vee \text{ʃ}(x) \vee \text{dʒ}(x) \vee \text{tʃ}(x) \vee \text{z}(x) \vee \text{s}(x) \vee \text{l}(x) \vee \ldots \tag{4.50}$$

$$\begin{aligned}\varphi^{4}_{license} \overset{\text{def}}{=} & \ (\text{m}(x) \vee \text{n}(x) \vee \text{ŋ}(x)) \vee (\text{p}(x) \vee \text{t}(x) \vee \text{k}(x) \vee \ldots \\ & \wedge \neg \exists y [y \lhd x \wedge \text{p}(x) \vee \text{t}(x) \vee \text{k}(x) \vee \ldots)]\end{aligned} \tag{4.51}$$

Figure 4.11 shows which domain elements are licensed in the workspace. Filled-in nodes are not licensed and therefore any relations that contain those nodes are removed from the output structure. This is what allows for overgeneration in the previous formulas. For example, the ordering relation for release gestures was defined for all domain elements between copy set one and copy set two. Instead, it would have been possible to define it for only consonant gestures, but then it would still be necessary to license only consonant gestures in copy set two. Overgenerating in the main formula allows for parsimony in the sense that we don't have to twice define that it is only consonant gestures that have a corresponding release gesture.

Given all of the formulas defined here, the input string structure for [læft] is turned exactly into (or *interpreted as...*) the coupling graph structure shown in Figure 4.6. Despite gestures technically being continuous parameters in a dynamic equation, by having fixed values they offer a discrete interpretation. This discreteness is what allows for symbolic

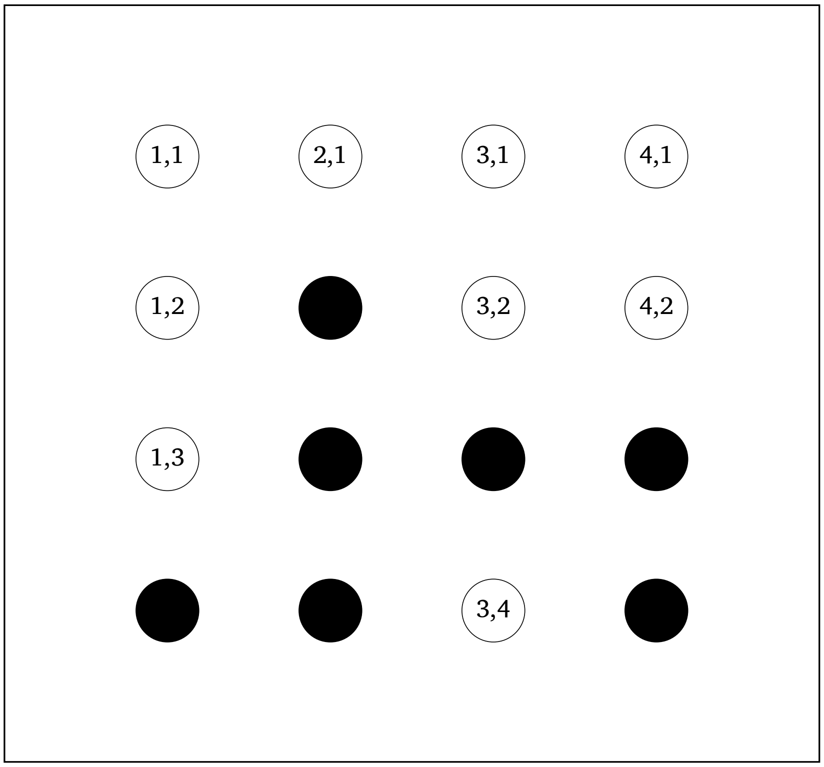


Figure 4.11: Workspace properties generated by the licensing for the string to coupling graph transduction of [læft].

representations to be given a gestural interpretation. The translation given above would ultimately be based on specific parameters for an English speaker, but there is room for variation in these parameters based on inter- and intra-language variation. In this sense, the formula provided here may define translations from strings to sets of coupling graphs. For example, the specific pronunciation of vowels and fricatives is known to vary cross-linguistically, even within a class that would be represented by the same IPA symbol. The symbolic representation can therefore be thought of as an equivalence class of certain gestural configurations. For now, I turn to the translation from coupling graph to string.

### 4.3.3 Translating from Coupling Graphs to Strings

The translation from strings to coupling graphs being first-order definable is itself a novel result, but not a totally unsurprising one. The GEST program discussed above translates strings into coupling graphs using a Perl program. Though the exact level of computational resources needed for the program to run is beyond the scope of this dissertation, it does act as a construction proof that all the information contained in a coupling graph can be inferred directly from the string representation. This aspect is somewhat surprising. Coupling graph representations in general are more expressive than string representations, but it seems that their full power is not needed to encode phonological structure. Projecting a lower-dimensional object into higher-dimensional space is always possible, but it brings with it questions about why the full higher-dimensional space is not being used.

Transduction in the opposite direction, coupling graph to string, seems like it would involve the loss of information. The richer coupling-graph representation is reduced to a symbolic string. The previous section showed that this is entirely not the case: all of the information encoded in a coupling graph could be directly inferred from the string representation. In this section, I show that identifying a string representation from a coupling graph is straightforward. This is possible because of definitions from Gafos

(2002) and Smith (2018) on *head* gestures. For Gafos, a head gesture is the oral gesture in the sequence. Certain segments, such as /l/, have multiple oral gestures. As Smith (2018) points out, defining what is the primary gesture in some cases is not straightforward. I determine the primary gesture in the following way: glottal and velum gestures are never primary, vowel gestures are always primary, and consonant gestures with a release are primary. In the case of /l/, this makes the tongue tip gesture the primary gesture. Coupling relations could also be used as secondary gestures are typically coupled only to the primary gesture and not neighboring gestures.

With this in mind, the strategy for inferring linear ordering from dynamic coupling involves what I call *spine identification*. This is done by looking at the head or primary gestures of the coupling graph. The technical aspect simply involves reducing the different phase relations to a single linear ordering successor relation. Figure 4.12 shows the spine for the coupling graph for [læft] highlighted in red. It is simply the subgraph that does not contain any of the release gestures, the glottal gesture, and the secondary articulation for /l/. I now discuss the translation in full.

Since the output structure in this instance needs to be smaller than the input structure, only one copy set is needed.

$$C \overset{\text{def}}{=} \{1\} \tag{4.52}$$

The labeling relations are essentially just the inverse of the labeling relations when going from string to coupling graph. Only now, we look for the intersection of the necessary properties. This echoes the discussion earlier about translating between segments and features. While gestural properties are not *features* in the way the term is standardly used in linguistics, they are still sub-segmental properties that combine to be interpreted as a segment or segment-like thing. The equations below show how to determine the segments in [læft] while the remaining equations for English segments are in the Appendix.

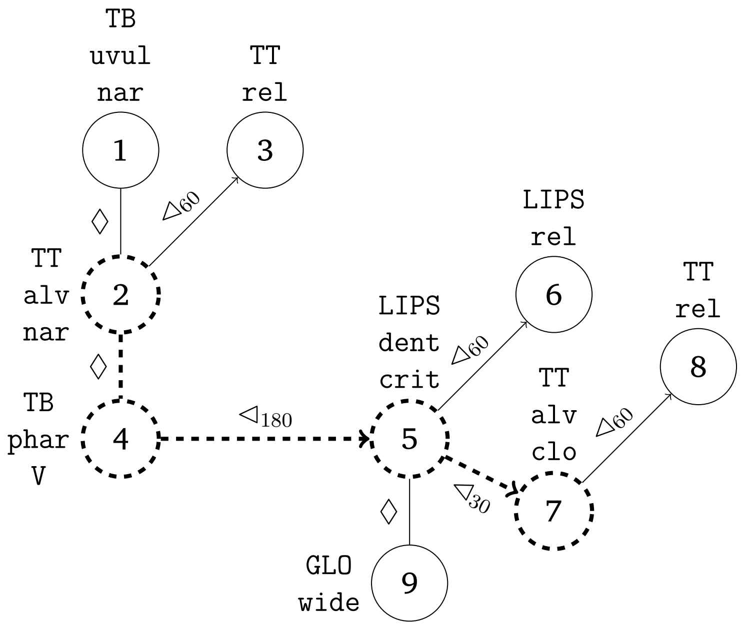


Figure 4.12: Subgraph (dashed) corresponding to the “spine” of the [læft] coupling graph model.

$$\varphi_{\mathrm{l}}(x) \overset{\text{def}}{=} \mathtt{TT}(x) \wedge \mathtt{alv}(x) \wedge \mathtt{nar}(x) \wedge \exists y[x \Diamond y \wedge \mathtt{TB}(x) \wedge \mathtt{uvul}(x) \wedge \mathtt{nar}(x)] \tag{4.53}$$

$$\varphi_{\text{æ}}(x) \overset{\text{def}}{=} \mathtt{TB}(x) \wedge \mathtt{phar}(x) \wedge \mathtt{V}(x) \tag{4.54}$$

$$\varphi_{\mathrm{f}}(x) \overset{\text{def}}{=} \mathtt{LIPS}(x) \wedge \mathtt{dent}(x) \wedge \exists y[x \Diamond y \wedge \mathtt{GLO}(y)] \tag{4.55}$$

$$\varphi_{\mathrm{t}}(x) \overset{\text{def}}{=} \mathtt{TT}(x) \wedge \mathtt{alv}(x) \wedge \mathtt{clo}(x) \wedge \exists y[x \Diamond y \wedge \mathtt{GLO}(y)] \tag{4.56}$$

The next step involves determining the output successor relation. This can be done by considering the ordering heuristics from above. First, anything coupled anti-phase with another gesture precedes that gesture. This accounts for post-vocalic consonants as well as onset consonant clusters. Second, anything eccentrically phased follows the thing it is eccentrically phased with (coda clusters). Finally, all onset consonants are in phase with the following vowel, but we only want the final onset consonant to directly precede the vowel. All of these can be accomplished with a single sentence of first-order logic.

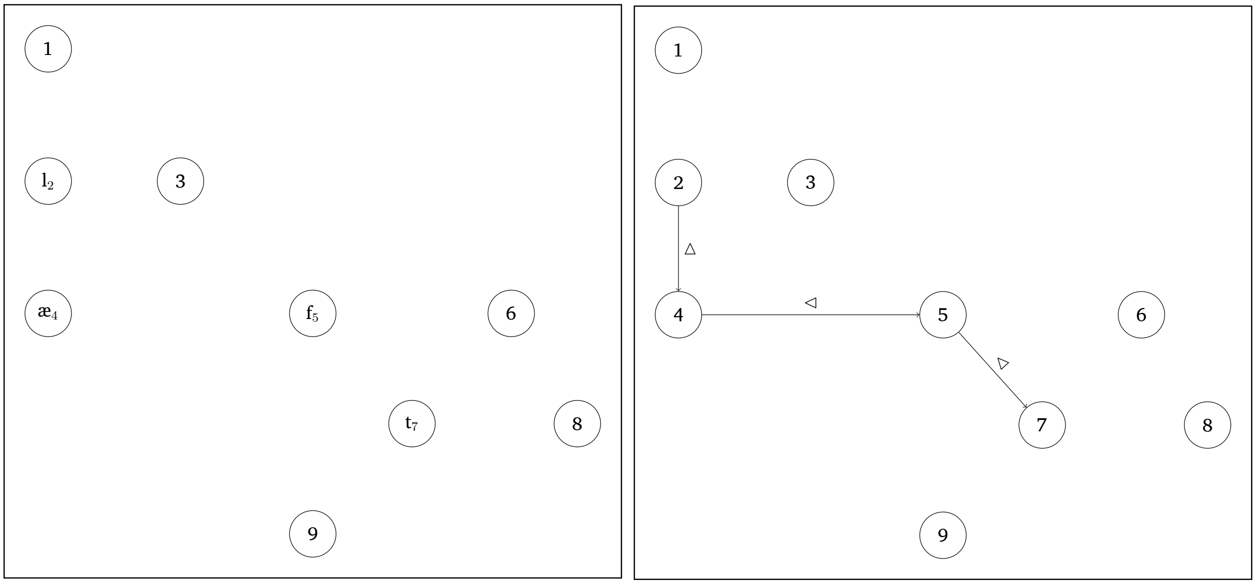
$$\varphi_{\lhd}(x,y) \overset{\text{def}}{=} (x \lhd_{180} y) \vee (x \lhd_{30} y) \vee (x \diamond y \wedge \mathtt{V}(y) \wedge \neg\exists z[x \lhd_{180} z]) \tag{4.57}$$

The final step involves only licensing the spine. This is done by explicitly licensing only things that do not have undesirable properties (glottal gesture, velum gesture, release gesture, secondary articulation). Since the actual types of secondary articulations vary, I use a predicate `SecArc` that I leave undefined for now.

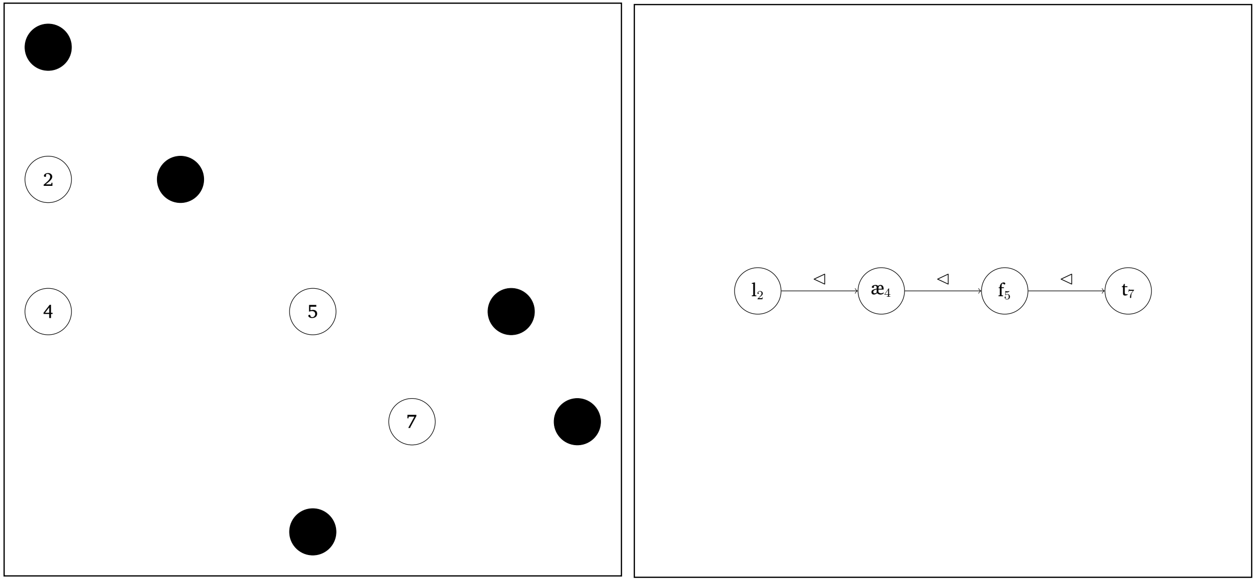
$$\varphi_{license}(x) \overset{\text{def}}{=} \neg\mathtt{rel}(x) \wedge \neg\mathtt{GLO}(x) \wedge \neg\mathtt{VEL}(x) \wedge \neg\,\mathtt{SecArc}\,(x) \tag{4.58}$$

This definition has an additional interesting property because it contains the conjunction of negated properties, making it reminiscent of the logical characterization of strictly local stringsets (Rogers et al., 2013). Figure 4.13 summarizes all parts of the transduction: labeling, ordering, licensing, and combining all of these properties to a final output structure. In many ways, inferring linear order from coupling graphs is straightforward. A segmental label is taken based on the intersection of gestural properties, ordering can be obtained based on the general heuristics for how gestures are aligned in Articulatory Phonology, and licensing eliminates unwanted secondary gestures.

