Ghost in the Cell:

Phonological Primes as Neural Symbols

Sayantan Mandal

A Thesis
In the Department of
Classics, Modern Languages and Linguistics

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Individualized Program)

at Concordia University
Montréal, Québec, Canada

August 2024

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Sayantan Mandal

Entitled: Ghost in the Cell: Phonological Primes as Neural Symbols

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Individualized Program)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Alan Bale	Chair
Dr. Heather Newell	External Examiner
Dr. Ida Toivonen	Arm's Length Examiner
Dr. Tobias Scheer	Examiner
Dr. Sigwan Thivierge	Examiner
Dr. Charles Reiss	Thesis Supervisor (s)

Approved by

Dr. Felice Yuen Chair of Department or Graduate Program Director

08.07.2024
Date of Defence

Dr. Effrosyni Diamantoudi Dean, of Faculty

Abstract for PhD

Ghost in the Cell: Phonological Primes as Neural Symbols

Sayantan Mandal, PhD

Concordia University, 2024

This dissertation investigates the role of abstractions in linguistic knowledge, particularly focusing on phonological representations and its interface with phonetics. It explores how abstract phonological grammar, which determines the "sound patterns" of languages, impacts speech detection, processing, and production. The work consists of a collection of independent articles, with some written specifically for this dissertation, centered on understanding the phonology-phonetics interface from a substance-free perspective.

Chapter 1 sets the historical context for substance free phonology, tracing its development from the cognitive revolution of the 1950s and the influence of Noam Chomsky's work, culminating in the biolinguistic program.

Chapter 2 reviews the foundations of substance-free phonology, arguing that its philosophical basis stems from the strong minimalist thesis and emphasizes reducing phonological content in Universal Grammar (UG). It lays the groundwork for constructing a model of the phonology-phonetics interface, emphasizing a partially veridical relationship between phonological form and phonetic substance.

Chapter 3 examines the implications of a substance free approach for the phonology-phonetics interface, discussing the challenges of a deterministic relationship between phonological form and phonetic realization. It proposes a hybrid framework that avoids both absolute determinism and complete arbitrariness, supported by neurobiological models of speech processing.

Chapter 4 presents an empirical study using dichotic listening to investigate how bilinguals manage phonological features in speech perception. The study finds that perceptual intrusions from an unattended language are influenced by voicing, supporting a feature specification approach based on privative features and underspecification.

Chapter 5 continues the behavioral investigation with experiments on velar palatalization in Malayalam. It reports significant vowel and suffix effects on palatalization, contributing to understanding the phonological and phonetic division of labor. An interface account, couched in the hybrid framework developed in Chapter 3, is provided.

Chapter 6 discusses neurobiological research on phonological representations, including a project examining event-related potentials (ERP) to differentiate phonological from other linguistic alternations. It also proposes future research on contrasting phonological and phonetic alternations in Hindi and French.

Chapter 7 concludes by reflecting on the dissertation's contributions, emphasizing the unifying theme of abstract substance-free neural symbols in phonological representations, and addressing feedback received throughout the research.

Acknowledgements

I began working on this dissertation in the winter of 2019, and since then I have been fortunate enough to consult and work with so many wonderful colleagues, teachers and friends that now writing an acknowledgment seems like the most daunting task. I am afraid I might leave someone out, and that I might not say enough. So, in advance, I leave my apologies for any oversight and the brevity of this section.

First and foremost, I am grateful for the companionship I have received from my primary supervisor, Charles Reiss. Believe it or not, I first found Charles' works when, not unlike Newton's proverbial apple, an edited volume of phonological works quite literally fell on my head. It is often said that Newton's apple is just a fanciful analogy, and it never quite dropped on his head. I assure you "my" book, and my head, was quite real. It was rather painful too, as it was a hardcover edition of *Phonological Knowledge*, edited by Noel Burton-Roberts, Philip Carr, and Gerard Docherty. I was browsing through a lower shelf of the library at the MARCS Institute of Brain, Behavior and Development in the fall of 2015, annoyed at the lack of internalist approaches in most of the works on language that I could get my hands on, when this rather happy, albeit painful, event occurred. Inside this volume I came across the chapter titled *Phonology as Cognition*, authored by Charles Reiss and Mark Hale. If I say this chapter was everything I have been looking for it would be both a lie, for I was not sure what I was looking for, and also an understatement. Starting there, I read everything I could find that had Charles' name on it, although it would be another four years before I finally wrote to him. This time the circumstances were not so happy, as I had just dropped out of a PhD program. But here I am, writing an acknowledgment to a doctoral dissertation, and I could not begin to thank Charles enough for giving me a second chance... for listening to me... for believing in me... especially in those times when I disagreed with him. This dissertation is every bit a work along the lines of Hale and Reiss, and where I have departed from their thought, I have sought only to pull the same to directions that I earnestly believe to be beneficial.

I am grateful, also, to Prof. Tobias Scheer, my second supervisor. My current work is almost entirely his brainchild, and I am immensely grateful for the countless hours he has spent talking to me. My first contact with Tobias was no less eventful, either. I had just had to abandon two successive experiments – a physiological experiment on the effects of sign-language articulation on breathing, and a fMRI study on phonological features, inundated by the COVID-19 pandemic and frustrated by the lack of relevant technical expertise in Concordia – when, perhaps in desperation, I reached out to Tobias. I had been vaguely familiar with his project, amusingly termed phonolEEGy, and I was ecstatic to learn that it was both underway, and I would be welcome to join. I cannot stress this enough – my primary, and perhaps *only*, interest in linguistic theory lies in the Chomskyan assertion that it tells us something about how our brains are wired, about what kind of creatures we are. Unlike many of my colleagues, who find joy and reward in languages themselves (and rightfully so!), I unfortunately care very little for languages. One is as good, or bad, as the other, and to quote Norbert Hornstein, they are all rather uninteresting. But they do say something about the Sapien brain, about why only us, and I will be eternally grateful to Tobias Scheer for giving me an avenue to pursue the issues concerned. It has given me a dissertation to write, but more importantly it has given me a vocation for life. A debt, I am afraid, I will never repay.

Speaking of Chomskyan assertions, I am indebted to Noam Chomsky for the many emails we have exchanged between 2016 and 2022. I would never have gone to, and through, graduate school if it wasn't for you. Your works have been, to large extent, the only consistent thing I have been pre-occupied with for most of my young-adult and adult life. I have written to you, sometimes needlessly, and you have always responded and every time with the utmost thoughtfulness and care. Every response I received have filled me, at once, with overwhelming joy and a sense of despair. Joy because your advice has always helped me look to the next day, in the most desparate of times. Despair at the thought of my ineptitude taking up your valuable time. I have learned too much, about too many things from you... but perhaps most importantly, you have taught me to never look away, and never forget the vulgar disparity of life around us.

Tom Bever and Anne Cutler, perhaps second only to Noam Chomsky, have been a constant source of encouragment and inspiration. I cannot count the number of emails we have exchanged, and every exchange has made me feel a little bit more competent. I do not how to thank Tom and Anne more, as teachers or as friends. But thank you both I shall, for being there every time I needed you.

Over the course of this work, I have had the opportunity to meet and learn from some of the brightest minds in the field. I will be forever grateful to Prof. William James Idsardi, Prof. Phillip J. Monahan and Prof. David Poeppel for answering my somewhat naive questions about neuroscience and language with feedback that were as decisive as they were easy to comprehend. Your works have inspired almost the entirely of this dissertation.

Norbert Hornstein deserves a special mention here. One of my earliest linguistic heroes, Norbert Hornstein took the time to listen to me and guide me while I was in the process of switching graduate programs. You were patient and understanding, your encouragment helped me get through some rather testing times.

One of the fondest experiences during my time at Concordia involves the two North American Phonology Conferences that we organized. These events were to me like drops of fresh water in the dessert, and I have lost count of the number of people I have met here who have inspired me to look beyond my own interests. Harry van der Hulst is one of the first names that springs to mind. One of the nicest person I have ever met, the conversation we had in NAPhCxii has been instrumental in shaping my understanding of phonological theory. Heather Newell is a close second. I still owe you a brain-scan, Heather!

Gratitude is also due to Chris Golston, Iris Berent, Tim O'Donnel, Brandon Gillon, Greg Hickok, David Corina, Laurel Lawyer, Jason Shaw, Fabian Mathy, Friederika Moltmann, Angela Friederici, Mathias Scharinger, Maria Polinsky, David Adger, David Medeiros and Massimo Piatelli-Palmarini. At one point or another, you have all been kind enough to discuss my work with me. Without your help, I would not be here today.

I will be remiss if I did not mention Rim Dabbous and Alexander Marx Chabot, my dearest colleagues. Your company have been no less influential, and indeed in some sense more so, than any of the names mentioned before. I am indebted to you for your friendship, and the immense reservoir of patience and understanding you have been to me. Without your help, I am afraid, I would have given up a second time. My love and gratitude is also owed to Armel Jacques Jolin... a constant friend and ever the source of encouragement. The time we spent together in Montreal were among the best, and you continue to inspire me to this day.