Given that I have now shown how to translate from string to coupling graphs and from coupling graphs to string, and that when doing so the output structures are isomorphic to the originally defined string or coupling graph, it appears that this translation satisfies the definition for bi-interpretability. In other words, given a string model of *laughed* ($\mathcal{M}^s$), a coupling graph model of *laughed* ($\mathcal{M}^g$), a string to coupling graph transduction ($\Gamma^{sg}$), and a coupling graph to string transduction ($\Gamma^{gs}$), we have $\mathcal{M}^s \equiv \Gamma^{gs}(\Gamma^{sg}(\mathcal{M}^s))$ and $\mathcal{M}^g \equiv \Gamma^{sg}(\Gamma^{gs}(\mathcal{M}^g))$. Bi-interpretability therefore holds. What does not hold is the stronger definition of notation variance put forth by Strother-Garcia (2019) since the translations require quantification. This aspect of the analysis will be discussed in the summary of this chapter. As noted at multiple points throughout this section, there is a



(a) Labeling relations added (b) Successor relation added



(c) Unlicensed elements ignored (d) Combined final structure

Figure 4.13: Visualization of transduction of word “laughed” from the symbolic generative phonology model to the coupling graph articulatory phonology model

local flavor to the translation, but some aspects of how the coupling graph models are defined prevent a complete QF approach.

### 4.3.4 Towards a Quantifier-Free Translation

Much work in the model-theoretic approach to phonological theory comparison has focused on the "quantifier-free" nature of the transductions. If a first-order transduction contains no quantifiers, then all of the information needed to perform the computation is guaranteed to be local. Chandlee and Lindell (2021) show that first-order logic without quantifiers that includes successor and predecessor functions rather than ordering relations was equivalent to the input strictly local functions, a weak computational class of functions. What gives it this power is the ability to offset queries about neighboring properties to functions. When relations are used, a quantifier must be introduced to determine any properties beyond a specific domain element. Even when determining the properties of a locally adjacent domain element, quantification is necessary. With string models, the successor and predecessor functions resolve this issue.

The problem becomes slightly more difficult when considering non-string structures. While the insight that a quantifier-free transduction constrains the transduction to be local rather than global, it's less clear what type of function could be used for the coupling graph models discussed in this chapter. It's not quite as simple as moving from ordering relations to an ordering function. One can quite easily imagine a function $f : (\mathcal{D} \times \pi) \rightarrow \mathcal{D}$ that provides the domain element that has a certain phase relationship with another domain element. However, the issue in this case does not stem from using relations rather than function but instead from the ability of the same node in the coupling graph to be connected with multiple nodes by the same phase relationship. For example, in the graph for [læft], the /l/ gesture contains a primary tongue tip gesture which is in phase with the vocalic gesture, but the primary tongue tip gesture is also in phase with the secondary tongue body gesture. The defining feature of a mathematical function is the fact that for any given input there is at most one output. In the coupling graph for [læft], $f(2, \diamond) = 4$ and $f(2, \diamond) = 1$. One option would be to say that the output is always the same, but the output is the set $\{1, 4\}$. This would make the functions be of type $f : \mathcal{D} \rightarrow \mathcal{P}(\mathcal{D})$.

Set membership is used in monadic second-order logic which suggests a quantifier-free subset of MSO rather than FO may be of interest. In general, it seems that the quantifier-free aspect is more important than variables ranging solely over individual elements of the domain rather than sets of individual domain elements. Under this type of approach, the transduction between strings and coupling graphs would be quantifier-free while also satisfying the conditions on isomorphism laid out by Friedman and Visser (2014). This would support strings and coupling graphs not only being inter-translatable but also being *notational variants*. Exploring this option is left for future work.

## 4.4 Case Study: English past tense

I move now to look at a specific case study: English past tense alternations. This is a standard pattern found in many introductory textbooks (with its twin flame the English plural alternation). The basic descriptive facts are that the past tense morpheme surfaces as [ɨd] when following /t/ or /d/, [d] when following and voiced sound other than /d/, and [t] when following any voiceless sound other than /t/. Examples are shown in (18).

(18) a. *voted* [voʊɾɨd]
b. *loaded* [loʊɾɨd]
c. *tagged* [tægd]
d. *sighed* [saɪd]
e. *greased* [gɹist]
f. *backed* [bækt]

In the following subsections, I describe accounts based on rules, Optimality Theory, and Articulatory Phonology. The first two accounts will assume symbolic representations, but the Articulatory Phonology account uses coupling graphs and dynamics. Because of the shared interpretations between symbolic representations and coupling graphs described above, logic can be used as a shared meta-language to compare each account.

Rule and constraint grammars have different ontological commitments for how the UR to SR mapping occurs, but in this example share a representation. The Articulatory Phonology account uses a dynamical systems implementation of an OT grammar and therefore shares with the OT account ontological commitments for how an output form is generated (constraint optimization). The shared language of formal logic can act as a way to computationally evaluate what each theory has to offer. In the realm of computation, seemingly disparate theoretical proposals can be effectively neutralized.

### 4.4.1 Rule-based account

In rule-based phonology, the following two rules explain the alternations in the English past tense morpheme:

(19) a. $\emptyset \rightarrow \text{ɨ} / \begin{bmatrix} -\texttt{son} \\ -\texttt{cont} \\ +\texttt{cor} \end{bmatrix} \_ \begin{bmatrix} -\texttt{son} \\ -\texttt{cont} \\ +\texttt{cor} \end{bmatrix}$ (epenthesis)

b. $[-\texttt{son}] \rightarrow [-\texttt{voi}]/[-\texttt{voi}]\_\#$ (voicing assimilation)

The first rule epenthesizes a vowel between two adjacent coronal stops. The second rule devoices an obstruent after a voiceless obstruent when they appear in a cluster at the end of a word. It is standardly assumed that the underlying representation of the past tense morpheme is /d/. Because of this, the epenthesis must be ordered before the devoicing rule, otherwise we would incorrectly predict an output form [voʊɾɨt] for *voted*. Table 4.5 shows example derivations.

| | /voʊt + d/ | /tæg + d/ | /gɹis + d/ |
|---|---|---|---|
| epenthesis | voʊtɨd | – | – |
| voicing assimilation | – | – | gɹist |
| | [voʊɾɨd] | [tægd] | [gɹist] |

Table 4.5: Example rule derivations for English past tense morpheme alternations.

$$
\begin{aligned}
\varphi_{\texttt{domain}} &\overset{\text{def}}{=} \texttt{TRUE} \\
C &\overset{\text{def}}{=} \{1,2\} \\
\varphi^{1}_{\texttt{license}}(x) &\overset{\text{def}}{=} \texttt{TRUE} \\
\varphi^{2}_{\texttt{license}}(x) &\overset{\text{def}}{=} \begin{array}{l}(\mathsf{t}(x) \vee \mathsf{d}(x)) \wedge \\ \exists y[x \lhd y \wedge (\mathsf{t}(y) \vee \mathsf{d}(y))]\end{array} \\
\varphi^{1}_{\texttt{t}}(x) &\overset{\text{def}}{=} \mathsf{t}(x) \\
\varphi^{1}_{\texttt{d}}(x) &\overset{\text{def}}{=} \mathsf{d}(x) \\
\varphi^{1}_{\texttt{ɨ}}(x) &\overset{\text{def}}{=} \texttt{FALSE} \\
\varphi^{1}_{\texttt{P}}(x) &\overset{\text{def}}{=} \mathsf{P}(x) \\
\varphi^{1}_{\texttt{B}}(x) &\overset{\text{def}}{=} \mathsf{B}(x)
\end{aligned}
$$

$$
\begin{aligned}
\varphi^{1,1}_{\lhd}(x,y) &\overset{\text{def}}{=} x \lhd y \wedge \neg\varphi^{2}_{\texttt{license}}(x) \\
\varphi^{1,2}_{\lhd}(x,y) &\overset{\text{def}}{=} x = y \wedge \varphi^{2}_{\texttt{license}}(x) \\
\varphi^{2,1}_{\lhd}(x,y) &\overset{\text{def}}{=} x \lhd y \wedge \varphi^{2}_{\texttt{license}}(x) \\
\varphi^{2,2}_{\lhd}(x,y) &\overset{\text{def}}{=} \texttt{FALSE} \\
\varphi^{2}_{\texttt{t}}(x) &\overset{\text{def}}{=} \texttt{FALSE} \\
\varphi^{2}_{\texttt{d}}(x) &\overset{\text{def}}{=} \texttt{FALSE} \\
\varphi^{2}_{\texttt{ɨ}}(x) &\overset{\text{def}}{=} \begin{array}{l}(\mathsf{t}(x) \vee \mathsf{d}(x)) \wedge \\ \exists y[x \lhd y \wedge (\mathsf{t}(y) \vee \mathsf{d}(y))]\end{array} \\
\varphi^{2}_{\texttt{P}}(x) &\overset{\text{def}}{=} \texttt{FALSE} \\
\varphi^{2}_{\texttt{B}}(x) &\overset{\text{def}}{=} \texttt{FALSE}
\end{aligned}
$$

Table 4.6: Logical formulas for epenthesis rule

Each rule can be turned into a logical interpretation as discussed in section 4.2.3. Rather than defining the rules for every symbol, I will instead look at equivalence classes. The subset of segments that will be defined are /t/, /d/, /ɨ/, /P/, and /B/, where /P/ is all voiceless sounds other than [t], and /B/ is all voiced sounds other than [d] and [ɨ]. Since rule (19a) involves epenthesis, there needs to be two copy sets. All of the necessary formulas are shown in Table 4.6. Of note is that most copy set formulas are simply `FALSE` since the only material being inserted is the epenthetic vowel. Unlike above, it is required to have redundancy between the licensing formula and the formula describing where the epenthetic vowel is inserted. Furthermore, the licensing predicate is used to define the output ordering. The input order is maintained unless an element is licensed on the second copy set. From here the method is just "down and over" to correctly insert the epenthetic vowel in the necessary position of the string.

Most of the formulas are either identity mappings or evaluate to `FALSE`. The one that matters is the formula dictating where the [ɨ] vowel is inserted. Since model-theoretic interpretations rely on identifying input structure as the source for determining output

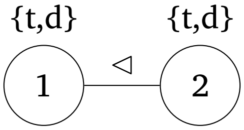


Figure 4.14: Sub-structure that triggers epenthesis based on rule account

$$
\begin{aligned}
\varphi_{\mathtt{domain}} &\overset{\text{def}}{=} \mathtt{TRUE} \\
C &\overset{\text{def}}{=} \{1\} \\
\varphi_{\mathtt{license}}(x) &\overset{\text{def}}{=} \mathtt{TRUE} \\
\varphi_{\triangleleft}(x,y) &\overset{\text{def}}{=} x \triangleleft y \\
\varphi_{\mathtt{t}}(x) &\overset{\text{def}}{=} \mathrm{t}(x) \vee (\mathrm{d}(x) \wedge \exists y[y \triangleleft x \wedge \mathrm{t}(y) \vee \mathrm{P}(y)]) \\
\varphi_{\mathtt{d}}(x) &\overset{\text{def}}{=} \mathrm{d}(x) \wedge \neg\exists y[y \triangleleft x \wedge \mathrm{t}(y) \vee \mathrm{P}(y)] \\
\varphi_{\mathtt{i}}(x) &\overset{\text{def}}{=} \mathtt{FALSE} \\
\varphi_{\mathtt{P}}(x) &\overset{\text{def}}{=} \mathrm{P}(x) \\
\varphi_{\mathtt{B}}(x) &\overset{\text{def}}{=} \mathrm{B}(x)
\end{aligned}
$$

Table 4.7: Logical formulas for voicing assimilation rule

structure, we can define a sub-graph that will lead to epenthesis occurring. Figure 4.14 shows this exact sub-structure. Despite the quantifier used in the formula, it is still a local object since the quantification is being used to find a neighboring element rather than scanning the entire string. The sub-structure is just a sequence of two coronal plosives. This will be useful when comparing across theories later in this chapter.

Table 4.7 shows the logical characterization of the assimilation rule. To maintain focus on the past tense alternation devoicing is only described as it affects /t/ and /d/. This is to facilitate analysis and to prevent getting any more bogged down in technical details. The rule technically targets the obstruent subset of /P/ and /B/ as well.

Again, we can discuss a local substructure that is driving devoicing. In this instance, it is a /d/ preceded by either a /t/ or a /P/. Because of rule ordering, and the assumption that the epenthesis rule happens first, there will never be a preceding /t/. But the rule

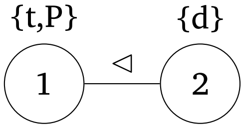


Figure 4.15: Sub-structure that triggers voicing assimilation/devoicing based on rule account

description has no knowledge of ordering and other rules in the system. It is simply a set of instructions to be carried out. Figure 4.15 shows this sub-structure in graph form.