I am grateful to the University of Concordia for supporting me with two concurrent scholarships during my tenure as a PhD student. Both of which has been instrumental in keeping me afloat. I am grateful also to MITACS, and the Government of Canada, for endowing me with a MITACS Globalink fellowship that has been instrumental in helping me carry out the Mind 2 Brain project in France. Finally, my eternal gratitude to the colleagues and friends at Université Côte d'Azur. Every single one of you have made my stay here a bliss... and I am eternally grateful that I got to write my dissertation at the beautiful French Riviera.

We all save the best for the end, and I could not possibly end this without thanking my family. My parents who spent their lives' savings so I could engage in abstractions, to Alice and Carol who made me believe that I was worthwhile. And lastly, to my didu, dadabhai and mama...

wish you were here. I miss you. And to Shikhamasi... who cooked me food every single time I was home. Without you all, I would not have survived to see this day!

Sayantan Mandal
February 26, 2024.
Côte d'Azur, France.

The common slogan that study of mind is neuroscience at an abstract level might turn out to be just as misleading as comparable statements about chemistry and physics ninety years ago. Unification may take place, but that might require radical rethinking of the neurosciences, perhaps guided by computational theories of cognitive processes, as Gallistel and King suggest.

Noam Chomsky; Science, Mind, and Limits of Understanding

So I'm increasingly concerned when I find that yet another Tuesday has come around, and the Times is still interested in the neural localisation of mental functions and I still can't figure out why. It occurs to me that maybe we're heavily invested in finding answers to which we don't know the corresponding questions.

Jerry Fodor; Why The Brain

Perhaps we see now a glimmer of unification among the notions that human symbolic representations have both an emotional and a computational component. What we may be working towards is a theory of the evolution of human expressions in general.

Thomas G. Bever; Noam's Ark

One of the troubling things about studying language is that almost everyone is willing to contribute an opinion on how language works, including scientists from other fields who should know better. This is not typically true when one studies, say, the kidney or motor control. Language is viewed as an area that licenses unconstrained speculation. Despite all lay intuitions, however, the scientific study of language requires technical expertise.

David Poeppel; Mind Over Chatter

"To my parents, who gave me life..
to my didu, who taught me to love...
to my dadabhai, who gave me curiosity...

& to Noam Chomsky, the wind beneath my wings.
"old father, old artificer, stand me now and ever in good stead".

Contribution of Authors

One of the chapters in this dissertation, specifically Chapter 4, is a journal article published in *Glossa: A Journal of General Linguistics* (Mandal et al., 2020). This chapter presents an empirical study and was co-authored by myself (first-author), and Prof. Catherine T. Best, Prof. Anne Cutler and Prof. Jason Shaw. The study was designed by me, with help from my co-authors, and the text itself is primary written by myself. The co-authors were my supervisors during my post-graduate study in Australia, and provided valuable inputs in refining the experimental design and statistical analyses. Prof. Catherine T. Best was particularly helpful in creating the experimental design, while Prof. Jason Shaw provided valuable feedback in constructing a theoretical framework to couch the experimental findings. I was primarily responsible for writing the published article, and the content has been presented in its original form. *Glossa* allows authors full rights and priviledges to use their work as part of their dissertations provided the original publication is cited and acknowledged. All authors reviewed the final manuscript and approved of the contents.

Contents

List of Figures **xix**

List of Tables **xxiii**

List of Abbreviations **xxv**

1 Ever since Chomsky **1**

1.1 Introduction 1
1.2 Background 2
1.3 The cognitive (counter-)revolution 3
1.4 New beginnings 4
1.5 Computation and cognition 7
1.6 Impact and aftermath 11
1.7 Conclusion 17

2 What phonology is: a twice-told tale **19**

2.1 Introduction 19
2.2 Phonology: the early years 26
  2.2.1 Sound patterns of English 30
  2.2.2 Substance abuse in generative grammar 35
  2.2.3 Substance-free phonology 38
2.3 The phonological grounding of phonetics 45

2.3.1 Things fall apart 51
2.3.2 The lower-interface: where we are (really) at, and where we can (realistically) go... 66
2.4 Conclusion 73

3 Partial Veridicality at the Lower Interface 75

3.1 Introduction 75
3.2 Modularity in generative grammar 77
3.2.1 Primes in substance-free phonology 78
3.2.2 Substance in substance-free phonology 84
3.3 The lower interface 93
3.3.1 Multiplexing 93
3.3.2 Analysis-by-Synthesis 95
3.4 Partial-veridicality 98
3.5 Conclusion 104

4 At The Lower Interface-I: Dichotic Listening in Bilingual Phonology 107

4.1 Introduction 107
4.2 English and Malayalam target consonants 109
4.3 Experimental Design 114
4.3.1 Stimuli 114
4.3.2 Participants 117
4.3.3 Procedure 118
4.3.4 Results 120
4.4 Discussion 123
4.5 Conclusion 129
4.6 Special acknowledgment 130

5 At the Lower Interface-II: Velar Palatalization 133

5.1 Introduction . . . 133

5.1.1 Language and linguistic theory: the "sound" domain . . . 136

5.1.2 I-Language . . . 138

5.1.3 Malayalam . . . 142

5.2 Psycholinguistic approaches to palatalization . . . 146

5.2.1 Perception . . . 147

5.2.2 Production . . . 148

5.2.3 Acquisition . . . 149

5.2.4 Form and substance in palatalization . . . 152

5.2.5 Two behavioral investigations . . . 154

5.3 Experiment I: Two forced-choice tasks . . . 157

5.3.1 Participants . . . 160

5.3.2 Stimuli . . . 161

5.3.3 Procedure . . . 163

5.3.4 Results . . . 165

5.3.5 Experiment II: Two modified WUG-tests . . . 170

5.3.6 Participants . . . 173

5.3.7 Stimuli . . . 173

5.3.8 Procedure . . . 174

5.3.9 Results . . . 175

5.4 Discussion . . . 179

5.4.1 Null-Hypothesis: modularity . . . 181

5.4.2 Grammar cannot be biased . . . 191

5.4.3 A phonological account . . . 194

5.4.4 The Lower interface: explaining gradience and variation . . . 205

5.5 Conclusion . . . 214

5.6 Special acknowledgment 215

6 What phonology is NOT **217**

6.1 Introduction 217

6.1.1 The issue of types of alternations/computations 218

6.2 Neurobiology of language 222

6.2.1 Correlational behavioral neuroscience 225

6.2.2 Towards a new neurobiology of language 226

6.3 Plastic and stone: neurobiological disambiguation of what phonology is not . . . 229

6.3.1 PhonolEEGical alternations 233

6.3.2 Mind 2 Brain: disambiguating phonology from analogy 237

Stimuli 239

Procedure 240

Analyses 242

Expected outcomes 243

6.3.3 What phonology is not 245

Experimental procedure 247

6.4 Final remarks 250

6.5 Special acknowledgment 252

7 Concluding Remarks **255**

7.1 Ghost in the Cell 255

7.2 Competence, performance, and relating verbal behavior to phonological features 257

7.3 On the ghosts 259

7.4 On the cells 261

7.5 Parting remarks 262

Bibliography **263**

List of Figures

1.1 The cognitive sciences 11

2.1 Temporal co-ordination of muscle activity. 48
2.2 Wetware for features 53
2.3 Phonological Feature 54
2.4 Transduction via CP 57
2.5 Transduction from speech signal to phonological representations 69

3.1 The upper-interface. From Scheer, 2010a 88
3.2 Arbitrary computations, deterministic transductions 89
3.3 Upper and lower interfaces of phonology. From Scheer, 2010a 92
3.4 Features as Neural Entrainments 95
3.5 Temporally synchronized features in the signal. 98
3.6 From vibrations in the ears to abstractions in the brain 100
3.7 Post-phonological lexicon 103

4.1 Distribution of stops and fricatives in English and Malayalam. Gray cells in a given language indicate phonological gaps in that language 110
4.2 Privative feature matrix contrasting the stimuli-initial consonants. 113
4.3 The dichotic nonce stimuli in each language, in which the stress-initial consonants are the target and distractor items in each respective language-attend condition. 115

4.4 Mean HNR scores in dB for English and Malayalam stimulus consonants. Standard deviations are provided in parentheses. 118
4.5 Examples of response-choice wheels. 120
4.6 Mean percent correct responses for English and Malayalam.. 121
4.7 Intrusions from unattended language/ear by language, ear and target consonant type. 122