### 4.4.2 Optimality Theory-based account

The Optimality Theory (OT; Prince and Smolensky, 1993) account follows Baković (2005). To account for the assimilation process, the Markedness constraint AGREE(VOICE) is ranked above the Faithfulness constraint IDENT(VOICE). To account for epenthesis, the constraint NOGEM, which is a Markedness constraint banning sequences of identical segments, is ranked above the Faithfulness constraint DEP(V). This alone cannot account for epenthesis when the stem ends in /t/. Baković points out that if NOGEM and AGREE(VOICE) are both ranked above DEP(V), then they work together to force epenthesis in stems ending in /t/. This is because a candidate where the past tense morpheme devoices ([tt]) would violate NOGEM while a candidate where the past tense morpheme stays faithful ([td]) would violate AGREE(VOICE). He also glosses over the fact that tapping for both /t/ and /d/ occurs in this position, but as shown in the tableaux below, this does not affect the current analysis as the tapping itself is a consequence of needing to satisfy NOGEM and AGREE(VOICE).

(20) *Epenthesis with stem-final /d/*

| /voʊt+d/ | NoGem | Agree(Voice) | Dep(V) | Ident(Voice) |
|---|---|---|---|---|
| a. voʊtd | | ∗! | | |
| b. voʊtt | ∗! | | | ∗ |
| ☞ c. voʊɾɨd | | | ∗ | |

(21) *Epenthesis with stem-final /t/*

| /loʊd+d/ | NoGem | Agree(Voice) | Dep(V) | Ident(Voice) |
|---|---|---|---|---|
| a. loʊdd | ∗! | | | |
| b. loʊdt | | ∗! | | ∗ |
| ☞ c. loʊɾɨd | | | ∗ | |

(22) *Faithful with stem-final voiced segment*

| /tæg+d/ | NoGem | Agree(Voice) | Dep(V) | Ident(Voice) |
|---|---|---|---|---|
| ☞ a. tægd | | | | |
| b. tægt | | ∗! | | ∗ |
| c. tægɨd | | | ∗! | |

(23) *Devoicing with stem-final voiceless segment*

| /gɹis+d/ | NoGem | Agree(Voice) | Dep(V) | Ident(Voice) |
|---|---|---|---|---|
| a. gɹisd | | ∗! | | |
| ☞ b. gɹist | | | | ∗ |
| c. gɹisɨd | | | ∗! | |

For the OT analysis we can make the following observations: if there is a sequence of two coronal plosives on the input then changing the voicing valuation will not make a difference in harmony. The only way to salvage a candidate is by epenthesizing due to higher ranking Dep. Once this is the case the determination of the voicing value for the past tense morpheme is controlled by IDENT(VOICE). Devoicing the morpheme is unnecessary as the epenthetic vowel removes the configuration where AGREE(VOICE) would come into play. Changing voicing is only necessary if epenthesis does not occur. When there is a sequence containing at least one non-coronal plosive, the high ranking NOGEM constraint does not even come into play, leaving the interaction of AGREE(VOICE) and IDENT(VOICE) to determine the winner. This time, because there is no epenthesis, AGREE(VOICE) fully determines the winning candidate by making sure the final obstruent matches in voicing with the preceding segment. Thus, determining the voicing of the past tense morpheme $x$ based on this OT grammar can be described in the following way:

$$\texttt{IF}\ \textsc{NoGem}(x)\ \texttt{THEN}\ \top\ \texttt{ELSE}\ \exists y[y \lhd x \land \texttt{voice}(y)] \tag{4.59}$$

Boolean Monadic Recursive Schemes (BMRS; Bhaskar et al., 2020; Chandlee and Jardine, 2021; Bhaskar et al., 2023) are a formalism that can be used to describe string to string transductions in the same way that first-order logic has been used thus far. While there are differences in the types of patterns that can be expressed with BMRS vs. first-order logic, one primary difference is the way in which logical connectives are encoded. In first-order logic, the connectives are the base primitives that expressions are built with. In BMRS, logical connectives emerge through the use of base `IF...THEN...ELSE` syntax. For example, logical `AND` (P∧Q) is equivalent to `IF` $P$ `THEN` $Q$ `ELSE` $\bot$. Logical `OR` is defined similarly and shows a parallel with the OT formulation above:

$$\texttt{IF}\ P\ \texttt{THEN}\ \top\ \texttt{ELSE}\ Q \tag{4.60}$$

Here, $P$ is equivalent to NOGEM$(x)$. In other words, $P$ can be thought as a statement that says there is a sequence of two coronal plosives. If the entire sentence determines the voicing property of an output structure, then when this is true, the second segment (the /d/ past tense morpheme) will automatically be voiced. $Q$ is equivalent to $\exists y[y \lhd x \land$

$\texttt{voice}(y)]$ which returns true or false depending on the voicing valuation of the preceding segment. This means it is possible to write an equivalent output voicing function with a disjunction using first-order logic:

$$\varphi_{\texttt{voice}}(x) \overset{\text{def}}{=} \textsc{NoGem}(x) \vee \exists y[y \lhd x \wedge \texttt{voice}(y)] \tag{4.61}$$

This formula can be updated to match the alphabetic symbols and approach taken in the rules section above. NOGEM can be expanded to $\exists y[y \lhd x \wedge ((\mathtt{t}(x) \wedge \mathtt{t}(y)) \vee (\mathtt{d}(x) \wedge \mathtt{d}(y)))]$[7] and $\exists y[y \lhd x \wedge \texttt{voice}(y)]$ can be expanded to $\exists y[y \lhd x \wedge (\mathtt{d}(y) \vee \mathtt{B}(y))]$. Furthermore, the second half can switch out d and B for t and P and flip the negation since these are just a voiceless/voiced partition that works either way. This leaves us with:

$$\varphi_{\mathtt{d}}(x) \overset{\text{def}}{=} \exists y[y \lhd x \wedge ((\mathtt{t}(x) \wedge \mathtt{t}(y)) \vee (\mathtt{d}(x) \wedge \mathtt{d}(y)))] \vee \neg\exists y[y \lhd x \wedge (\mathtt{t}(y) \vee \mathtt{P}(y))] \tag{4.62}$$

The left disjunct here targets the substructure shown in Figure 4.14 while the right disjunct targets essentially the substructure in Figure 4.15 above. What is potentially noteworthy here is that for sequences containing the past tense morpheme (Bd, Pd, td, dd), the only part of the left disjunct that matters is $\mathtt{d}(x)$ which would reduce the formula to exactly the same conditions maintaining faithful voicing as in $\varphi_{\mathtt{d}}(x)$ in the rule-based account above. It is left to be seen if the other aspects of the formula can be reduced across the board or if it makes different predictions from the rule-based account.

A similar approach can be taken to understand the structural conditions on when epenthesis occurs. Baković (2005) frames epenthesis as an instance of *anti-gemination* driven by similarity avoidance. This is a specific case where it is partial identity that drives the process. The ranking NOGEM, AGREE(VOICE) $\gg$ DEP(V) $\gg$ IDENT(VOICE) ensures that epenthesis happens to prevent either identical obstruent sequences or obstruent sequences that match for all features except for [voice]. In this instance, we don't need a BMRS

---

[7] NOGEM, despite its name, is satisfied when the geminate structure exists.

statement since it will evaluate to `True` if the condition holds and `False` otherwise which is the same as an FO statement with just the condition. Limiting the definition to coronals, the statement which captures the OT constraint ranking is *if there is a sequence of [tt] or [dd] that violates NOGEM or if there is a sequence [td] or [dt] that violates IDENT(VOICE) then epenthesize a vowel, otherwise don't.* As pointed out by Chandlee and Jardine (2021, p. 487), these types of statements capture the particular structures that *license* or *block* certain processes from occurring. This statement licenses epenthesis but blocks voicing assimilation when two coronals are adjacent in the same way intended by the constraint ranking. Its logical representation is as follows.

$$\varphi_{\mathtt{i}}(x) \overset{\text{def}}{=} (t(x) \vee d(x)) \wedge \exists y[x \lhd y \wedge (t(y) \vee d(y))] \tag{4.63}$$

It is instructive to compare 4.63 with $\phi^2_{\mathtt{i}}(x)$ in Table 4.6. 4.63 describes the same set of substructures as $\phi^2_{\mathtt{i}}(x)$ in Table 4.6 (which is visualized in Figure 4.14) and triggers epenthesis if and only if the input string contains one of these substructures. This highlights the fact that the base relation between input and output structure in both OT and rule-based analyses of this process are the same. The difference therefore seems to exactly lie in the (pseudo)algorithm used to determine the relation between the pairs: string rewriting vs. optimizing over an infinite candidate set. This brings about further computational distinctions at a less abstract level, but discussion of this lies outside the scope of this dissertation. Furthermore, the typological predictions will vary between the two accounts, but once again the point being made here is about the declarative computation of a particular process, and at that level there is convergence.

While an OT grammar is defined in terms of constraint interaction, the maps that it defines can still be described declaratively in the same spirit as the rule grammar from above. This point is made by Chandlee and Jardine (2021) in relation to BMRS but it extends to first-order logic as well. Furthermore, since the OT analysis can be couched in the same formal language as the other analyses, it was possible to discuss the relationship between the rule account and the constraint account. I move now to the Articulatory Phonology account.

### 4.4.3 Articulatory Phonology-based Account

Goldstein (2011) provides an overview of how to account for the past-tense alternation in Articulatory Phonology. First, he establishes that the past-tense morpheme has a single form in terms of a coupling graph: a tongue tip closure gesture with corresponding release gesture.[8] Second, he uses modeling to show that this single coupling graph can capture the acoustic behavior of the three forms. It relies on the following three rules for combining the past tense coupling graph with the coupling graphs for stems:

1. Couple the TT closure gesture of the suffix coupling graph anti-phase to a stem-final vowel constriction gesture.
2. Couple the TT closure gesture of the suffix eccentric phase to a stem-final non-coronal obstruent constriction gesture.
3. Couple the TT closure gesture of the suffix eccentric phase to a stem-final coronal obstruent release gesture.

The first two rules are the standard ways in which coda consonants combine with previous material: a coda C is in phase if it directly follows a, otherwise it is eccentrically-phased with the preceding C's constriction gesture. These account for the non-epenthetic cases of past tense formation. Epenthesis in this account results from the extra space between gestures provided by coupling the past-tense coupling graph to the *release* portion of the stem-final C rather than the constriction portion. Unlike the previous accounts, there is

---

[8] There is also a closed velum gesture in his analysis. The inclusion of velum closing gestures versus just velum opening gestures varies throughout the Articulatory Phonology literature. Since I neglected velum closing gestures above, I will continue to do so now. From a representational perspective, the absence of this gesture causes no issues, but from a gestural/dynamic perspective, it does make a difference.

no true insertion of a target here, but instead, a targetless vocalic element emerges from the coordination pattern.

To account for the deviation, Goldstein suggests using Gafos and Benus's (2006) implementation of constraint interaction in dynamical systems. The tilted anharmonic oscillator has form $\dot{x} = f(x, R) = R + x - x^3$ where $R$ is a continuous control parameter. When $R$ is 0, there are two possible attractor states that the system can end up in. When $R$ is negative, the system ends up in one stable state such as CLO-CLO coupling, and when $R$ is positive, the system ends up in the other state such as CLO-REL coupling. Goldstein therefore suggests relating $R$ to the relative coupling strength of the past tense closure gesture to the preceding consonants constriction or release gestures. The reason for using the ratio is to ensure that $R$ is a continuous parameter, but it also involves stating that the past tense coupling graph is coupled to both gestures, albeit with different strengths. When it has a stronger coupling relation with the closure gesture, it ends up in the CLO-CLO state, and vice versa when there is a stronger coupling relation with the release gesture.

Abstracting away from the specific dynamical system implementation, the descriptive generalization is that coupling graphs where the constriction degree does not change anywhere in the vocal tract between the stem and affix are dispreferred/marked/banned/etc. Procedurally, the process first involves combining a stem coupling graph with the past tense coupling graph. Suppose this process follows the standard rules of the language: couple the constriction gesture of the past tense coupling graph anti-phase with a vowel-final stem gesture or eccentric phase with *any* consonant final stem's constriction gesture. This process operates the same way as morphological concatenation in string-based analyses. Marked structures emerge because this point of the process simply involves basic combination. The next step evaluates the structure and makes any necessary changes to avoid the banned structure.

While viewing this as a banned substructure is one way to interpret the phenomenon,

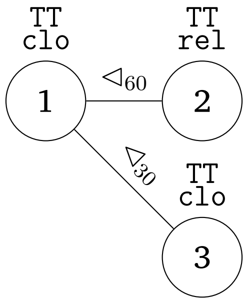


Figure 4.16: CLO-CLO banned substructure

it can be similarly viewed as conditions on a transformation. As was the case for previous analyses, it is possible to use declarative statements in first-order logic to define the interpretation of one structure in terms of another. In this case, the output structure is one in which the TT CLO gesture of the affix is coupled eccentrically to the TT REL gesture. Therefore, the logical statement should evaluate true for an output predicate $\varphi_{\triangleleft_{30}}(x, y)$ only for nodes $x = 2$ and $y = 3$. This can be defined as follows.

$$\varphi_{\triangleleft_{30}}(x, y) \overset{\text{def}}{=} \texttt{TT}(y) \wedge \texttt{clo}(y) \wedge \exists z[z \triangleleft_{30} y \wedge \texttt{TT}(z) \wedge \texttt{clo}(z) \wedge \exists w[z \Diamond w] \wedge z \triangleleft_{60} x] \tag{4.64}$$

To rule out coronal nasal sequences, a second set of quantifiers can be used to ensure there is no in-phase velum gesture with either $x$ or $y$. What is possible now that this is in the language of first-order logic is to translate this structure driving the recoupling of the past tense morpheme with the structure driving epenthesis in the rule-based and constraint-based accounts. One thing to note is that this translation will describe multiple structures since it occurs at the level of a segment for the rule-based account and not at the level of the feature. But this is desirable. The structure in figure 4.14 contains a set of substructures since it accounts for /tt/, /td/, /dt/, and /dd/. The structure in figure 4.16 is expanded in the following way. First, the release gesture is not licensed in the output structure by equation (4.58). Second, a general successor ordering relation is determined between domain elements 1 and 3 by equation (4.57). Finally, domain elements 1 and 3 can be considered to be sub-properties of both /t/ and /d/ since they are not connected to a velum gesture and are the only tongue-tip closer gestures in the translation. Therefore, the same coupling graph structure that leads to epenthesis is interpreted in first-order logic as the same structure that drives epenthesis in the symbolic string approach.