5.1 Malayalam velar paltalization rule revised for single-melody velars 145
5.2 From "Learning Phonology with Substantive Bias", by Colin Wilson (Wilson, 2006), pp. 949 148
5.3 Test Block. From "Learning Phonology with Substantive Bias", by Colin Wilson (Wilson 2006), pp. 965 151
5.4 Velar From "Malayalam: a Grammatical Sketch" by Haoway Jiang (Jiang, 2010) . 155
5.5 Experiment 1: Proportion of Palatalized Responses 167
5.6 Experiment 2: Proportion of Palatalized Responses 169
5.7 Experiment 2A: Proportion of Palatalized Responses 176
5.8 Experiment 2A: Proportion of Palatalized Responses 178
5.9 Biolinguistic typology 185
5.10 Modularity of Input Systems. From Scheer, 2010a. 190
5.11 Malayalam velar palatalization rule. From Mohanan and Mohanan, 1984 197
5.12 Nasal Place Assimilation as DAX Spreading. From Halle (2005) 200
5.13 Feature tree for [−Back] vowels 201
5.14 Malayalam velar palatalization after[i] 201
5.15 Malayalam velar palatalization after [e] 202
5.16 Malayalam velar palatalization after [a] 202
5.17 Prespecification of /k/ 204
5.18 Transductive interfaces 206

5.19 Upper and lower interfaces of phonology. From Scheer, 2010a 208
5.20 Processin steps in AxS model. 210
5.21 Post-phonological lexicon and spell-out 212

6.1 ERP entrainment for English velar-softening results. From Chabot, 2021, p. 281 . 236
6.2 ERP experiment timeline for Mind 2 Brain 242

List of Tables

1.1 Marr's (1982) Tri-level hypothesis 13

2.1 Muscle activity assignment schema. 47
2.2 Lenneberg's neuromsucular schema. 49

3.1 Modules and vocabularies 87

5.1 Surface place contrasts in Malayalam 143
5.2 Distribution of Velar and Palatals 144
5.3 Exceptions in Velar Palatalization 145
5.4 Examples of experimental stimuli used in exp.1A 162
5.5 Examples of experimental stimuli used in exp.1B 163
5.6 Module-Specific Vocabulary 205

6.1 Conditions in Sahin et al. (2009) 234
6.2 List of -aux] type words 240
6.3 Conditions in Scheer et al. forthcoming 245
6.4 Phonetics and phonology in Hindi and French 247
6.5 Proposed word-blending task 249

List of Abbreviations

ANN	**A**rtificial **N**eural **N**etworks
C-I	**C**onceptual **I**ntentional systems
C-R	**C**omputational **R**epresentational
CTM	**C**omputational **T**heory of **M**ind
CP	**C**ognitive **P**honetics
EEG	**E**lectro**e**ncephalo**g**raphy
ERP	**E**vent-**r**elated**P**otentials
FL	**F**aculty of **L**anguage
ICA	**I**ndependent **C**omponents **A**nalyses
IPS	**I**nformation **P**rocessing **S**ystem
LP	**L**ogical **P**honology
LF	**L**ogical **F**orm
NL	**N**atural **L**anguage
PF	**P**honological **F**orm
SPE	**S**ound **P**atterns of **E**nglish
SFP	**S**ubstance **F**ree **P**honology
STRF	**S**pectro **T**emporal **R**eceptive **F**ields
S-M	**S**ensory-**M**otor systems
UG	**U**niversal **G**rammar

Chapter 1

Ever since Chomsky

1.1 Introduction

This dissertation is an exercise in (re)applying to the phonological module of Universal Grammar (UG) some of the core insights that led to, and lay at the heart of, what has come to be called *The Cognitive Revolution* (Miller, 2003; Mandal, 2021). This revolution, by all accounts still ongoing (Piattelli-Palmarini and Uriagereka, 2008), was marked by a series of events and development that sought to liberate psychology from the dominant behaviorist dogma prevalent in the 1950s, and re-orient it to what William James (James, 1890) had envisioned it to be – a study of the mentalistic events and principles that guides observable behavior. By all accounts (Miller, 2003) the onset of the revolution was marked by Noam Chomsky's scathing review of B.F. Skinner's *Verbal Behavior* (Skinner, 1957), and perpetuated by his continuing assault on the behavioral-taxonomic methodology (Chomsky, 1953; Chomsky and Schützenberger, 1959; Chomsky, 1959; Chomsky and Halle, 1965). I published this first chapter, as part of The Encyclopedia of Evolutionary Psychological Sciences (Mandal, 2021), and here it serves to initiate the reader to a series of paradigm shifting events to which all of the Cognitive Sciences, in my opinion, trace their roots back to, ever since Chomsky, 1959.

1.2 Background

In the early twentieth century, psychology had wandered a long way from being the "study of mind" that William James (James, 1890; James, 2007[1890]) had envisioned it to be. Quite the contrary, psychologists had all but given up on issues concerning the mind and the mental, focusing instead on behavior as responses to physical stimuli. The behaviorists argued that mental events, such as beliefs and representations, were not publicly observable. Since my internal beliefs, say "I like red cars", are not objectively available to others for observation, independently of my introspective recollections, behaviorists argued that such concepts cannot be studied scientifically. Behavior, on the other hand, is very publicly observable and recordable. It is possible, for instance, to observe me repeatedly buying cars which are red and a scribe my decision to do so to similarly observable stimuli that reinforces my behavior, such as approving attitude of my friends and neighbors. As such recording observable behavior, documenting what factors in the environment (stimuli) correspond to them, and converting them to a behaviorist jargon were asserted by behaviorists to be the only means to bringing about scientific objectivity to the study of what living organisms do and why (Miller 2003). Primarily an American phenomena, behaviorism had an over whelming influence on experimental psychology. For many behaviorists, consequently, there ceased to be any meaningful dichotomy between perception and discrimination, memory and learning, etc. This,however, had the negative side effect of erasing any distinction between describing a phenomena(e.g., If dropped, a ball falls downward.) and causally explaining the precise mechanisms underlying said phenomena (e.g., the effects, and origins, of gravitational pull). In 1951, George Miller published *Language and Communication* (Miller, 1951) a study of language and linguistic phenomena, acknowledging the well-established behaviorist bias of the time in his preface. Miller's work, however, was much less radical in its behaviorism compared to B.F. Skinner's soon-to-follow Verbal Behavior (Skinner, 1957), which practically reduced language to operant-conditioned verbal behavior. It was Skinner's book that sparked what has become, possibly,the most well-known criticism of behaviorism and one of the major manifestos of the cognitive revolution in the form of Noam Chomsky's scathing review of Skinner (Chomsky, 1959).

1.3 The cognitive (counter-)revolution

Behaviorism in psychology was a very influential movement in North America, dominating almost every aspect of the discipline and turning it into "a study of behavior". The behaviorist shad a narrow aim of predicting and controlling behavior, and in the case of psychologists like J.B. Watson, the latter took center field and equated all psychology with applied psychology. The motto of the methodological behaviorist was to deny all discussions of internal mental states or belief systems and focus simply on documenting observable behavior in order to isolate specific stimuli that could be used to either re-elicit, suppress, or modify said behavior in some shape or form. Watson's insistence on using a stimulus-response approach resonated well with radical empiricists because of its reliance on the associative basis of all behaviors.Allegedly, every behavior could be associated with some stimuli that could reinforce its learning. While the notion of "observable" was central to behaviorists in declaring their methodology as "science", the history of science itself is rather sparse in endorsing such definitions of the term (Chomsky, 1959; Kuhn, 2012[1967]; Whitehead and Russell, 1926[1912]).

Watson's influence was seminal in most aspects of psychology, flooding it with the standard stimulus-response approach; there still were notable exceptions found in E.C. Tolman's "cognitive maps" and B.F. Skinner's "functional behaviorism". Skinner used a functional definition of stimuli and responses as eliciting/discriminative conditions and operant behavior (Skinner, 1957), but not unlike stimulus-response behaviorists, much of his early work was focused on nonhuman behaviors. This was, directly and indirectly, a result of the theoretical failure of behaviorism in accounting for the complex cognitive functions found in humans (Miller 2003). As interest rose in explicating the biological basis of human cognition, and isolating the factor that creates a species boundary between Homo sapiens and other primates, the gaps in behaviorist reasoning started becoming more apparent. Having been banished from mainstream psychology for decades, the writing was on the wall for an oncoming paradigm shift in the studyof mind and brain that was going to take folk psychological notions of concepts, representations, and knowledge with the seriousness it deserved and couch its defense in some form of physicalist naturalist framework, without regressing into a sort of Cartesian dualism.

Reminiscing on his personal role in these events, George Armitage Miller recalls that the foundational principles of the cognitive revolution were not so much novel, as they were Noam Chomsky's bold return to penetrating insights of Kant, Descartes and Humboldt. Miller, 2003, thus concludes, that the behaviorist fervor had occluded insights to which psychology could now return, and terms the events of the mid-1950s an anti-behaviorist counter-revolution.