There is one problematic aspect to this account as it pertains to the physical epenthetic material. The translation from a string to a coupling graph will currently not produce the same coupling graph because coupling the coda closure gesture to the preceding release gesture is not standardly defined. This shows how the model-theoretic approach can also highlight differences. In the account from Goldstein (2011), the epenthetic vowel is not a target gesture and instead emerges from the short gap between gestures provided by this special coupling relation. There is support for this as well as Flemming and Johnson (2007) show that the formant properties of intra-coronal reduced vowels are different from other reduced vowels in English. That being said, the use of [ɨ] rather than [ə] in this account provides a way to update the string to coupling graph representation. A symbolic sequence of [ɨ] between two coronals can be interpreted as requiring this special type of coupling rather than the insertion of a gestural node in the graph. I leave this for future work.

This section highlighted the similarities and differences between different theories of the past tense alternation in English by using model-theoretic representations in first-order logic. Strikingly, while the Articulatory Phonology approach couches its analysis in continuous dynamical systems, the computational properties driving the restructuring of the coupling graph can be reduced to the same as the symbolic approach: a locally ordered sequence of coronal plosives. Therefore, the actual input-output computation needed to account for this alternation is no different whether it is performed with symbolic string representations or with coupling graph representations. One area where this account may diverge has to do with coupling strength. Goldstein (2011) does derive the crucial parameter by taking a ratio of coupling strengths between the closure gesture of the past tense morpheme and the closure and release gestures of the stem-final consonants. But ultimately what determines the state of the system is the discrete aspect of this ratio: is it more coupled to the closure? If so, then “epenthesize” by restructuring the coupling graph. Is it more coupled to the release? If so, then don’t restructure.

## 4.5 Summary

This chapter began with an overview of Articulatory Phonology and a discussion of how model theory can act as a formal system for understanding phonological knowledge. I then showed how to use model theory to represent the dynamic coupling graph representations used in Articulatory Phonology. The outcome of this was that it provided a way to use first-order logic to interpret coupling graph representations as segmental strings and vice-versa. This chapter specifically showed how to translate between strings and coupling graphs in both directions. The logical programs I defined provide isomorphisms between the two types of structures, therefore showing that the same distinctions can be made in both representation schemes within the power of first-order logic. Furthermore, this means that insights from each approach can be translated into the other.

After showing how to translate between the different types of structures, I then looked at the specific phenomenon of English past tense morphophonological alternations. I showed how to turn rule and OT-based approaches to the alternations into logical statements. The logic-based approach to phonological transformations is about satisfaction. Given this substructure exists in the input, give the output this property. I showed how the rule and OT accounts reduce to looking for the same type of substructures, regardless of differences in how the theories algorithmically determine the output. More importantly, I discussed the Articulatory Phonology account of these alternations and found that to account for epenthesis, one must also posit a grammar that changes the coupling graph structure. This can also be interpreted logically and the outcome is that the input properties that need to be satisfied for the computation in dynamic coupling graphs are the exact same substructure as in the symbolic approaches: a locally ordered sequence of coronal plosives. The translations between string and coupling graph structures are what allow for this to be verifiably true.

What is the consequence for theories of the phonetics-phonology interface given this result? First, it provides concrete support for phonetic properties such as articulatory gestures not needing to be directly encoded into long-term memory representations as the properties can be inferred from linear strings. This follows the trend from previous chapters of pointing out that systematic continuous behavior can emerge from discrete knowledge given we have a rigorous understanding of how phonological and phonetic information are integrated. On the flip side, since the string and coupling graphs are bidirectionally translatable, it also supports one of the main points Articulatory Phonology uses in arguing that discrete behavior is present in the dynamical systems that it uses. While this is true, and the dynamical equations used in Articulatory Phonology certainly do give us low and high-level descriptions of speech sounds, this chapter shows concretely, as was discussed in Chapter 2, the segmental string is available as an even higher level of description, even if indirectly.

That being said, the classic version of AP still struggles with accounting for categorical alternations. The logical approach discussed in this chapter provides a path forward in this area. Given that I have provided the programs needed to translate symbolic representations into gestural representations, any phonological process that can be encoded into a symbolic rule can now be translated into an equivalent set of logical equations using only gestural representations. These will provide the substructures that are used to compute the desired output forms. Of course, this will be at the qualitative/discrete level. A fully dynamical approach would need some type of continuous parameter controlling the categorical behavior. In this chapter, I showed that to account for these types of phenomena, AP essentially implements the same type of grammar that maps an input structure to an output structure based on the properties of the input structure. Doing this over strings or doing it over coupling graphs doesn't change the base computational complexity. Regardless of the representational encoding, we therefore need some type of structure-to-structure function that acts as the "phonology" in the classic sense. It likely will come down to ontological commitment and aesthetic preferences for what type of representations should be used in these cases.

While I take the personal view that the results of this chapter argue in favor of not encoding the articulatory properties directly in the long-term memory representations, there are many results from the AP research program that need to be taken seriously by symbolic theories of phonology. First of all, AP has shown that many allophonic phenomena can emerge directly from the gestural dynamics (Browman and Goldstein, 1992). Processes like aspiration or vowel nasalization in English likely do not need to be encoded into the grammar directly as these phenomena will show up through the phonetic implementation rules of a given language. Second, AP has identified phenomena like speech errors, even categorical seeming *spoonerisms*, to be driven largely by timing errors rather than wholesale symbolic rearrangement.

Relatedly, there has been work in the AP tradition on gestural hiding (Browman and Goldstein, 1990; Beckman, 1996). For example, Browman and Goldstein (1990, p. 6) point to English phrases like *must be*, *ground pressure*, and *perfect memory* where interconsonantal /t/ is not present in casual/quick speech. Examination of x-ray microbeam data (Fujimura et al., 1973; Miller and Fujimura, 1982) containing these types of phrases shows the presence of a /t/ gesture even when there is no /t/ acoustic percept. The magnitude of this gesture was similar to a perceived /t/ gesture in other recordings. Therefore, it has been argued that this is not a "deletion" rule and instead just gestural overlap.

At the same time, work on word-final/t/ assimilation in phrases like *late calls* (Nolan, 1992), it was observed that the magnitude of the alveolar gesture was reduced. While the presence of the gesture in the output falls in line with the AP story of gestures not being "deleted" in the classic sense, the reduction in magnitude is harder to straightforwardly analyze in the framework. On the other hand, these types of data are straightforwardly predicted if we assume a production architecture like the blueprint model of production (Chapter 3) and a symbolic rule of place assimilation. The retention of some /t/-like gestural properties emerges from the phonetic output being a blend of input and output form. Therefore, a gestural representation is not needed to account for the retention of a lexical gesture specification.

This reiterates the point that there is a critical interaction between computational structure and representational choices. The modular feed-forward architecture of the interface does not offer a solution for the phenomenon of gesture magnitude reduction since there can be no influence of lexical representation on the final phonetic output. On the other hand, the blueprint model of production architecture allows for additional support in discerning the role of representation in accounting for this phenomenon. The ability to retain lexical information in the phonetic output under the architecture of the blueprint model of production is predicted *regardless of the choice of representation*. It is sufficient, once again, to assume a discrete and symbolic phonological representation, and still account for these types of systematic continuous phenomena.

# Chapter 5

# Conclusion

## 5.1 Summary of Results

In this dissertation, I took a computational approach to understanding the phonetics-phonology interface. This took the form of both providing a characterization of the computational structure of the interface in terms of language production as informed by generative theories of linguistics and using tools from the theory of computation to argue that phonetic substance such as articulatory plans do not need to be encoded directly into long term memory representations. While the demarcation of phonological and phonetic knowledge is hardly a new problem for linguists, the computational approach taken here does provide a novel way of understanding this relationship. I further argued that computational formalization is necessary because it provides a way to identify hidden factors affecting theoretical interpretations. Given that the proliferation of phonetic substance into phonological theories is largely driven by experimental findings, I also argued that empirical investigations can and should be driven by strong theoretical commitments. The best way to have these commitments stated clearly is through the type of computational work done in this dissertation.

Specifically, I introduced the Blueprint Model of Production. It is characterized based on typed functions and provides a high-level, “computational”, description of how the different functions (lexicon, phonology, phonetics) interact. The classical modular feed-forward view is shown to be a composed function where phonology acts as an intermediary between the lexicon and phonetics, while in the blueprint model both the lexicon and phonology are separate arguments to a higher-order phonetics function. This characterization allows for the view of phonology as a function mapping URs to SRs to remain but also provides a way for lexical information to directly affect the phonetic exponent. By modeling the interface this way, it was shown that phenomena like incomplete neutralization and variation in homophone duration do not invalidate discrete theories of phonological knowledge as previously argued. This was supported with various simulations of different natural language phenomena such as final devoicing in German, tonal alternations in Cantonese, and vowel epenthesis in Lebanese Arabic. In each case, the data were simulated assuming discrete phonological knowledge but could generate the observed systematic continuous values. While discrete models of phonology largely do not appear compatible with modular feed-forward models of production in accounting for these types of phenomena, these phenomena do appear quite natural under the blueprint model of production.

Furthermore, I showed the equivalency of string-based representations of lexical items and dynamic coupling graph-based representations of lexical items within first-order logic. The model-theoretic interpretation shows that all of the dynamical gestural information that is encoded in coupling graphs can be inferred directly from the string-based representation using a relatively restricted amount of computational machinery. It was speculated that the computational power needed to define the interpretation could be further reduced to a class that would suggest not only equivalency within first-order logic but match what previous authors have argued for satisfying *notational equivalency*. I also showed that grammatical accounts of alternations using the coupling graph representations ultimately recapitulate analyses based on string representations. While there are ontological and philosophical distinctions between traditional phonological accounts using rules or constraints and the dynamical-based accounts championed by Articulatory Phonology, what this dissertation showed is that there is quite a bit of overlap when we abstract to the actual structure-to-structure computation involved in all cases.

The computational account also shows that it is possible to compare what are often thought to be disparate theories and therefore opens up an avenue for further inquiry. Model theory acts as a shared meta-language that provides insights into the expressivity and computational power required to represent both individual theories of phonological and phonetic knowledge, but also between different theories. Ultimately it provides a tool that allows for formal reasoning about which parts of a given theory are unique and which parts might be more universal. By looking for universal properties across theories, we can come to have a clearer picture of the platonic object of study. The computational approach taken in this dissertation provides these types of insights in ways that are difficult for other approaches.

## 5.2 Looking Ahead

This dissertation has laid a foundation for future research both specifically in terms of immediate follow-ups to the main findings in Chapters 3 and 4, as well as more long-term projects in related domains. The BMP provides a new framework to look at other phenomena not covered in Chapter 3 such as boundary effects, deletion processes, and absolute neutralization. In Chapter 4, I suggested that the logical interpretations could be made quantifier-free by using a logic that contained set membership. Both of these ideas can be immediately pursued. There are also new domains and phenomena that I believe can be fruitfully pursued using the broad idea that a formal, computational characterization of the phonetics-phonology interface will provide meaningful insight. These include models of speech perception, sound change, loanword adaptation, integrating sociolinguistic information, and developing phonological and phonetic learning algorithms. In the remainder of this section, I discuss each of these in more detail.

The next direct step will involve characterizing the input/perception function using typed functions in the same way that the blueprint model of production characterized the output/production function. The goal once again is to characterize the way that different types of information are integrated during the perception process. In general, there are varying views on how phonological information plays a role in perception, with very few examples of a phonological perception grammar. Boersma (1998), Boersma et al. (1999), Boersma (2009), Boersma et al. (2020) and Broselow (2009) are notable exceptions to the lack of perception grammars. In general, the view I currently hold is largely one that perception affects the phonological grammar indirectly as exemplified in Hume and Johnson (2001).

Deciding how exactly phonological and phonetic information combine to influence perception is therefore the next primary question I seek to answer. A starting point here may be the general architecture and ideas proposed by Kazanina et al. (2018, figure 2). The phenomenon that I think will be most fruitful for doing the characterization is illusory vowel perception (Dupoux et al., 1999). The classic example of this is Japanese speakers being played the nonceword [ebzo] but hearing [ebɯzo]. Illusory vowels are a worthwhile topic for thinking about how different types of information are integrated in the perception process because there is a wide rang of previous research on their relationship to phonotactic constraints (Kabak and Idsardi, 2007), phonological alternations (Durvasula and Kahng, 2015), phrasal/prosodic phonology (Durvasula and Kahng, 2016), as well as factors such as transition probability (Kilpatrick et al., 2021) and frequency (Wilson, 2017). The remainder of this section lays out other interesting avenues to pursue following the work done in this dissertation.

Sound change is a phenomenon that has been argued to be both phonetically driven as well as phonologically driven. Therefore, this is an area where an explicit formulation of the phonetics-phonology interface will have implications on how each type of information affects the outcome. Furthermore, there is both *phonological* sound change where broad categories change and *phonetic* sound change where the phonetic exponent of a category changes.

Two types of sound change that are commonly referenced are Neogrammarian sound change where a sound change is regular and exceptionalness and lexical diffusion where change spreads throughout the lexicon at varying rates. As Bermúdez-Otero (2007) points out, there is a third view of sound change as being both lexically gradual and phonetically gradual. Exemplar models of the grammar propose that each lexical item is represented as a cloud of previously heard tokens of that word (Bybee, 2001). Different types of exemplar models have been proposed to explain sound change (*e.g.,* Wedel, 2006; Tupper, 2015; Todd et al., 2019), but there is no uniform proposal on how phonology should be included in these types of models. Evidence for there being both phonological and phonetic sound change suggests that radical exemplar models with no formal phonology are likely incapable of modeling the endpoints of the s-shaped curve often found in language change. If there is no underlying phonological system then it begs the question what is speeding up the change in the physical qualities of the sound change as it nears completion?

Merger of vowel contrasts is an area where phonological and phonetic aspects collide and is an empirical phenomenon that may prove to be fruitful for this type of work. In Chapter 3, certain cases of near-merger were already given a novel characterization based on the blueprint model structure of the interface. Another phenomenon to investigate is the merger of initial /n/ and /l/ in Hong Kong Cantonese (Zee, 1999). Rather than acoustic similarity driving merger as is the case for vowel mergers, this merger has been argued to be a result of low functional load maintained by the contrast (Wedel et al., 2013). This provides a second example to show how different types of information, both grammatical and extra-grammatical, can play a role in sound change.