1.4 New beginnings

Breakthrough disciplines often bring about a transformation in science in general, and this was no less true of the cognitive revolution.A number of developments would ultimately result in the actual shift causing psychologists and philosophers to rethink what the object of inquiry in a scientific study of human nature ought to be. This includes a growing realization of the shortcomings of the dominant paradigm of the day; the rise of then nascent disciplines like computer science, artificial intelligence,and neuroscience; and new developments in statistical and mathematical theories as well as breakthroughs in mainstream biology.The emerging developments in genetics and evolutionary biology were making it aptly clear that even in animals endowed with less complicated cognitive processes, a simple stimulus-response model could not account for their behavioral patterns.For instance, the works of biologists such as E.O. Wilson showed that an animal itself makes significant contributions toward curving out an ecological niche within which its experiential learning could take place. This contribution made by the organism is encoded in its genetic code, and a causal theory of behavior will then have to account for the ways in which the organism concerned interprets and represents its environment to itself (Wilson, 2000). As exciting an adventure as behaviorism had been for psychology, it seemed poised not to succeed in attaining the goals it had set for itself.

Observing the obsession in psychologists of the day with recording a complete history of input stimulus and accompanying behavior, Chomsky, 1959 writes:

> Putting it differently, anyone who sets himself the problem of analyzing the causation of behavior will(in the absence of independent neurophysiological evidence) concern himself with the only data avail-able, namely the record of inputs to the organism and the organism's present response, and will try to describe the function specifying the response in terms of the history of inputs. This is nothing more than the definition of the problem. There are no possible grounds for argument here, if one accepts the problem as legitimate, though Skinner has often advanced and defended this definition of a problem as if it were a thesis which other investigators reject. The differences that arise between those who affirm and those who deny the importance of the specific" contribution of the organism" to learning and performance concern the particular character and complexity of this function and the kinds of observations and research necessary for arriving at a precise specification of it. If the contribution of the organism is complex, the only hope of predicting behavior even in a gross way will be through a very indirect program of research that begins by studying the detailed character of the behavior itself and the particular capacities of the organism involved.

The premise of Chomsky's criticism here is rather simple – it is not enough to merely describe what an organism does and when, unless we have an account of how the circumstantial/ environmental information of the event is filtered by the organisms' genetically limited processing abilities and the precise nature of the information extracted therein. In simpler words, if we observe that bats echo-locate their ways around space, an explanatory account of the same should include the precise mechanisms that allow a bat to create sound waves and receive the echoed waves back in order to elicit spatial information from them.This ability is the bat's unique contribution to echo-location is independent of echoes as physical phenomena and cannot be inductively learned bya species unless they are born with it. What is difficult still, however, is coming up with such scientific accounts for infinitely more complex phenomena such as human mental processes(e.g., beliefs, desires, concepts, language, etc.).

The introductory lines of Syntactic Structures (Chomsky, 2002[1957]), a book that grew out of his initial attacks on behaviorism, aptly summarize Chomsky's intended alternative to approaching the psychological basis of mental processes:

> Syntactical investigation of a given language has as its goal the construction of a device for producing the sentences of the language under investigation. [...] The ultimate outcome of [such]investigations should be a theory of linguistic structures in which the descriptive devices utilized in particular grammars are presented and studied abstractly. [...] One function of this theory is to provide a general method for selecting a grammar for each language, given a corpus of this language

Unlike the behaviorists, Chomsky's rationalist approach is concerned less with attested patterns in behavior and more with the open-ended creativity with which children seem to manifest their linguistic abilities (Aarsleff, 1970; Bever, 2009; Chomsky, [1966] 2009). This is reminiscent of Bertrand Russell's concern with how it is that human beings, whose contact with nature is so brief and transient, come to learn so much of their environments? This would lead scientists of the day to look elsewhere in search of possible alter-natives, and here Chomsky's seminal review of behaviorism comes to play an important role.Applying this general observation to linguistic phenomena, Chomsky asks how is it possible for a child to instinctively create new utterances to which it has never been exposed specifically, based only on exposure to a fraction of the possible sentences of a language. The conclusion that Chomsky drew from this line of thinking was already a staple in mainstream evolutionary biology (Mayr, 1961; Mayr, 1982) and amounted only to an acknowledgment that all epigenetic developments follow a trajectory that is genetically pre-determined. In other words, to learn something properly, an organism needs to have at least a preliminary outline of what it is that needs to be learned and what constitutes evidence for such learning. Not unlike David Hume, Chomsky points out that though organisms no doubt derive much of their understanding of nature through experimental reasoning, that ability itself is something that they derive through the original hand of nature, and on this, they improve little to nothing throughout their lifetime.

All biological organisms are rather rigidly constrained in things that they can, and cannot,make sense of (Mayr, 1961; Wilson, 2000),and this constrain shapes and determines the way an organism would represent its knowledge of the world to itself. Psychologists of the day,however, were still reluctant to admit any notion of representations, as such talks were considered mentalistic , and being unobservable externally, mentalistic discussions were little more than mathematical obscurities to the behaviorist. This would change radically, however, with the rise of computer science and the growing realization that mathematical computations were in fact realized in nature on physical substrates. Here, at last, was a way to give a scientific expression within psychology to the link between cognition and behavior in a manner that was unlikely to fall short of biological complexity.

1.5 Computation and cognition

The discussions so far have straddled various scattered bits and pieces of reasoning, with a common unbroken thread running through them – why do organisms behave the way they do, and what causes their behavior? The fundamental problem with behaviorism had been the problem of causality. It is not sufficient evidence of a causal relationship that event Y is often preceded by event X. Missing from such correlational observations, among other things,is the notion of levels of interaction that often lead to emergent phenomena. A scientific theory can only be evaluated at the appropriate level it is addressing, and while behaviorism flourished at locating correlations between X and Y, for example, its ability to tease apart whether the two events are correlated directly or as a result of intervening phenomena at levels not addressed by its stimulus-response methodologies was limited. This problem concerned not just linguistic behavior but held true for most issues concerning complicated cognitive phenomena of any kind.One of the first major empirical findings to put a finger on this issue involved Claude Shannon's information theory. Shannon's theory, based on Markovian processes, had an inherent allure for behaviorists and their stimulus-response methodology. Miller, 1951; Miller, 1956; Miller and Chomsky, 1963 was among the first to illustrate that Shannon's theory was drastically lacking in explanatory prowess beyond elementary analysis of written sequences. Recounting his early efforts in this direction, Miller, 2003, p. 1 writes:

> But information measurement is based on probabilities and increasingly the probabilities seemed more interesting than their logarithmic values,and neither the probabilities nor their logarithms shed much light on the psychological processes that were responsible for them.

Miller's empirical findings would lead him to look elsewhere for explanatory accounts of cognition, and this would eventually lead him, and others, to Chomsky's syntactic theory. This shift in perspective, while slow, opened the door for reconnecting the two sides of the Atlantic and made possible meaningful collaborations with research groups that were unaffected by the primarily American behaviorism. These would include, among others, Bartlett's works on memory in the UK, Jean Piaget's seminal research on child acquisition in Geneva, as well as A.R. Luria's foundational attempts at unifying the mind and the brain in Moscow. Being unaffected by the behaviorist aversion to abstractions, these researchers were willing to look beyond quantifiable behavior for explanations. Chomsky's syntactic theories seemed suited to become the strongest contender for an alternative framework, primarily because they were not statements of behavior but rather formal mathematical statements of cognition with rules of cognition proving the means for transformation of one cognitive state of affair into others.

This change in attitude would coincide with breakthrough developments in computer science. On the 10th of September, 1956, The Special Interest Group in Information Theory organized their seminal symposium at MIT, which directly lead to some of the foundational concepts in computation and cognition being tabled for discussion. Newell and Simon, 1956 discussed their logic machines, while Miller (1956) discussed the nature and organization of human memory architecture. Alongside these scholars, John McCarthy and Marvin Minsky (Minsky, 1961; Minsky and Papert, 2017) were laying the foundations of what would become artificial intelligence. Finally, while Chomsky was revolutionizing linguistics by demonstrating that language exploits all the precision of mathematics and explicating its computational nature (Chomsky, 1953; Chomsky, 1959; Chomsky et al., 1963), Bever, Fodor,and colleagues (Bever, Fodor, and Weksel, 1965b; Bever, Fodor, and Weksel, 1965a; Fodor and Bever, 1965; Abrams and Bever, 1969) produced foundational research highlighting the cognitive subtleties of acquisition . Miller, 2003 writes that computer scientists like Newell were primarily invigorated by Chomsky's exposition of the algorithmic beauty of language, as it was finally making available for formal mathematical investigation one of the most complex human behaviors. Other related developments included Szikali's experiments on perception and processing and Birdsall's work on application of signal detection theory to perceptual studies.