Loanword adaptation is potentially influenced by a multitude of factors (Kang, 2011). Two interacting factors are phonological and phonetic information. One explanation for loanword adaptation patterns is that they are a result of interpreting foreign structures using native language speech perception processes (Peperkamp and Dupoux, 2003; Peperkamp, 2004; Peperkamp et al., 2008, *inter alia*). In other words, adaptation patterns are based on acoustic properties rather than, say, the phonological properties of a given word in the lending language. LaCharité and Paradis (2005) argue against the perception-based account of adaptation partly due to how certain English vowels are adapted into French. They show that in the Project CoPho database, English /ɪ/ and /ʊ/ are overwhelmingly adapted as /i/ and /u/ when adapted into French. For Parisian French, this mapping occurs 100% of the time for both vowels. In Quebec French, this mapping occurs $>$98% of the time for both vowels. So in both dialects, this adaptation pattern is very robust. Using vowel formant data from Delattre (1969) for Parisian French and Martin (2002) for Quebec French, they show that this adaptation pattern does not make sense since the vowels /e/ or /ɛ/ and /o/ or /ɔ/ are closer in the F1-F2 vowel space.

This is another area of inquiry that relies on the specific characterization of the phonetics-phonology interface. One specific aspect that could be explored more is how different dialects may vary in their adaptation patterns for the same sound. Sticking with French, there is variation in the adaptation of /θ/. It is adapted as /s/ in European French while it is adapted as /t/ in Quebec French. Both dialects of French have the same consonant inventories, but the phonetic properties of the sounds in the two languages vary (Brannen, 2012). These facts also suggest a possibility of production and perception interacting which has not been addressed up to this point. I have given a characterization of the interface where production and perception are in fully separate streams, but the production-perception loop is a well-known phenomenon that can be addressed (Poeppel and Idsardi, 2011).

Another direction to take this work is in describing how sociolinguistic variation can be synchronically realized. Some variation has been analyzed using OT grammars. For example, Anttila and Cho (1998) show how different dialects that mix /r/-dropping and intrusive /r/ can be described by re-ranking three basic constraints. On the other hand, some variation has to do with the phonetic exponent of a given phoneme. For example, the realization of /ow/ in Philadelphia varies based on gender and socioeconomic class (Labov, 2001; Labov et al., 2013). Despite this phonetic variation, there is no evidence of phonological variation. These two types of variation largely are split between categorical phonological variation and continuous phonetic variation. Most sociophonetic variation is now accounted for with exemplar models (*e.g.* Abramowicz, 2007) and a primary motivation for developing the blueprint model of production was to show that exemplar models were not needed to describe certain phenomena. Showing that the blueprint model can also account for individual variation seems like a logical next step.

Given the characterization of the interface developed in this dissertation, one can ask how each of the subcomponents is learned. Part of this is determining which aspects of the production/perception process are innate and which are learned. For example, Hale et al. (2007) make the strong claim that the transduction processes between phonetic substance and phonological structure are universal and therefore unlearned while Kingston and Diehl (1994) argue that much of phonetic implementation is learned. It is an open question what the formal nature of learning for the phonetics-phonology interface and how the different aspects (output function, input function, representation) interact in learning. To further reiterate, the answers again depend on the specific characterizations given. One advantage of the model-theoretic approach taken throughout this dissertation is that it has ties to computational complexity theory and learnability (Strother-Garcia et al., 2016; Vu et al., 2018; Chandlee et al., 2019), which should help facilitate this type of project.

## 5.3 Final Thoughts

This dissertation showed that careful consideration of the computational structure of the phonetics-phonology interface is important for understanding phonetic and phonological phenomena more broadly. The insights that are gained by approaching things in this manner are not easily obtainable using other methods. Therefore, this dissertation has also provided a new perspective for considering the relationship between phonological and phonetic knowledge. The computational approach was also argued for as a way to guide future empirical inquiry. As I have laid out in this chapter, there are many directions where this type of work can be taken, but they are all unified by depending on careful and specific characterizations of the interface. The Blueprint Model of Production is one formalization that can now be tested empirically. Furthermore, the computational approach provided insight into the nature of long-term representations. While it is possible to *account* for certain phenomena by articulatory dynamics, it is not necessary to *encode* them directly. In many ways, this dissertation is a defense of the general idea that it is possible for a theory of language to maintain more general/abstract representations as long as one is specific about the way the computations work. To emphasize this point, I end with a passage from Lewis Carroll's (1893) *Sylvie and Bruno Concluded*:

> "What do you consider the largest map that would be really useful?"
>
> "About six inches to the mile."
>
> "Only six inches! [...] We very soon got to six yards to the mile. Then we tried a hundred yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of a mile to the mile!"
>
> "Have you used it much? [...]"
>
> "It has never been spread out, yet [...] the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well."

# Bibliography

Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, 12(1):1–13.

Abramowicz, Ł. (2007). Sociolinguistics meets exemplar theory: Frequency and recency effects in (ing). *University of Pennsylvania Working Papers in Linguistics*, 13(2):3.

Albright, A. and Hayes, B. (2006). Modelling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In Fanselow, G., Féry, C., Schlesewsky, M., and Vogel, editors, *Gradience in Grammar*. Oxford University Press.

Alderete, J. and Finley, S. (2023). Probabilistic phonology: A review of theoretical perspectives, applications, and problems. *Language and Linguistics*, 24(4):565–610.

Altmann, G. and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3):191–238.

Anderson, J. and Jones, C. (1974). Three theses concerning phonological representations. *Journal of linguistics*, 10(1):1–26.

Anttila, A. (2007). Variation and optionality. In de Lacy, P., editor, *The Cambridge handbook of phonology*, pages 519–536. Cambridge.

Anttila, A. and Cho, Y.-m. Y. (1998). Variation and change in optimality theory. *Lingua*, 104(1-2):31–56.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5(2):183–207.

Archangeli, D. B. (1984). *Underspecification in Yawelmani phonology and morphology*. PhD thesis, Massachusetts Institute of Technology.

Avery, P. and Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In Hall, T. A., editor, *Distinctive feature theory*, pages 41–70. de Gruyter, Berlin.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1996). The CELEX lexical database (cd-rom).

Baese-Berk, M. and Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and cognitive processes*, 24(4):527–554.

Baković, E. (2005). Antigemination, assimilation and the determination of identity. *Phonology*, 22(3):279–315.

Beaton, M. E. (2016). Revisiting incomplete neutralization: The case of Puerto Rican Spanish. *University of Pennsylvania Working Papers in Linguistics*, 22(1):5.

Beckman, M. E. (1996). When is a syllable not a syllable. *Phonological structure and language processing: Cross-linguistic studies*, pages 95–123.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.

Bermúdez-Otero, R. (2007). Diachronic phonology. In de Lacy, P., editor, *The Cambridge handbook of phonology*, pages 497–517. Cambridge.

Bhaskar, S. (2023). Finite model theory, interpretations, and recursive schemes. Unpublished Course Notes.

Bhaskar, S., Chandlee, J., and Jardine, A. (2023). Rational functions via recursive schemes. arXiv preprint arXiv:2302.03074.

Bhaskar, S., Chandlee, J., Jardine, A., and Oakden, C. (2020). Boolean monadic recursive schemes as a logical characterization of the subsequential functions. In *Language and Automata Theory and Applications: 14th International Conference*, pages 157–169. Springer.

Bird, S. (1995). *Computational Phonology: A Constraint-Based Approach*. Cambridge University Press.

Bird, S., Coleman, J. S., Pierrehumbert, J., and Scobbie, J. M. (1992). Declarative phonology. In *Proceedings of the XVth International Congress of Linguists. Université Laval, Québec*. Université Laval, Québec.

Bishop, J. B. (2007). Incomplete neutralization in Eastern Andalusian Spanish: Perceptual consequences of durational differences involved in s-aspiration. In *Proceedings of the International Congress of Phonetic Sciences*, volume 16, pages 1765–1768.

Bloomfield, L. (1933). *Language*. Holt, Rinehart and Winston.

Boersma, P. (1998). *Functional phonology*. Netherlands Graduate School of Linguistics.

Boersma, P. (2009). Cue constraints and their interactions in phonological perception and production. In Boersma, P. and Hamann, S., editors, *Phonology in Perception*.

Boersma, P., Benders, T., and Seinhorst, K. (2020). Neural network models for phonology and phonetics. *Journal of Language Modelling*, 8(1):103–177.

Boersma, P. et al. (1999). On the need for a separate perception grammar. Available as ROA-358 from the Rutgers Optimality Archive.

Boersma, P. and Pater, J. (2016). Convergence properties of a gradual learning algorithm for harmonic grammar. In McCarthy, J. and Pater, J., editors, *Harmonic Grammar and Harmonic Serialism*. Equinox.

Bolyanatz, M. A. (2020). Evidence for incomplete neutralization in Chilean Spanish. *Phonetica*, 77(2):107–130.

Brannen, K. (2012). Differential substitution: Phonetic distance in the perception of interdental fricatives. Paper presented at Second Language Acquisition of Phonology.

Braver, A. (2014). Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua*, 152:24–44.

Braver, A. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology*, 36(1):1–36.

Braver, A. and Kawahara, S. (2016). Incomplete neutralization in Japanese monomoraic lengthening. In *Proceedings of the Annual Meetings on Phonology*, volume 2.

Broselow, E. (1984). Default consonants in Amharic morphology. *MITWPL*, 7:15–31.

Broselow, E. (2009). Stress adaptation in loanword phonology: Perception and learnability. *Phonology in perception*, 15:191–234.

Broselow, E., Chen, S.-I., and Huffman, M. (1997). Syllable weight: convergence of phonology and phonetics. *Phonology*, 14(1):47–82.

Browman, C. P. (1994). Lip aperture and consonant releases. In Keating, P. A., editor, *Papers in laboratory phonology III: Phonological structure and phonetic Form*, pages 331–353.

Browman, C. P. and Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45(2-4):140–155.

Browman, C. P. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, pages 341–376.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Browman, C. P. and Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3:219–252.

Browman, C. P. and Goldstein, L. M. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2):201–251.

Browman, C. P. and Goldstein, L. M. (1995). Dynamics and articulatory phonology. In *Mind as Motion*, chapter 7, pages 175–193. MIT Press.

Büchi, J. R. (1960). Weak second-order arithmetic and finite automata. *Mathematical Logic Quarterly*, 6(1-6).

Burton, M. W. and Robblee, K. E. (1997). A phonetic analysis of voicing assimilation in Russian. *Journal of phonetics*, 25(2):97–114.

Bybee, J. (1999). Usage-based phonology. *Functionalism and formalism in linguistics*, 1:211–242.

Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.

Carroll, L. (1893). Sylvie and Bruno concluded. In *Literature and Philosophy in Nineteenth Century British Culture*, pages 303–307. Routledge.

Cedergren, H. J. and Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, pages 333–355.

Chabot, A. (to appear). What phonology is and why it should be. In Dupre, G., Nefdt, R., and Stanton, K., editors, *The Oxford Handbook on the Philosophy of Linguistics*. Oxford University Press.

Chandlee, J. (2014). *Strictly local phonological processes*. PhD thesis, University of Delaware.

Chandlee, J., Eyraud, R., Heinz, J., Jardine, A., and Rawski, J. (2019). Learning with partially ordered representations. In *Proceedings of the 16th Meeting on the Mathematics of Language*, pages 91–101, Toronto, Canada. Association for Computational Linguistics.

Chandlee, J. and Jardine, A. (2021). Computational universals in linguistic theory: Using recursive programs for phonological analysis. *Language*, 97(3):485–519.

Chandlee, J. and Lindell, S. (2016). Local languages. Paper presented at the 4th Workshop on Natural Language and Computer Since.

Chandlee, J. and Lindell, S. (2021). Logical perspectives on strictly local transformations. unpublished manuscript.

Chao, Y. R. (1934). The non-uniqueness of phonemic solutions of phonetic problems. *Bulletin of the Institute and History of Philology*, 4(4):363–397.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on information theory*, 2(3):113–124.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English.* ERIC.

Church, A. (1932). A set of postulates for the foundation of logic. *Annals of mathematics*, pages 346–366.

Church, A. (1933). A set of postulates for the foundation of logic. *Annals of mathematics*, pages 839–864.

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, 2(1):225–252.

Clopper, C. G. and Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., and Zellers, M., editors, *Rethinking reduction: interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, pages 25–72. Mouton de Gruyter.

Coats, H. S. and Harshenin, A. P. (1971). On the phonological properties of Russian. *The Slavic and East European Journal*, 15(4):466–478.

Coetzee, A. W. (2016). A comprehensive model of phonological variation: Grammatical and non-grammatical factors in variable nasal place assimilation. *Phonology*, 33(2):211–246.

Coetzee, A. W. and Kawahara, S. (2013). Frequency biases in phonological variation. *Natural Language & Linguistic Theory*, 31:47–89.

Coetzee, A. W. and Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory*, 26:289–337.

Cohn, A. C. (1993). Nasalisation in English: phonology or phonetics. *Phonology*, 10(1):43–81.

Cohn, A. C. (2007). Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory*, 16:1–31.

Coleman, J. and Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*.

Cooper, R. P. and Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research*, 27:42–49.

Courcelle, B. (1994). Monadic second-order definable graph transductions: a survey. *Theoretical Computer Science*, 126(1):53–75.

Cummins, R. (1983). *The Nature of Psychological Explanation*. MIT Press.

Cummins, R. (2000). “how does it work?” vs. “what are the laws?” two conceptions of psychological explanation. pages 117–145.

Curry, H. B. (1980). Some philosophical aspects of combinatory logic. In Barwise, J., Keisler, H. J., and Kunen, K., editors, *The Kleene Symposium*, volume 101 of *Studies in Logic and the Foundations of Mathematics*, pages 85–101. Elsevier.

Danis, N. (2020). Yorùbá vowel deletion involves compensatory lengthening: Evidence from phonetics. *Stellenbosch Papers in Linguistics Plus*, 60(1):1–12.

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *IRAL: International Review of Applied Linguistics in Language Teaching*, 7(4):295.