The major formal recognition of some thing akin to a scientific revolution involving multiple disciplines would come with the Sloan Foundation's interests in funding extensive research into what was by now being referred to as the cognitive sciences. Following the establishment of the cognitive studies group at Harvard, information-processing psychology at Carnegie Mellon, and an international meeting at MIT in 1974 organized by Massimo Piattelli-Palmarini bringing together biologists, like the Nobel laureate Salvador Luria, and linguists like Chomsky and

Lenneberg, the foundation facilitated the formation of a highly interdisciplinary committee to summarize the state of the art in cognitive science. Miller (2003) writes that the driving concepts behind the emerging areas of research were summarized in a hexagonal figure,with the six corners representing the six major fields that would contribute expertise, while the lines joining them stood for an interdisciplinary methodology combining the tools of the disciplines. Thus, the union of computer science with the computational interpretation of linguistic phenomena would lead to computational linguistics, while computational methodologies would model neural functions elucidated by neuroscientists to create cybernetics. The Sloan Foundation accepted the report in 1978 (Miller 2003), and subsequently, funding was made available to various universities for recruitment of scholars and facilitation of scholar mobility, seminars, and collaborations. This directly resulted in an outburst of scholarly works that pushed the limits in a multitude of disciplines and brought about methodological paradigmatic shifts in all six disciplines. Linguistics rapidly matured from being the red-haired stepchild of philology to a proper scientific study of a biological phenomenon (Massimo Piattelli-Palmarini coined the term biolinguistics in 1974 to describe this emerging new field); philosophers left their armchairs and indulged in sophisticated empirical pursuits, while computational sciences established once and for all that computation was more than a mathematical convenience and constituted a fundamental natural phenomena. The study of thinking was no longer the domain of metaphysics and speculation, and the human mind was ready for a naturalistic inquiry into its functional and architectural limitations.

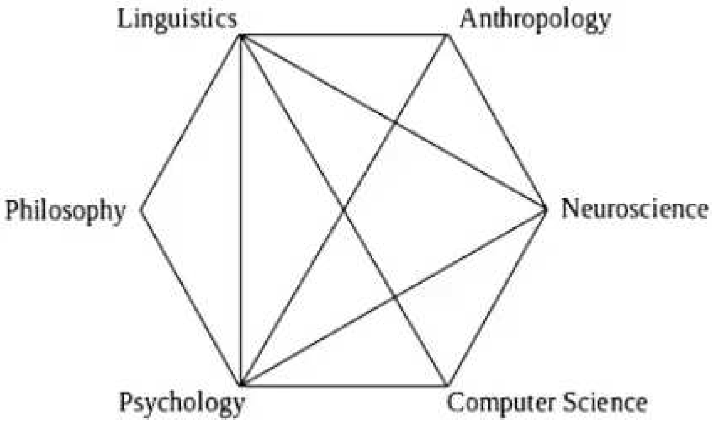


FIGURE 1.1: The cognitive sciences

1.6 Impact and aftermath

The impact of the events that unfolded in the1950s, and through the 1960s and 1970s, has been profound enough that the series of events have been termed the cognitive revolution. While the term revolution is prone to imply a somewhat violent detachment from all things past, a number of scholars have pointed out that in reality, it was more of a return to unanswered questions and unexplored problems yesteryears (Chomsky, [1966] 2009; Fodor, 2000). The cognitive revolution was primarily driven by a need to better understand the nature of information and the role of a biological organ, the brain, in packaging and storing it in a fashion that could feed cognitive processes. The foundational frameworks were laid by the works of Chomsky and Halle, 1965; Chomsky, Halle, and Encrevé, 1973; Chomsky, 1980, Fodor, Bever, Garrett, et al., 1974; Fodor, 1983, David Marr (Marr, 1977; Marr, 2010), and Lenneberg, 1953; Lenneberg, 1967 among others. Chomsky 's attack on behaviorism opened the door for a return to the old questions concerning universality of the human mind. In Cartesian Linguistics,Chomsky, [1966] 2009 notes that the idea that there must be something universal about the structure of the human mind, and its unique abilities including language, date back to Descartes and featured prominently in the works of the Port Royal Grammarians. Chomsky's formalist investigation of the information structuring of natural languages (Chomsky, 1985), as well as his works on formal languages and logic (Chomsky and Schützenberger, 1959), helped firmly establish the need to approach the functioning brain as a mathematically describable information processing system. Early research in computer science would then help establish that computations are, indeed, executable on physical substrates further lending credence to the plausibility that the brain could really be something akin to a computer. It was, however, the philosopher and cognitive scientist Jerry Fodor who would propose the most impactful analysis of the mind as a computer with his series of expositions of the computational theory of mind (Fodor, Bever, Garrett, et al., 1974; Fodor, 1983). Fodor argued that mental states, such as beliefs and desires, are relations between individuals and mental representations. He maintained that these representations can only be correctly explained in terms of a language of thought (LOT) in the mind. Furthermore, this language of thought itself is an actually existing thing that is codified in the brain and not just a useful explanatory tool. Fodor adhered to a species of functionalism, maintaining that thinking and other mental processes consist primarily of computations operating on the syntax of the representations that make up the language of thought. For Fodor, significant parts of the mind, such as perceptual and linguistic processes, are structured in terms of modules, or "organs," which he defines by their causal and functional roles.These modules are relatively independent of each other and of the "central processing" part of the mind, which has a more global and less "domain-specific" character.

Likewise, the neuroscientist David Marr wasa pioneering influence in furthering theunderstanding of the biological basis of information processing (Marr, 1970; Marr, 1977; Marr, 2010) . Marr's seminal book, *Vision* (Marr, 2010), argued for what is often referred to as the tri-level hypothesis – one must understand information processing systems at three distinct, complementary levels of analysis(Table 1).

Level	Description
Computational	What problem does the system solve?
Algorithmic	How is the problem solved? What data structures are required?
Implementational	How is such a system physically realized?

TABLE 1.1: Marr's (1982) Tri-level hypothesis

The convergence of the sort of psychological nativism and computationalism advocated by Chomsky, Fodor, and Marr has been influential in, some senses, completely refocusing the research questions that concern scientists involved in understanding the biological basis of cognitive functions (Bever, Fodor, and Weksel, 1965a; Carey and Smith, 1993; Fodor and Pylyshyn, 1988; Bever and Poeppel, 2010; Embick and Poeppel, 2015; Hauser, Chomsky, and Fitch, 2002).A complete review of its influences is beyond the scope of this chapter, but briefly, its effects have been dominant in four major strands of emerging research – the evolution of complex cognition, the realization of highly formal computations on physical/biological substrates,the constraints on such computations and their realizations stemming from general laws of nature, as well as in more applied researches into bioinformatics and artificial intelligence.

The discussion of evolution of cognitive abilities has a long and complicated history in the scientific literature. While curiosities and speculations about the universal nature of the mind can be traced back to Plato (cf. Plato's Problem, in Chomsky, [1966] 2009), with significant contributions by Kant, Descartes,and Hume (Chomsky and Schützenberger, 1959; Chomsky, 1959), discussions of evolutionary origins of such abilities like language were declared taboo by the Société de Linguistique de Paris in 1866. With the advent of the cognitive revolution, however, a naturalist framework for such discussions was made avail-able that did not have to regress to Cartesian dualism (Chomsky and Dinozzi, 1972). This effect was not limited to linguistics in particular either, as the fundamental basis of the revolution was predicated upon the issue of mental representations. A representational system, linguistic or otherwise, is an algebraic system of symbolic variables (Gallistel, 1981). The symbols are the representational primitives that form representations that are isomorphic to another system, the represented system.The isomorphism implies that the mathematical relations that hold between the symbols bear a similar form to the ones holding between the things being represented.Because the forms of the relations that the variables enter into are the same between the two systems, it ensures the validity of the conclusions drawn about the represented system through an analyses of the symbols of a representational system.Computational processes are algorithmic instructions that dictate the manipulation of representational symbols, leading to generativity. Accordingly, (Gallistel and King, 2011) have argued that the ability to represent numbers, and to recall them from long-term memory, is the most important mechanism that underlies higher cognition. In a similar vein, many researchers argue that a science of cognition is essentially the science of representational systems and their inner workings (Gallistel, 2001; Marcus, 2003).The exact nature of the evolutionary process that led to such computational representational systems is still a hotly debated topic. While Pinker and colleagues have argued that the computational mind is a result of adaptation through natural selection (Plotkin, 1997; Pinker, 2005), yet others have criticized the adaptationist approach (Fodor, 2000; Hauser, Chomsky, and Fitch, 2002; Fitch and Cutler, 2005; Fodor and Piattelli-Palmarini, 2011). The growing interest has contributed significantly to the rise of evolutionary psychology as a major scientific discipline, with its insistence on a massively modular mind that is the result of specific adaptational pressures (Pinker, 2003). While maintaining its commitments to computational-representational theories, the evolutionary psychology approach has none the less attracted criticisms directed both at its conception of modularity and computationalism from other computationalists (Fodor, 2000; Fodor and Piattelli-Palmarini, 2011).

Effects of the cognitive revolution have been also prominently felt in neuroscience and psychology, especially with the advent of easily accessible neuroimaging tools. The ease of observing the functioning brain in action has led to a plethora of imaging studies that have shed light on the many intricacies involved in cognitive processing (Poeppel et al., 2004; Poeppel and Idsardi, 2011; Hickok and Poeppel, 2004; Hickok and Poeppel, 2007; Mesgarani et al., 2014). As useful and insightful as imaging studies of various kinds have been, however, the cognitivist perspective has exerted an even greater influence over theoretical neuroscience. For instance, David Poeppel and colleagues have argued in a series of influential publications that in order for cognitive neuroscience to be successful, appropriate linking hypotheses must be posited between cognitive and neural phenomena (Bever and Poeppel, 2010; Poeppel, 2012). Because neuroscience and formal cognitive investigations operate on different sizes of conceptual granularity, the authors argue that there is no direct translation of phenomena from one domain in terms of the lexicon of the other. Two important factors are pointed out – (i) the granularity mismatch issue points out that the phenomena that concern the two disciplines, and their primitives, are not similar in the size of their conceptual granularity, and (ii) the ontological incommensurability problem is a reminder against greedy reduction of cognitive primitives to neural substrates (Poeppel, 2012; Embick and Poeppel, 2015; Poeppel and Embick, 2017). Likewise, Krakauer and colleagues have argued that in order to attain a properly explanatory account of the neural basis of cognitive processes,neural examinations should be informed by theoretical insights as well as thorough behavioral examinations (Krakauer et al., 2017)

A similar concern is seen in the refocusing of the goals of linguistic theory through the emerging biolinguistic program (Piattelli-Palmarini and Uriagereka, 2004; Piattelli-Palmarini and Uriagereka, 2008; Hornstein and Boeckx, 2009; Martins and Boeckx, 2016; Yang et al., 2017). In a series of much-debated publications, Hale, Reiss, and colleagues (Hale and Reiss, 2008; Isac and Reiss, 2013; Reiss, 2017; Bale and Reiss, 2018) have argued for an I-language perspective, a speaker's implicit knowledge of her language, that which yields her observable utterances, is a system of computation that operates over symbolic variables. The rules of this computation constitute her knowledge of language, and a study of grammar constitutes an investigation of the shared properties of all such possible states of knowledge. Likewise, building on Reiss' arguments of modularity (Reiss, 2007), the current work (chapters 4 and 6) propose a framework for exploring the partial veridicality that holds between physical speech stimuli and mental representations of sounds and meaning. It is argued that cognitive representation and processing of sounds of natural language are purely algebraic computations and that the primitives of such computations are independent of the stimuli that accompany their externalization. Similar arguments have also been made by cognitive scientists and psychologists exploring other types of information processing in the brain (Gallistel, 1981; Pylyshyn, 1984; Fodor and Pylyshyn, 1988; Marr, 2010).

It is perhaps a sign that aforementioned developments have been able to bridge the gaps between the cognitive sciences and the more basic natural sciences, that in recent years there has been an increase in research activities focusing on the place of cognition within nature and the status of cognitive sciences as natural science.For instance, Medeiros and colleagues have reported that linguistic computations manifest putative universal patterns observed generally in nature (Carnie and Medeiros, 2005; Medeiros, Piattelli-Palmarini, and Bever, 2016) and that linguistic structures like the x-bar schema are results of optimal computations in a very general sense. Krivochen and Saddy, 2018, like-wise, report empirical evidence to support the same position. Further, Piattelli-Palmarini and colleagues (Piattelli-Palmarini and Vitiello, 2017) have reported remarkable isomorphisms between various cognitive processes and field quantum theory, while seminal research in bioinformatics have recently reaffirmed the feasibility of classical computations on biological substrates (Green et al., 2017).

1.7 Conclusion

In the end, it remains a matter of perspectives whether or not the cognitive revolution was a single event localized to a single year in history,or a series of events over the years, and whether it had set for itself a fixed set of goals and agendas. What can be said definitively, however, is that a sort of Kuhnian paradigm shift has gradually come to pass and has exerted enormous influence across several disciplines (psychology, philosophy, neuroscience, linguistics, anthropology, etc.) and resulted in the prosperity of newer ones (artificial intelligence, cognitive science). It was marked by a shift in perspectives from observation of behavior and correlational stimulus-response behaviorism to more causally explanatory theories of cognition. It has not,however, been a uniform movement, having given rise to many contending frameworks (cf. Hauser, Chomsky, and Fitch, 2002; Fodor and Piattelli-Palmarini, 2011 vs. Plotkin, 1997; Pinker, 2005 and Prince and Smolensky, 2004 vs. Hale and Reiss, 2008) , that nonetheless can be argued to share broad assumptions regarding the object of inquiry. As an intellectual movement, it was sparked by a break from a dominant paradigm, was inspired by newly emerging fields of research in computation, and has been fundamentally influential in refocusing several research questions even if it has not always provided conclusive answers. There are few conclusive ends in science, being that as a project, science is characterized by continually evolving understandings of nature. However, to the extent that scientific inquiry is dependent on scrutiny and periodic readjustment of its methods and tools, the cognitivist phenomena have been nothing short of a revolution.

Chapter 2

What phonology is: a twice-told tale

2.1 Introduction

The previous chapter outlined a brief history of *The Cognitive Revolution* (Mandal, 2021; Miller, 2003), a series of events that, beginning in the mid-1950s, had a transformative effect on the study of the mind (Chomsky, 1956; Miller, 1951), machines (Newell and Simon, 1956) and computing intelligence (Turing, 2004; Turing, 2009), through the gradual fusion of six independent disciplines – linguistics, anthropology, psychology, philosophy, computer science and neuroscience (Miller, 2003) – leading to the birth of what has come to be known as *cognitive science* or, perhaps more justifiably, the *cognitive sciences*. Cognitive science was to be a science of the mental capacities of animals, with a particular focus on *homo sapiens* and specifically the uniquely human capacity for *natural languages* (NL), but with an overt and explicit acknowledgment that there is more to the *mind* than meets the eye manifested through *behavior*. Surprisingly enough, this conception of the mind was not quite birthed in first half of the twentieth century as almost half a decade ago, in 1890, William James (James, 1890; James, 2007[1890]) had introduced the concept of the "*stream of consciousness*" – the continuous flow of thoughts, sensations, and perceptions in an individual's mind – emphasizing the internal and dynamic ever-changing nature of the phenomena. James was a psychologist of colossal importance and yet, as the previous chapter discussed, psychology itself had turned away from the study of the mind and towards the cataloguing of behavior under the influence of *behaviorism*. Psychology in the 1950's was in no position to participate in cognitive science (Miller, 2003), and the cognitive revolution itself was, in effect, a counter-revolution of sorts, and overt attempt at unlearning the behaviorist dogma that acknowledged little in the way of mental phenomena divorced from behaviour, either punished (leading to inhibition) or rewarded (causing reinforcement) externally. For the behaviorist only that exists which is behaviorally observable, and thus any admissible evidence for the mental capacities of an organism should, and indeed *must*, be limited to behaviour. This focus on the external, Miller, 2003, p. 141 points out (emphasis added), had a crippling effect on the scope of psychological investigations, and indeed on that of psychologists' research.

> The behavioral revolution transformed experimental psychology in the US. *Perception became discrimination, memory became learning, language became verbal behavior, intelligence became what intelligence tests test.* By the time I went to graduate school at Harvard in the early 1940s the transformation was complete. I was educated to study behavior and I learned to translate my ideas into the new jargon of behaviorism. As I was most interested in speech and hearing, the translation sometimes became tricky. *But one's reputation as a scientist could depend on how well the trick was played*.

Miller, 1951's seminal book, *Language and Communication*, by his own account had an overt behaviorist bias and in the preface Miller (1951) writes,

> The bias is behavioristic – not fanatically behavioristic, but certainly tainted by a preference. There does not seem to be a more scientific kind of bias, or, if there is, it turns out to be behaviorism after all.

Yet, Miller, 2003, pp. 141–142 acknowledges that all his attempts to converge at the architecture of natural language (NL) through probabilistic dissection of verbal behavior failed to yield any causally explanatory account, and by the mid-1950s Miller had come to accept that a scientific theory of NL cannot be a theory of behavior, but rather a theory of the mental cognitive processes that underlie observed behavior (emphasis mine).