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

Deshmukh, N., Ganapathiraju, A., Gleeson, A., Hamaker, J., and Picone, J. (1998). Resegmentation of SWITCHBOARD. In *ICSLP*. Syndey.

Dinnsen, D. and Garcia-Zamor, M. (1971). The three degrees of vowel length in German. *Research on Language & Social Interaction*, 4(1):111–126.

Dinnsen, D. A. and Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of phonetics*, 12(1):49–60.

Dmitrieva, O., Jongman, A., and Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of phonetics*, 38(3):483–492.

Dresher, B. E. (2009). *The contrastive hierarchy in phonology*. Number 121. Cambridge University Press.

Du, N. and Durvasula, K. (2024a). Phonetically incomplete neutralisation can be phonologically complete: evidence from Huai'an Mandarin. *Phonology*, pages 1–37.

Du, N. and Durvasula, K. (2024b). Psycholinguistics and phonology: the forgotten foundations of generative phonology. https://ling.auf.net/lingbuzz/007737.

Duhem, P. M. M. (1954). *The Aim and Structure of Physical Theory*. Princeton University Press.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in japanese: A perceptual illusion? *Journal of experimental psychology: human perception and performance*, 25(6):1568.

Durvasula, K. and Kahng, J. (2015). Illusory vowels in perceptual epenthesis: The role of phonological alternations. *Phonology*, 32(3):385–416.

Durvasula, K. and Kahng, J. (2016). The role of phrasal phonology in speech perception: What perceptual epenthesis shows us. *Journal of phonetics*, 54:15–34.

Elgot, C. C. (1961). Decision problems of finite automata design and related arithmetics. *Transactions of the American Mathematical Society*, 98(1):21–51.

Engelfriet, J. and Hoogeboom, H. J. (2001). MSO definable string transductions and two-way finite-state transducers. *ACM Transactions on Computational Logic (TOCL)*, 2(2):216–254.

Ernestus, M. (2011). Gradience and categoricality in phonological theory. In *The Blackwell companion to phonology*, pages 2115–2136. Wiley-Blackwell.

Ernestus, M. and Baayen, R. H. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. *Laboratory phonology*, 8(1):27–49.

Ernestus, M. T. C. and Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, 79(1):5–38.

Feldman, J. A. and Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive science*, 6(3):205–254.

Ferreira, F. and Nye, J. (2017). The modularity of sentence processing. In de Alimeida, R. G. and Gleitman, L. R., editors, *On concepts, modules, and language: Cognitive science at its core*, pages 63–86. Oxford University Press.

Feyerabend, P. (1965). Reply to criticism: comments on smart, sellars and putnam. In Cohen, R. and Wartofsky, M., editors, *Proceedings of the Boston Colloquium for the Philosophy of Science II*, pages 223–261. Springer.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, 18(1):7–44.

Flemming, E. and Johnson, S. (2007). Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association*, 37(1):83–96.

Fodor, J. A. (1974). Special sciences, (or: the disunity of science as a working hypothesis). *Synthese*, 28:97–115.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Fosler-Lussier, E. and Morgan, N. (1998). Effects of speaking rate and word frequency on conversational pronunciations. In *Modeling Pronunciation Variation for Automatic Speech Recognition*.

Fougeron, C. and Steriade, D. (1997). Does deletion of French schwa lead to neutralization of lexical distinctions? In *Fifth European Conference on Speech Communication and Technology*.

Fourakis, M. and Port, R. (1986). Stop epenthesis in English. *Journal of Phonetics*, 14(2):197–221.

Fowler, C. A. (2015). The segment in articulatory phonology. pages 23–43.

Fox, R. A. and Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics*, 5(1):27–34.

Friedman, H. M. and Visser, A. (2014). When bi-interpretability implies synonymy. *Logic Group Preprint Series*, 320:1–19.

Frixione, M. (2001). Tractable competence. *Minds and Machines*, 11:379–397.

Fujimura, O., Kiritani, S., and Ishida, H. (1973). Computer controlled radiography for observation of movements of articulatory and other human organs. *Computers in Biology and Medicine*, 3(4):371–384.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural language & linguistic theory*, 20(2):269–337.

Gafos, A. I. (2006). Dynamics in grammar: Comment on Ladd and Ernestus & Baayen. *Laboratory phonology*, 8(4):51.

Gafos, A. I. and Benus, S. (2006). Dynamics of phonological cognition. *Cognitive science*, 30(5):905–943.

Gafos, A. I., Charlow, S., Shaw, J. A., and Hoole, P. (2014). Stochastic time analysis of syllable-referential intervals and simplex onsets. *Journal of Phonetics*, 44:152–166.

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

Gerfen, C. (2002). Andalusian codas. *Probus*, 14(2):247–277.

Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, volume 1, pages 517–520. IEEE Computer Society.

Goldrick, M. (2000). Turbid output representations and the unity of opacity. In *North East Linguistics Society*, volume 30.

Goldrick, M. and Cole, J. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics*, 99:101254.

Goldsmith, J. A. (1976). *Autosegmental phonology*. PhD thesis, Massachusetts Institute of Technology.

Goldstein, L. (2011). Back to the past tense in English. In Gutiérrez-Bravo, R., Mikkelsen, L., and Potsdam, E., editors, *Representing language: Essays in honor of Judith Aissen*, pages 69–88.

Goldstein, L. and Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. *Phonetics and phonology in language comprehension and production: Differences and similarities*, pages 159–207.

Gordon, M. (2004). *Syllable weight*, chapter 9, pages 277–312. Cambridge University Press Cambridge.

Gorman, K. and Sproat, R. (2021). *Finite-State Text Processing*. Morgan & Claypool Publishers.

Gouskova, M. and Hall, N. (2009). Acoustics of epenthetic vowels in Lebanese Arabic. *Phonological argumentation: Essays on evidence and motivation*, pages 203–225.

Grodner, D., Gibson, E., and Watson, D. (2005). The influence of contextual contrast on syntactic processing: Evidence for strong-interaction in sentence comprehension. *Cognition*, 95(3):275–296.

Guest, O. and Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4):789–802.

Guest, O. and Martin, A. E. (2023). On logical inference over brains, behaviour, and artificial neural networks. *Computational Brain & Behavior*, pages 1–15.

Gussmann, E. (2007). *The phonology of Polish*. Oxford University Press.

Hale, M., Kissock, M., and Reiss, C. (2007). Microvariation, variation, and the features of universal grammar. *Lingua*, 117(4):645–665.

Hale, M. and Reiss, C. (2000). “Substance abuse” and “dysfunctionalism”: current trends in phonology. *Linguistic inquiry*, 31(1):157–169.

Hale, M. and Reiss, C. (2008). *The phonological enterprise*. Oxford University Press.

Hall, N. (2010). Articulatory phonology. *Language and Linguistics Compass*, 4(9):818–830.

Hall, N. (2013). Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic. *Journal of Phonetics*, 41(2):133–143.

Hall, N. (2017). Phonetic neutralization in Palestinian Arabic vowel shortening, with implications for lexical organization. *Glossa: a journal of general linguistics*, 2(1).

Halle, M. (1995). Feature geometry and feature spreading. *Linguistic inquiry*, 26(1):1–46.

Halle, M. (2003). *From memory to speech and back*. De Gruyter Mouton.

Halle, M. and Stevens, K. (1971). A note on laryngeal features. *MIT Res. Lab. Electron. Q. Prog. Rep.*, 101:198–213.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive science*, 8(3):191–219.

Harris, J. and Lindsey, G. (2014). The elements of phonological representation. In *Frontiers of phonology*, pages 34–79. Routledge.

Hayes, B. (2016). Varieties of noisy harmonic grammar. In *Proceedings of the annual meetings on phonology*.

Hayes, B. (2022). Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics*, 8(1):473–494.

Hayes, B., Kirchner, R., and Steriade, D. (2004). *Phonetically Based Phonology*. Cambridge University Press.

Hayes, B., Siptár, P., Zuraw, K., and Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, pages 822–863.

Heinz, J. (2018). The computational nature of phonological generalizations. In Hyman, L. and Plank, F., editors, *Phonological Typology*, Phonetics and Phonology, chapter 5, pages 126–195. De Gruyter Mouton.

Heinz, J. (Forthcoming). *Doing Computational Phonology*. Oxford University Press.

Herd, W., Jongman, A., and Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics*, 38(4):504–516.

Hinton, G. E. and Anderson, J. A., editors (1981). *Parallel models of associative memory: updated edition*. Erlbaum.

Hockett, C. F. and Hockett, C. D. (1960). The origin of speech. *Scientific American*, 203(3):88–97.

Hodges, W. (1993). *Model theory*. Cambridge university press.

Hodges, W. (1997). *A shorter model theory*. Cambridge university press.

Hopcroft, J. E., Motwani, R., and Ullman, J. D. (2006). *Introduction to Automata Theory, Languages, and Computation*. Pearson, third edition edition.

Hume, E. and Johnson, K., editors (2001). *The role of speech perception in phonology*. Brill.

Inkelas, S. and Shih, S. S. (2016). Re-representing phonology: consequences of q theory. In *Proceedings of NELS*, volume 46, pages 161–174.

Iverson, G. K. and Salmons, J. C. (1995). Aspiration and laryngeal representation in Germanic. *Phonology*, 12(3):369–396.

Jakobson, R., Fant, C. G., and Halle, M. (1951). *Preliminaries to speech analysis: The distinctive features and their correlates*. MIT press.

Jardine, A., Danis, N., and Iacoponi, L. (2021). A formal investigation of q-theory in comparison to autosegmental representations. *Linguistic Inquiry*, 52(2):333–358.

Jassem, W. and Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, 17(4):317–325.

Johnson, K. et al. (1997). Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, pages 145–165.

Jurafsky, D., Bell, A., and Girand (2002). The role of the lemma in form variation. In Gussenhoven, C. and Warner, N., editors, *Laboratory phonology VII*, pages 1–34.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 229–254. John Benjamins.

Kabak, B. and Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and speech*, 50(1):23–52.

Kang, Y. (2011). Loanword phonology. *The Blackwell companion to phonology*, pages 1–25.

Kaye, J. (1985). The internal structure of phonological elements: A theory of charm and government". *Phonology Yearbook*, 2.

Kazanina, N., Bowers, J. S., and Idsardi, W. (2018). Phonemes: Lexical access and beyond. *Psychonomic bulletin & review*, 25(2):560–585.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. phonetic linguistics: Essays in honor of Peter Ladefoged, ed. Victoria A. Fromkin.

Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, 5(2):275–292.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of phonetics*, 18(3):321–334.

Kenstowicz, M. (2010). The cambridge handbook of phonology. *Language*, 86(1):216–219.

Kilbourn-Ceron, O. and Goldrick, M. (2021). Variable pronunciations reveal dynamic intra-speaker variation in speech planning. *Psychonomic Bulletin & Review*, 28(4):1365–1380.

Kilpatrick, A., Kawahara, S., Bundgaard-Nielsen, R., Baker, B., and Fletcher, J. (2021). Japanese perceptual epenthesis is modulated by transitional probability. *Language and Speech*, 64(1):203–223.

Kim, H. and Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, 24(3):295–312.

Kingston, J. (2019). The interface between phonetics and phonology. *The Routledge handbook of phonetics*, pages 359–400.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language*, 70(3):419–454.

Kodner, J., Payne, S., and Heinz, J. (2023). Why linguistics will thrive in the 21st century: A reply to piantadosi (2023). https://ling.auf.net/lingbuzz/007485.

Kopkalli, H. (1994). *A phonetic and phonological analysis of final devoicing in Turkish*. PhD thesis, University of Michigan.

Kowalski, R. (1974). Predicate logic as programming language. In *IFIP congress*, volume 74, pages 569–544.

Labov, W. (2001). *Principles of linguistic change, volume 2: Cognitive and cultural factorSocial Factors. Language in Society*, volume 2. John Wiley & Sons.

Labov, W., Rosenfelder, I., and Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, pages 30–65.

Labov, W., Yaeger, M., and Steiner, R. (1972). *A quantitative study of sound change in progress*, volume 1. US Regional Survey.

LaCharité, D. and Paradis, C. (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic inquiry*, 36(2):223–258.

Lahiri, A. and Hankamer, J. (1988). The timing of geminate consonants. *Journal of Phonetics*, 16(3):327–338.

Lakatos, I. (1970). *Falsification and the Methodology of Scientific Research Programmes*, volume 4, page 91–196. Cambridge University Press.

Lambert, D., Rawski, J., and Heinz, J. (2021). Typology emerges from simplicity in representations and learning. *Journal of Language Modelling*, 9.

Leslau, W. (1969). Frequency as determinant of linguistic changes in the Ethiopian languages. *Word*, 25(1-3):180–189.

Li, Y. (2016). Complete and incomplete neutralisation in Fuzhou tone sandhi. *Tonal Aspects of Languages 2016*, pages 116–120.

Lionnet, F. (2017). A theory of subfeatural representations: the case of rounding harmony in Laal. *Phonology*, 34(3):523–564.

Lloyd, J. W. (2012). *Foundations of logic programming*. Springer.

Lohmann, A. (2018a). Cut (n) and cut (v) are not homophones: Lemma frequency affects the duration of noun–verb conversion pairs. *Journal of Linguistics*, 54(4):753–777.

Lohmann, A. (2018b). Time and thyme are not homophones: A closer look at Gahl's work on the lemma-frequency effect, including a reanalysis. *Language*, 94(2):e180–e190.

Losiewicz, B. (1995). Word frequency effects on the acoustic duration of morphemes. *Journal of the Acoustic Society of America*, 97:32–43.

Łukaszewicz, B. (2021). The dynamical landscape: phonological acquisition and the phonology–phonetics link. *Phonology*, 38(1):81–121.

Mai, A., Riès, S., Ben-Haim, S., Shih, J., and Gentner, T. (2022). Phonological contrasts are maintained despite neutralization: an intracranial eeg study. In *Proceedings of the Annual Meetings on Phonology*, volume 9.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. CUMINCAD.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198):226–228.

Martin, P. (2002). Le système vocalique du français du québec. de l'acoustique à la phonologie. *La linguistique*, 38(2):71–88.

McCarthy, J. J. (1988). Feature geometry and dependency: A review. *Phonetica*, 45(2-4):84–108.

McCarthy, J. J. (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. Equinox Publishing (Indonesia).

McCarthy, J. J. and Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. In *North East Linguistics Society*, volume 24.

McClelland, J. L. and Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, 18(1):1–86.

McClelland, J. L., Mirman, D., and Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in cognitive sciences*, 10(8):363–369.

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological science*, 2(6):387–395.

McNaughton, R. and Papert, S. A. (1971). *Counter-Free Automata*. The MIT Press.

Miller, J. and Fujimura, O. (1982). Graphic displays of combined presentations of acoustic and articulatory information. *Bell System Technical Journal*, 61(5):799–810.

Mohanan, K. P. (1986). *The theory of lexical phonology*, volume 6. Springer.

Munson, B. and Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research*, 47(5):1048–1058.

Myers, S. (2000). Boundary disputes: The distinction between phonetic and phonological sound patterns. Oxford University Press.

Nam, H. (2007a). *A gestural coupling model of syllable structure*. PhD thesis, Yale University.

Nam, H. (2007b). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In *Laboratory phonology*, volume 9, pages 483–506. Mouton de Gruyter Berlin.

Nam, H., Browman, C., Goldstein, L., Proctor, M., Rubin, P., and Saltzman, E. (2022). TADA: Task dynamic model of inter-articulator speech coordination. Computer Software.

Nam, H. and Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th international congress of phonetic sciences*, pages 2253–2256.

Nelson, S. (2022). A model theoretic perspective on phonological feature systems. *Proceedings of the Society for Computation in Linguistics*, 5(1).

Nelson, S. and Heinz, J. (in press). The blueprint model of production. *Phonology*.

Newell, A. and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological review*, 65(3):151.

Nicenboim, B., Roettger, T. B., and Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in german. *Journal of Phonetics*, 70:39–55.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In Docherty, G. J. and Ladd, D. R., editors, *Papers in Laboratory Phonology 2: Gesture, Segment, Prosody*. Cambridge University Press.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3):189–234.

Nycz, J., Blaho, S., Vicente, L., and Schoorlemmer, E. (2005). The dynamics of near-merger in accommodation. *Proc. ConSOLE XIII*, pages 273–285.

Oakden, C. (2020). Notational equivalence in tonal geometry. *Phonology*, 37(2):257–296.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech*, pages 189–216. Springer.

Ohala, J. J. (1990). There is no interface between phonology and phonetics: a personal view. *Journal of phonetics*, 18(2):153–172.

Osborn, H. A. (1966). Warao I: phonology and morphophonemics. *International Journal of American Linguistics*, 32(2):108–123.

Peng, S.-H. (2000). Lexical versus 'phonological' representations of Mandarin sandhi tones. *Papers in laboratory phonology V: Acquisition and the lexicon*, pages 152–167.

Peperkamp, S. (2004). A psycholinguistic theory of loanword adaptations. In *Annual Meeting of the Berkeley Linguistics Society*, volume 30, pages 341–352.

Peperkamp, S. and Dupoux, E. (2003). Reinterpreting loanword adaptations: the role of perception. In *Proceedings of the 15th international congress of phonetic sciences*, pages 367–370.

Peperkamp, S., Vendelin, I., and Nakamura, K. (2008). On the perceptual origin of loanword adaptations: experimental evidence from Japanese. *Phonology*, pages 129–164.

Pierce, B. C. (2002). *Types and programming languages*. MIT press.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of phonetics*, 18(3):375–394.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory phonology*, 7.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2:33–52.

Poeppel, D. and Idsardi, W. (2011). Recognizing words from speech: the perception-action-memory loop. *Lexical representation: A multidisciplinary approach*, pages 171–196.

Popper, K. (1959). *The logic of scientific discovery*. Routledge.

Port, R. and Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of phonetics*, 17:257–282.

Port, R., Mitleb, F., and O'Dell, M. (1981). Neutralization of obstruent voicing in German is incomplete. *The journal of the Acoustical Society of America*, 70(S1):S13–S13.

Port, R. F. and Leary, A. P. (2005). Against formal phonology. *Language*, 81(4):927–964.

Port, R. F. and O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of phonetics*.

Pouplier, M. (2003). The dynamics of error. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 2245–2248. Citeseer.

Pouplier, M. and Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, 33(1):47–75.

Prince, A. and Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in phonology*.

Priva, U. C. and Gleason, E. (2020). The causal structure of lenition: A case for the causal precedence of durational shortening. *Language*, 96(2):413–448.

Pulleyblank, D. (1988). Underspecification, the feature hierarchy and Tiv vowels. *Phonology*, 5(2):299–326.

Putnam, H. (1967). Psychological predicates. In Capitan, W. H. and Merill, D. D., editors, *Art, mind, and religion*, pages 37–48. University of Pittsburgh Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60(1):20–43.

Ramer, A. M. (1996). A letter from an incompletely neutral phonologist. *Journal of Phonetics*, 4(24):477–489.

Rapp, B. and Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review*, 107(3):460–499.

Rawski, J. and Baumont, L. (2023). Modern language models refute nothing. https://ling.auf.net/lingbuzz/007203.

Reiss, C. (2017). Substance free phonology. In *The Routledge handbook of phonological theory*, pages 425–452. Routledge.

Reiss, C. (2018). Substance free phonology. *The Routledge handbook of phonological theory*, pages 425–452.

Reiss, C. and Volenec, V. (2020). Conquer primal fear: Phonological features are innate and substance-free. https://ling.auf.net/lingbuzz/005683.

Roark, B. and Sproat, R. (2007). *Computational approaches to morphology and syntax*, volume 4. Oxford University Press.

Rogers, J., Heinz, J., Bailey, G., Edlefsen, M., Visscher, M., Wellcome, D., and Wibel, S. (2010). On languages piecewise testable in the strict sense. In *The Mathematics of Language: 10th and 11th Biennial Conference, MOL 10, Los Angeles, CA, USA, July 28-30, 2007, and MOL 11, Bielefeld, Germany, August 20-21, 2009, Revised Selected Papers*, pages 255–265. Springer.

Rogers, J., Heinz, J., Fero, M., Hurst, J., Lambert, D., and Wibel, S. (2013). Cognitive and sub-regular complexity. In *Formal Grammar: 17th and 18th International Conferences, FG 2012, Opole, Poland, August 2012, Revised Selected Papers, FG 2013, Düsseldorf, Germany, August 2013. Proceedings*, pages 90–108. Springer.

Roon, K. D. and Gafos, A. I. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language*, 89:222–243.

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group (1988). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. MIT press.

Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology*. PhD thesis, Massachusetts Institute of Technology.

Saltzman, E., Nam, H., Krivokapic, J., and Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of the 4th international conference on speech prosody*, pages 175–184.

Saltzman, E. L. and Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological psychology*, 1(4):333–382.

Sankoff, D. and Labov, W. (1979). On the uses of variable rules. *Language in society*, 8(2-3):189–222.

Sannella, D., Fourman, M., Peng, H., and Wadler, P. (2021). *Introduction to Computation: Haskell, Logic and Automata*. Springer, first edition edition.

Schane, S. A. (1984). The fundamentals of particle phonology. *Phonology*, 1:129–155.

Scobbie, J. M. (2005). The phonetics phonology overlap. In *QMU Speech Science Research Centre Working Papers, WP-1*.

Scobbie, J. M., Coleman, J. S., and Bird, S. (1996). Key aspects of declarative phonology. *Current Trends in Phonology*.

Shaw, J. and Tang, K. (2023). A dynamic neural field model of leaky prosody: proof of concept. In *Proceedings of the Annual Meetings on Phonology*, volume 10.

Shrager, M. (2012). Neutralization of word-final voicing in Russian. *Journal of Slavic linguistics*, pages 71–99.

Simonet, M., Rohena-Madrazo, M., and Paz, M. (2008). Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish. In *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, pages 72–86. Cascadilla Proceedings Project Somerville, Massachusets.

Sipser, M. (2013). *Introduction to the Theory of Computation*. Cengage Learning, third edition edition.

Slowiaczek, L. M. and Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of phonetics*.

Smith, C. M. (2018). *Harmony in gestural phonology*. PhD thesis, University of Southern California.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and brain sciences*, 11(1):1–23.

Smolensky, P. and Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of French liaison. *Ms. Johns Hopkins University and Northwestern University. Available as ROA*, 1286.

Solé, M.-J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and speech*, 35(1-2):29–43.

Solé, M.-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 38(1):1–23.

Solé, M.-J. (2007). Controlled and mechanical properties in speech. *Experimental approaches to phonology*, pages 302–321.

Sproat, R. and Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of phonetics*, 21(3):291–311.

Stemberger, J. P. (1985). An interactive activation model of language production. *Progress in the psychology of language*, 1.

Steriade, D. (1987). Redundant values. *CLS*, 23(2):339–362.

Steriade, D. (1993). Closure, release, and nasal contours. In Krakow, R. A. and Huffman, M. K., editors, *Nasals, nasalization, and the velum*, pages 401–470. Elsevier.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, 17(1-2):3–45.

Strother-Garcia, K. (2019). *Using model theory in phonology: a novel characterization of syllable structure and syllabification*. PhD thesis, University of Delaware.

Strother-Garcia, K., Heinz, J., and Hwangbo, H. J. (2016). Using model theory for grammatical inference: a case study from phonology. In Verwer, S., van Zaanen, M., and Smetsers, R., editors, *Proceedings of The 13th International Conference on Grammatical Inference*, volume 57 of *JMLR: Workshop and Conference Proceedings*, pages 66–78.

Strycharczuk, P. (2019). Phonetic detail and phonetic gradience in morphological processes.

Tabor, W. and Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, 23(4):491–515.

Tanner, J., Sonderegger, M., and Wagner, M. (2017). Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology*, 8(1).

Thomas, W. (1982). Classifying regular events in symbolic logic. *Journal of Computer and System Sciences*, 25(3):360–376.

Tilsen, S. (2023). Probability and randomness in phonology: Deep vs. shallow stochasticity. *Studies in Phonetics, Phonology and Morphology*, 29(2):247–269.

Todd, S., Pierrehumbert, J. B., and Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, 185:1–20.

Trakhtenbrot, B. A. (1961). Finite automata and the logic of single-place predicates. In *Doklady Akademii Nauk*, volume 140, pages 326–329. Russian Academy of Sciences.

Tuller, B., Case, P., Ding, M., and Kelso, J. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human perception and performance*, 20(1):3.

Tupper, P. F. (2015). Exemplar dynamics and sound merger in language. *SIAM Journal on Applied Mathematics*, 75(4):1469–1492.

Turvey, M. T. (1990). Coordination. *American psychologist*, 45(8):938–953.

van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua*, 118(9):1362–1374.

van Rooij, I. (2008). The tractable cognition thesis. *Cognitive science*, 32(6):939–984.

van Rooij, I. and Baggio, G. (2020). Theory development requires an epistemological sea change. *Psychological Inquiry*, 31(4):321–325.

van Rooij, I. and Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*.

van Rooij, I. and Blokpoel, M. (2020). Formalizing verbal theories. *Social Psychology*, 51(5):285–298.

van Rooy, B., Wissing, D., and Paschall, D. D. (2003). Demystifying incomplete neutralisation during final devoicing. *Southern African Linguistics and Applied Language Studies*, 21(1-2):49–66.

Vaux, B. (2008). Why the phonological component must be serial and rule-based. In Vaux, B. and Nevins, A., editors, *Rules, constraints, and phonological phenomena*. Oxford University.

Volenec, V. and Reiss, C. (2017). Cognitive phonetics: The transduction of distinctive features at the phonology-phonetics interface. *Biolinguistics*, 11:251–294.

Vu, M. H., Zehfroosh, A., Strother-Garcia, K., Sebok, M., Heinz, J., and Tanner, H. G. (2018). Statistical relational learning with unconventional string models. *Frontiers in Robotics and AI*, 5(76):1–26.

Wagner, M. (2012). Locality in phonology and production planning. *McGill working papers in linguistics*, 22(1):1–18.

Walsh, T. and Parker, F. (1983). The duration of morphemic and nonmorphemic /s/ in English. *Journal of Phonetics*, 11:201–206.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of phonetics*, 32(2):251–276.

Wedel, A., Kaplan, A., and Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2):179–186.

Wedel, A. B. (2006). Exemplar models, evolution and language change.

Westbury, J. R. and Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of linguistics*, pages 145–166.

Wheeler, M. W. (2005). *The phonology of Catalan*. Oxford University Press on Demand.

Wilson, C. (2017). *Lexical statistics determine the choice of epenthetic vowel in Japanese loanword adaptation*. PhD thesis, Yale University.

Winter, B. and Röettger, T. (2011). The nature of incomplete neutralization in German: Implications for laboratory phonology. *Grazer Linguistische Studien*, 76:55–74.

Yang, C. (2016). *The Price of Productivity*. MIT Press.

Yu, A. C. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, 24(1):187–214.

Yu, A. C. (2011). Mergers and neutralization. *The Blackwell companion to phonology*, pages 1–27.

Zee, E. (1999). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics*, 27(1):120–167.

Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. *Phonetically based phonology*, pages 157–190.

Zsiga, E. C. (1997). Features, gestures, and Igbo vowels: An approach to the phonology-phonetics interface. *Language*, pages 227–274.

Zuraw, K. (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory*, 28:417–472.