> Five years later, inspired by such colleagues as Noam Chomsky and Jerry Bruner, I had stopped pretending to be a behaviorist. So I date the cognitive revolution in psychology to those years in the early 1950s. [...] I was therefore ready for Chomsky's alternative to Markov processes. Once I understood that Shannon's Markov processes could not converge on natural language, I began to accept syntactic theory as a better account of the cognitive processes responsible for the structural aspects of human language. *The grammatical rules that govern phrases and sentences are not behavior. They are mentalistic hypotheses about the cognitive processes responsible for the verbal behaviors we observe.*

The history of what followed, in general, has been elaborated upon in the previous chapter. But I repeat this short reminder here for a reason. The focus of this work is phonological theory, purportedly a core component of any overarching scientific theory of NL (Chomsky and Halle, 1968; Halle, 1992; Anderson and Lightfoot, 2000; Bromberger and Halle, 2000; Fitch, 2000; Scheer, 2010a; Chabot, 2021). Yet, starting with SPE (Chomsky and Halle, 1968) the overt trend in mainstream generative phonology (Sagey, 1986; Archangeli and Pulleyblank, 1994; Prince and Smolensky, 2004; Wilson, 2006) has seen an increased focus on tasking phonological theory with a neat-fit around behavioral and typological patterns attested in speech. Arguing for a substantively-biased phonology Wilson, 2006, p. 968, for instance, states:

> The substantively biased model yields detailed qualitative and quantitative fits to the pattern of behavioral data [...]

Wilson's (2006) study is concerned with cross-linguistic trends in palatalization in the context of vowels with varying degrees of phonetic frontedness, and learners' ability to generalize based on exposure to palatalization in the context of relatively more (e.g. [i]) vs. less fronted (e.g. [e]) vowels. Based on observed asymmetry in learners' generalization of context – being exposed to palatalization in the context of [e] is more likely to induce palatalization in the context of [i] than in the other direction – and the cross-linguistic fact that languages that palatalize in the context of [e] are typically likely to also palatalize in the context of [i] Wilson concludes that phonological grammar in inherently biased in favor of substantively motivated patterns. Wilson's (2006) experimental data is, of course, impeccable. However, there appears to be at least two assumptions in Wilson's (2006) discussion of his results that are a priori, in the sense that they do not follow from the data but are rather axiomatic positions in light of which he evaluates his experimental results. First, Wilson (2006) appears to assume that typology and acquisition are constrained only by phonology. This is clearly not the case as acquisition of language crucially depends on both grammatical and extra-grammatical (second and third) factors (Chomsky, 2005). Here Wilson (2006) appears to straightforwardly violate the notion of *modularity* (Fodor, 1983), a cornerstone of generative thought since its inception (Scheer, 2010a). Second, Wilson (2006) assumes, without providing any explicit arguments against a modular approach to grammar, that typological and behavioral data form explanandum for phonological theory. This is a surprisingly behaviorist-esque stance on grammar, and one that was clearly rejected at the very onset of generative thought. Chomsky, 1959, for instance, writes (emphasis added):

> Putting it differently, anyone who sets himself the problem of analyzing the causation of behavior will (in the absence of independent neurophysiological evidence) concern himself with the only data available, namely the record of inputs to the organism and the organism's present response, and will try to describe the function specifying the response in terms of the history of inputs. *This is nothing more than the definition of the problem.* There are no possible grounds for argument here, if one accepts the problem as legitimate, though Skinner has often advanced and defended this definition of a problem as if it were a thesis which other investigators reject.

> The differences that arise between those who affirm and those who deny the importance of the specific "contribution of the organism" to learning and performance concern the particular character and complexity of this function and the kinds of observations and research necessary for arriving at a precise specification of it. If the contribution of the organism is complex, the only hope of predicting behavior even in a gross way will be through a very indirect program of research that begins by studying the detailed character of the behavior itself and the particular capacities of the organism involved.

As Chomsky aptly points out, describing an organism's behavior in terms of a historical record of its received inputs does not amount to an account of what drives said behavior, but merely provides a definition of a problem that is to be solved. An explanatory account, on the other hand, elaborates on the causal underpinnings that drive an organism's behavior, and explains the biological endowments that form the pre-requisites for its ability to process environmental inputs. In the case of phonology, this amounts to asking what facts about the architecture of phonology, if any at all, leads to the asymmetries in acquisition and typology that are reported, what factors suggest that such asymmetries are attributable to phonology at all, and if they are not then how else can such asymmetries be accounted for? Unfortunately, post-SPE phonology has been characterized by an overt desire to limit the scope of phonological computations by trying to attain a tighter fit around only observed typological patterns and behavioral tendencies through an increased focus on making computations as natural as possible, whenever possible (Archangeli and Pulleyblank, 1994; Wilson et al., 2004; Wilson, 2006). In other words, any computation that is not immediately natural is suspect, while at the same time, as Chabot, 2021, p. 152 aptly points out:

> However, the review of naturalness [...] shows that not only was phonetic information not subject to the same scrutiny, it was specifically given formal status in an effort to provide explanatory adequacy.

> One effect of enshrining naturalness as a formal element was to give it status in UG itself, since what is in formal theory is by nature in UG. The result is a commensurate enlargement of UG, resulting in a rich structure and a suite of innate mechanisms used to describe individual languages or linguistic properties, such as the battery of markedness constraints typical of OT.

The position adopted in the rest of this work will be that a neat fit around typological and behavioral data is neither the objective of phonological theory, nor does it amount to explanatory adequacy. A tighter theory is just that, tighter. But not necessarily more explanatory. Moreover, considering the fact that universal grammar (UG) is explicitly realist in its conception (Chomsky, 1956; Chomsky, 1959; Chomsky, 1980; Chomsky, 2002) – i.e. it is predicated on the assumption that grammatical artefacts are *real* neurobiological events and phenomena that are biological qualities of the human brain, encoded in (partially) in the human genome – a tighter theory, attained at the cost of a richly articulated UG, adds explanatory burden rather than providing explanation. If the wealth of markedness factors that are usually attributed to grammar in order to make it more natural are, in fact, grammatical, then it follows that any theory of UG must account for how such considerations evolved. A richer UG is, thus, harder to validate scientifically. The more that is contained *in* UG, or posited as part of UG, the more that the biolinguistic grammarian must explain in terms of selection pressure (Dennett, 2002; Fitch, Hauser, and Chomsky, 2005). In this sense, adding more to UG, then, does not so much explain the problem of language architecture as it pushes it back a step – the theory is tighter because UG richer, but this richly articulated UG itself requires explaining. Likewise, Chabot, 2021, p. 143, aptly points out that the problem with assuming that grammar must fit tightly around observed patterns is that there is always the possibility that a hitherto unattested pattern could be found in a newly discovered language. The point behind this argument, what Chabot calls the *remote island* argument is not hard to see. What constitutes linguistic typology, after all, is not some scientifically exhaustive repository of all the possible forms natural language could take, but rather a subset of possible languages limited by, among other things, the relatively short recorded history of linguistic research, biases of linguists, historical accidents (genocide, for instance, could have wiped entire languages), availability of research grants etc. (Reiss, 2007; Hale and Reiss, 2008; Chabot, 2021). The remote island argument simply points out this fallacy that if grammar is fit tightly around the set of attested (i.e. documented) languages then any "remote island", with a possibly undocumented language, could require either adding more stipulations to the theory, or throwing out stipulations that were hitherto considered a core component of it.

Contrary to this trend, the position adopted in this work will be the one espoused by substance-free phonology (SFP) (Hale and Reiss, 2000a; Hale and Reiss, 2008; Blaho, 2008; Chabot, 2022a). Substance-free phonology is the idea that phonological theory is a theory of knowledge and not behavior, of computation and not typology, and that functional motivations do not belong within its purview. In the context of the preceding discussions, this entails two related positions that run directly against the position advocated for by, among others (Sagey, 1986; Browman and Goldstein, 1992; Flemming, 2013; Gallagher, 2011; Wilson, 2006). First, SFP assumes that typological trends and verbal behavior are affected by a myriad of factors that are not phonological in kind. SFP being *modular* (Fodor, 1983) in kind this implies that a myriad of properties that are attributed to phonology are, in fact, in need of removal from its purview. A direct outcome of this is that phonological grammar shrinks in size. The SFP position, as Scheer, 2010a aptly puts it, is *small is beautiful*. As the size of grammar shrinks its scope, however, enlarges – a grammar with less stipulative statement predicts more in the way of possible sound patterns than one with more stipulations. This, also, is a core assumption of SFP – grammar predicts, or at least should predict, patterns that are humanly computable, and not just those that are attested. While against the grain of much of post-SPE generative phonology, the rest of this chapter traces the history of the development of phonological theory, from the early structuralists to present day SFP, and attempts to establish that far from being novel, the SFP story is really a twice-told tale. And one that was taken to be a cornerstone of the cognitive (counter)revolution.