# Appendix

## Formulas for Translating from String to Coupling Graph

### Articulators

$$\begin{aligned}
\varphi^{1}_{\texttt{LIPS}} &\overset{\text{def}}{=} \text{b}(x) \lor \text{p}(x) \lor \text{m}(x) \lor \text{v}(x) \lor \text{f}(x)\\
\varphi^{2}_{\texttt{LIPS}} &\overset{\text{def}}{=} \text{b}(x) \lor \text{p}(x) \lor \text{m}(x) \lor \text{v}(x) \lor \text{f}(x)\\
\varphi^{3}_{\texttt{LIPS}} &\overset{\text{def}}{=} \text{w}(x) \lor \text{ɹ}(x)\\
\varphi^{4}_{\texttt{LIPS}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\texttt{TT}} &\overset{\text{def}}{=} \text{d}(x) \lor \text{t}(x) \lor \text{n}(x) \lor \text{ð}(x) \lor \text{θ}(x) \lor \text{z}(x) \lor \text{s}(x) \lor \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x) \lor \text{ɹ}(x) \lor \text{l}(x)\\
\varphi^{2}_{\texttt{TT}} &\overset{\text{def}}{=} \text{d}(x) \lor \text{t}(x) \lor \text{n}(x) \lor \text{ð}(x) \lor \text{θ}(x) \lor \text{z}(x) \lor \text{s}(x) \lor \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x) \lor \text{ɹ}(x) \lor \text{l}(x)\\
\varphi^{3}_{\texttt{TT}} &\overset{\text{def}}{=} \text{ɚ}(x)\\
\varphi^{4}_{\texttt{TT}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\texttt{TB}} &\overset{\text{def}}{=} \text{g}(x) \lor \text{k}(x) \lor \text{ŋ}(x) \lor \text{j}(x) \lor \text{w}(x) \lor \texttt{Vowel}(x)\\
\varphi^{2}_{\texttt{TB}} &\overset{\text{def}}{=} \text{g}(x) \lor \text{k}(x) \lor \text{ŋ}(x) \lor \text{j}(x) \lor \text{w}(x) \lor \texttt{Vowel}(x)\\
\varphi^{3}_{\texttt{TB}} &\overset{\text{def}}{=} \text{z}(x) \lor \text{s}(x) \lor \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x) \lor \text{l}(x)\\
\varphi^{4}_{\texttt{TB}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\texttt{VEL}} &\overset{\text{def}}{=} \texttt{False}\\
\varphi^{2}_{\texttt{VEL}} &\overset{\text{def}}{=} \texttt{False}\\
\varphi^{3}_{\texttt{VEL}} &\overset{\text{def}}{=} \texttt{False}\\
\varphi^{4}_{\texttt{VEL}} &\overset{\text{def}}{=} \text{m}(x) \lor \text{n}(x) \lor \text{ŋ}(x)
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\texttt{GLO}} &\overset{\text{def}}{=} \texttt{False}\\
\varphi^{2}_{\texttt{GLO}} &\overset{\text{def}}{=} \texttt{False}\\
\varphi^{3}_{\texttt{GLO}} &\overset{\text{def}}{=} \texttt{False}\\
\varphi^{4}_{\texttt{GLO}} &\overset{\text{def}}{=} \text{p}(x) \lor \text{t}(x) \lor \text{k}(x) \lor \text{f}(x) \lor \text{θ}(x) \lor \text{s}(x) \lor \text{ʃ}(x) \lor \text{tʃ}(x) \lor \text{h}(x) \lor \text{ʔ}(x)
\end{aligned}$$

## Constriction Degree

$$\begin{aligned}
\varphi^{1}_{\mathtt{clo}} &\overset{\text{def}}{=} \text{b}(x) \lor \text{p}(x) \lor \text{m}(x) \lor \text{d}(x) \lor \text{t}(x) \lor \text{n}(x) \lor \text{g}(x) \lor \text{k}(x) \lor \text{ŋ}(x)\\
\varphi^{2}_{\mathtt{clo}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{3}_{\mathtt{clo}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{4}_{\mathtt{clo}} &\overset{\text{def}}{=} \text{ʔ}(x)
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{crit}} &\overset{\text{def}}{=} \text{v}(x) \lor \text{f}(x) \lor \text{ð}(x) \lor \text{t}(x) \lor \text{z}(x) \lor \text{s}(x) \lor \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x)\\
\varphi^{2}_{\mathtt{crit}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{3}_{\mathtt{crit}} &\overset{\text{def}}{=} \text{z}(x) \lor \text{s}(x) \lor \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x) \lor \text{l}(x)\\
\varphi^{4}_{\mathtt{crit}} &\overset{\text{def}}{=} \mathtt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{nar}} &\overset{\text{def}}{=} \text{j}(x) \lor \text{w}(x) \lor \text{ɹ}(x) \lor \text{l}(x)\\
\varphi^{2}_{\mathtt{nar}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{3}_{\mathtt{nar}} &\overset{\text{def}}{=} \text{ʒ}(x) \lor \text{ʃ}(x) \lor \text{dʒ}(x) \lor \text{tʃ}(x) \lor \text{l}(x) \lor \text{w}(x) \lor \text{ɹ}(x) \lor \text{l}(x) \lor \text{ɔ}(x) \lor \text{u}(x) \lor \text{ʊ}(x) \lor \text{oʊ}(x)\\
\varphi^{4}_{\mathtt{nar}} &\overset{\text{def}}{=} \mathtt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{wide}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{2}_{\mathtt{wide}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{3}_{\mathtt{wide}} &\overset{\text{def}}{=} \text{z}(x) \lor \text{s}(x)\\
\varphi^{4}_{\mathtt{wide}} &\overset{\text{def}}{=} \text{p}(x) \lor \text{t}(x) \lor \text{k}(x) \lor \text{f}(x) \lor \text{θ}(x) \lor \text{s}(x) \lor \text{ʃ}(x) \lor \text{tʃ}(x) \lor \text{h}(x) \lor \text{m}(x) \lor \text{n}(x) \lor \text{ŋ}(x)
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{V}} &\overset{\text{def}}{=} \text{i}(x) \lor \text{ɪ}(x) \lor \text{eɪ}(x) \lor \text{ɛ}(x) \lor \text{æ}(x) \lor \text{u}(x) \lor \text{ʊ}(x) \lor \text{oʊ}(x) \lor \text{ɔ}(x) \lor \text{ɑ}(x) \lor \text{ʌ}(x) \lor \text{ə}(x) \lor \text{ɚ}(x)\\
\varphi^{2}_{\mathtt{V}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{3}_{\mathtt{V}} &\overset{\text{def}}{=} \text{i}(x)\\
\varphi^{4}_{\mathtt{V}} &\overset{\text{def}}{=} \mathtt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{rel}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{2}_{\mathtt{rel}} &\overset{\text{def}}{=} \mathtt{True}\\
\varphi^{3}_{\mathtt{rel}} &\overset{\text{def}}{=} \mathtt{False}\\
\varphi^{4}_{\mathtt{rel}} &\overset{\text{def}}{=} \mathtt{False}
\end{aligned}$$

## Constriction Location

$$\begin{aligned}
\varphi^{1}_{\mathtt{pro}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{2}_{\mathtt{pro}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{3}_{\mathtt{pro}} &\overset{\text{def}}{=} \text{u}(x) \vee \text{ʊ}(x) \vee \text{oʊ}(x) \vee \text{ɔ}(x) \\
\varphi^{4}_{\mathtt{pro}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{dent}} &\overset{\text{def}}{=} \text{v}(x) \vee \text{f}(x) \vee \text{ð}(x) \vee \text{θ}(x) \\
\varphi^{2}_{\mathtt{dent}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{3}_{\mathtt{dent}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{4}_{\mathtt{dent}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{alv}} &\overset{\text{def}}{=} \text{d}(x) \vee \text{t}(x) \vee \text{n}(x) \vee \text{z}(x) \vee \text{s}(x) \vee \text{l}(x) \\
\varphi^{2}_{\mathtt{alv}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{3}_{\mathtt{alv}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{4}_{\mathtt{alv}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{palv}} &\overset{\text{def}}{=} \text{ʒ}(x) \vee \text{ʃ}(x) \vee \text{dʒ}(x) \vee \text{tʃ}(x) \\
\varphi^{2}_{\mathtt{palv}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{3}_{\mathtt{palv}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{4}_{\mathtt{palv}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{pal}} &\overset{\text{def}}{=} \text{j}(x) \vee \text{ɹ}(x) \vee \text{i}(x) \vee \text{ɪ}(x) \vee \text{eɪ}(x) \vee \text{ɛ}(x) \vee \text{ɚ}(x) \\
\varphi^{2}_{\mathtt{pal}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{3}_{\mathtt{pal}} &\overset{\text{def}}{=} \text{ʒ}(x) \vee \text{ʃ}(x) \vee \text{dʒ}(x) \vee \text{tʃ}(x) \\
\varphi^{4}_{\mathtt{pal}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{vel}} &\overset{\text{def}}{=} \text{g}(x) \vee \text{k}(x) \vee \text{ŋ}(x) \\
\varphi^{2}_{\mathtt{vel}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{3}_{\mathtt{vel}} &\overset{\text{def}}{=} \text{z}(x) \vee \text{s}(x) \\
\varphi^{4}_{\mathtt{vel}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\begin{aligned}
\varphi^{1}_{\mathtt{uval}} &\overset{\text{def}}{=} \text{w}(x) \vee \text{u}(x) \vee \text{ʊ}(x) \vee \text{ə}(x) \\
\varphi^{2}_{\mathtt{uval}} &\overset{\text{def}}{=} \texttt{False} \\
\varphi^{3}_{\mathtt{uval}} &\overset{\text{def}}{=} \text{l}(x) \\
\varphi^{4}_{\mathtt{uval}} &\overset{\text{def}}{=} \texttt{False}
\end{aligned}$$

$$\varphi^1_{\texttt{uvuphar}} \overset{\text{def}}{=} \text{oʊ}(x)$$

$$\varphi^2_{\texttt{uvuphar}} \overset{\text{def}}{=} \texttt{False}$$

$$\varphi^3_{\texttt{uvuphar}} \overset{\text{def}}{=} \texttt{False}$$

$$\varphi^4_{\texttt{uvuphar}} \overset{\text{def}}{=} \texttt{False}$$

---

$$\varphi^1_{\texttt{phar}} \overset{\text{def}}{=} \text{æ}(x) \vee \text{ɑ}(x) \vee \text{ɔ}(x)$$

$$\varphi^2_{\texttt{phar}} \overset{\text{def}}{=} \texttt{False}$$

$$\varphi^3_{\texttt{phar}} \overset{\text{def}}{=} \texttt{False}$$

$$\varphi^4_{\texttt{phar}} \overset{\text{def}}{=} \texttt{False}$$

# Formulas for Translating from Coupling Graph to String

## Consonants

$$
\begin{aligned}
\varphi_{\text{b}}(x) &\overset{\text{def}}{=} \texttt{LIPS}(x) \wedge \texttt{clo}(x) \wedge \neg\exists y[x \Diamond y \wedge (\texttt{GLO}(y) \vee \texttt{VEL}(y))] \\
\varphi_{\text{p}}(x) &\overset{\text{def}}{=} \texttt{LIPS}(x) \wedge \texttt{clo}(x) \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{m}}(x) &\overset{\text{def}}{=} \texttt{LIPS}(x) \wedge \exists y[x \Diamond y \wedge \texttt{VEL}(y)]
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{\text{d}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{alv}(x) \wedge \texttt{clo}(x) \wedge \neg\exists y[x \Diamond y \wedge (\texttt{GLO}(y) \vee \texttt{VEL}(y))] \\
\varphi_{\text{t}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{alv}(x) \wedge \texttt{clo}(x) \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{n}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \exists y[x \Diamond y \wedge \texttt{VEL}(y)]
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{\text{g}}(x) &\overset{\text{def}}{=} \texttt{TB}(x) \wedge \texttt{vel}(x) \wedge \texttt{clo}(x) \wedge \neg\exists y[x \Diamond y \wedge (\texttt{GLO}(y) \vee \texttt{VEL}(y))] \\
\varphi_{\text{k}}(x) &\overset{\text{def}}{=} \texttt{TB}(x) \wedge \texttt{vel}(x) \wedge \texttt{clo}(x) \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{ŋ}}(x) &\overset{\text{def}}{=} \texttt{TB}(x) \wedge \exists y[x \Diamond y \wedge \texttt{VEL}(y)]
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{\text{v}}(x) &\overset{\text{def}}{=} \texttt{LIPS}(x) \wedge \texttt{dent}(x) \wedge \neg\exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{f}}(x) &\overset{\text{def}}{=} \texttt{LIPS}(x) \wedge \texttt{dent}(x) \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{ð}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{dent}(x) \wedge \neg\exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{θ}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{dent}(x) \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{z}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{alv}(x) \wedge \texttt{crit} \wedge \neg\exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{s}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{alv}(x) \wedge \texttt{crit} \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{ʒ}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{palv}(x) \wedge \texttt{crit} \wedge \neg\exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{ʃ}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{palv}(x) \wedge \texttt{crit} \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)]
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{\text{dʒ}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{palv}(x) \wedge \texttt{clo} \wedge \neg\exists y[x \Diamond y \wedge \texttt{GLO}(y)] \\
\varphi_{\text{tʃ}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{palv}(x) \wedge \texttt{clo} \wedge \exists y[x \Diamond y \wedge \texttt{GLO}(y)]
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{\text{j}}(x) &\overset{\text{def}}{=} \texttt{TB}(x) \wedge \texttt{pal}(x) \wedge \texttt{nar}(x) \\
\varphi_{\text{w}}(x) &\overset{\text{def}}{=} \texttt{TB}(x) \wedge \texttt{uvul}(x) \wedge \texttt{nar}(x)
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{\text{ɹ}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{pal}(x) \wedge \texttt{nar}(x) \\
\varphi_{\text{l}}(x) &\overset{\text{def}}{=} \texttt{TT}(x) \wedge \texttt{alv}(x) \wedge \texttt{nar}(x)
\end{aligned}
$$

$$
\begin{aligned}
\varphi_{\text{h}}(x) &\overset{\text{def}}{=} \texttt{GLO}(x) \wedge \texttt{wide}(x) \\
\varphi_{\text{ʔ}}(x) &\overset{\text{def}}{=} \texttt{GLO}(x) \wedge \texttt{clo}(x)
\end{aligned}
$$

## Vowels

A closer inspection of the TADA manual shows that the “V” parameter, which corresponds with constriction degree, has different continuous values for each vowel. They are shown in the table below. This can be accounted for by assuming labeling relations like $\texttt{V}_\texttt{n}$ where the specific parameter is encoded. While the space of *possible* $n$’s is continuous, the space of *used* $n$’s is finite meaning there are a fixed set of possible `V` values for English vowels.

| *Vowel* | V | Location | Secondary Art |
|---|---|---|---|
| [i] | 5 | PAL | LIPS-V |
| [ɪ] | 8 | PAL | – |
| [eɪ] | 8.5 | PAL | – |
| [ɛ] | 11.5 | PAL | – |
| [æ] | 17 | PHAR | – |
| [ɑ] | 11 | PHAR | – |
| [ɔ] | 11 | PHAR | LIPS-PRO |
| [u] | 4 | UVUL | LIPS-PRO |
| [ʊ] | 6 | UVUL | LIPS-PRO |
| [oʊ] | 5 | UVUPHAR | LIPS-PRO |
| [ʌ] | 6.5 | UVUPHAR | – |
| [ə] | 8.5 | UVUL | – |
| [ɚ] | 11 | PHAR | TT |