2.2 Phonology: the early years

The generative conceptualization of natural language (NL) as a system of abstract structural properties, indirectly related to meaning, was inherited directly from Neogrammarian and structuralist thought. Chabot, 2021 points to two ideas specifically that would move the focus of linguistics from diachronic philology to the study of abstract mental, and ultimately neurobiological, properties of humans. The first of these was Courtenay, 1870[1972]'s insistence on moving away from diachronic philology and towards the study of abstract linguistic structures, all of which de Courtenay held to be intrinsic properties of the mind itself. In other words, these abstract properties were not cultural artefacts of language use, but rather imposed by the mind from inside out. A second related development, Chabot (2021) points out, was Saussure, 1916[1967]'s conceptualization language, itself a two-prongued framework. First, Saussure saw linguistic inquiry as a synchronic affair, a study of a systematic relationship between abstract *forms* and the arbitrary meanings that are assigned to them. Second, Saussure was the first to make a distinction between *parole* – the externalized utterances, speech forms and tokens, that ultimately lend themselves to the socio-cultural notion of externalized language in use – and *langue* – the abstract structural properties, the forms, that Baudouin de Courtenay argue emerge as properties of the mind. Inherent in de Courtenay and Saussure's conceptualization of language was the assertion that the underlying structural properties of NL are only partially visible in externalised speech forms. This position is taken up and further developed in generative grammar (Chomsky, 1953; Chomsky, 2002[1957]; Chomsky, 2014[1965]) where the foundational importance placed on abstractions is, perhaps, most visible today in the domain of (phrasal/clausal) syntax. Syntacticians do not contest that linguistic knowledge (of the syntactic variety) is almost never visible in linearized strings, nor are constituent structures available as substantive cues in speech. It is not, in other words, a matter of surprise that syntactic knowledge is a form of abstraction that is layers removed from the speech signals that convey them. But when it comes to phonology this issue is still contentious. Chabot, 2023, pp. 2–3, thus, points out that even within generativism phonology occupies a fraught position:

> Phonology however, is strung somewhere between a theory of unobservable properties of the mind and directly measurable physical phenomena observable in the senses—compared to syntax, it is not as immediately clear what the empirical remit of phonology is.

As mentioned previously generative thought inherited a lot of its philosophical underpinnings from Neogrammarian and structuralist works of old. Within the domain of phonology this is perhaps most visible in the Saussurian (Saussure, 1916[1967]) position that the primes of phonological representation are not high resolution encoding of the details of acoustic and articulatory information, but rather symbolic markers that only capture the linguistically relevant aspects of speech sounds. For the structuralists the prime was the phoneme, but the phoneme was to be understood distinctly from the prototype of a particular kind of speech sound. Indeed, for the structuralists a phoneme is defined not by the articulatory mechanisms that produce its physical tokens, nor by the acoustic percepts that aid its recognition, but rather by its relation to all the other phonemes in the inventory of the language. Thus, a phoneme stands in opposition to all the other phonemes in the inventory, and the inventory itself is a system of oppositions. This system of opposition lends a phoneme its distinctive property, and this property is not reducible to any physical quality of the externalized tokens of the phoneme. Chabot, 2023, p. 3 writes:

> In Sapir's 1907 grammar of Takelma, for example, he argued that a glottalized affricate [tˀs] stood in opposition to a plain sibilant [s], and that the former's phonological status was in fact /sˀ/, despite its opaque realization as an affricate [...] Sapir's argument was based on the theoretical system of contrasts between [pˀ tˀ kˀ] and [b d g]: [tˀs] is phonologically /sˀ/ because it stands in opposition to phonological

> /s/ in the same paradigmatic manner. The existence of /s^ʔ/ in the minds of speakers of Takelma is revealed not just through careful observation, it requires reasoned consideration of those observations through the lens of an abstract theory.

Thus, for Saussure, the physical (articulatory and acoustic) properties of speech sounds are abstracted away in the phoneme, which stands uniquely as a mental symbolic marker of how said phoneme relates to all the other phonemes in a network of linguistic contrasts. In Saussure' (1916 [1967]:166) words, *"dans la langue il n'y a que des différences"*. This is also a view argued by Edward Sapir (Sapir, 1925), for whom the phoneme espouses two distinct qualities – first it stands as a mental symbol for a set of contextually varying sounds, and second the symbol for this set of contextually varying sounds occupies a distinctive position within a network of opposing units. The phoneme Sapir describes as being *"psychologically aloof"* precisely because the distinctive quality of it derives not from some combination or subset of its physical qualities, but from the uniquely linguistic contrast it helps establish. Thus, the [t^{ʔ}s] in Saussure's grammar of Takelma is physically realised as an affricate, but stands phonologically as /s^ʔ/. This same sound, with similar physical qualities, can indeed be a phonological affricate in yet another language. Thus, Sapir points out that the psychological aloofness of the phoneme means that speakers can either perceive physically different sounds as the same phoneme, while the same phoneme can be physically realised as varying acoustic events in a context-sensitive manner. It is the form, rather than the physical qualities, that concerns phonology which, in turn, is about mental instantiation of linguistic knowledge. Put another way, the phoneme belongs in the realm of *langue*, while articulation and perception, the concern of phonetics, belong to that of *parole*.

For the structuralists, thus, there existed a visibly pronounced bulkhead between the mental symbols that were the phoneme, and the acoustic events and articulatory mechanisms that manifested its physical forms, or the "sounds" of speech. Linguistic inquiry, likewise, was the study of the symbolic mentalistic function of phonemes, abstracted away from the functional aspects of perception and articulation of speech sounds. Roman Jakobson's work would then begin the shift away from this purely formal conceptualization, while simultaneously decomposing the phoneme into sub-phonemic distinctive features. Here we begin to see the groundwork for generative conceptualization of phonology starting to take shape. First, through a decomposition of the phoneme into sub-phonemic units, or distinctive features. But perhaps more crucially by beginning the gradual dismantling of the bulkhead between physical qualities of speech sounds and the purely formal mental symbols. For Jakobson (1949) the binary oppositions marked by the distinctive features were not derived solely from a system of contrastive relationships like the structuralist phoneme was, rather they were based on the specific articulatory and acoustic correlates of those contrasts. Indeed, recognising this Halle, 2005, p. 28 writes:

> A major consequence of Jakobson's 1928 proposals was that phonetics and phonology were seen as a single field of study: that is, it was recognized that not only do phonetic facts shed important light on the phonology of a language, but also facts of phonology provide insights into the nature of phonetics.

That phonetics and phonology are connected in non-trivial ways is not a matter of debate. Phonology governs the linguistic patterning of contrasts in a language's inventory, while phonetics provides acoustic and articulatory shape to phonological forms. However, the issue under contest is to how the two modules interact. Since its inception generative grammar has been modular (Fodor, 1983) in kind (Scheer, 2010a; Chomsky, 2014[1965]). By this logic, then, phonology and phonetics should be considered separate, and independent, modules for the same reason that morphosyntax and phonology, or for that matter logical form (LF) and phonological/phonetic form (PF), are considered separate modules. However, Fodor's (1983) articulation of modularity of mind is significantly younger than Jakobsonian phonology, and indeed the *Sound Patterns of English* (Chomsky and Halle, 1968), and while modular conceptions of the mind were not necessarily unheard of prior to 1983, it would hardly be fair to judge Jakobson and the Sound Patterns of English by the same standards of modularity. The following sections outlines the development of phonological theory, with a particular focus on the relationship to phonetics, and attempts to trace the roles modular conceptions of mind have played in phonology leading up to the newly re-emerging substance-free program.

2.2.1 Sound patterns of English

The Sound Patterns of English (SPE; Chomsky and Halle, 1968) builds on and extends the Jakobsonian approach to phonology, adopting both the sub-phonemic zeitgeist and the substantively laden features. While SPE attempted to replace many of Jakobson's acoustic features with the now familiar articulatory features (Halle, 2005), the reliance on substance remained intact. In fact, in the now famous Chapter 9 of SPE, Chomsky and Halle, 1968, p. 400 bemoan the fact that the system of grammar they have developed is capable of overgenerating – i.e. it generates grammars that are never attested by linguists.

> The entire discussion of phonology in this book suffers from a fundamental theoretical inadequacy. Although we do not know how to remedy it fully, we feel that the outlines of a solution can be sketched, at least in part. The problem is that our approach to features, to rules, and to evaluation has been overly formal. Suppose, for example, that we were systematically to interchange features or to replace $[\alpha F]$ by $[-\alpha F]$ (where $[\alpha = +$, and F is a feature) throughout our description of English structure. There is nothing in our account of linguistic theory to indicate that the result would be the description of a system that violates certain principles governing human languages.

The authors are concerned that this overgeneration implies that their theory is lacking in explanatory prowess as it is unable to distinguish possible from impossible languages. Their solution to this issue, one that is taken up with great enthusiasm by virtually every post-SPE theory of phonology, is to allow phonological computations direct access to phonetic substance

Chomsky and Halle, 1968, p. 400, what they call the *"intrinsic content"* of features, besides of course a wealth of morphological and syntactic information.

> To the extent that this is true, we have failed to formulate the principles of linguistic theory, of universal grammar, in a satisfactory manner. In particular, we have not made any use of the fact that the features have intrinsic content. By taking this intrinsic content into account, we can, so it appears, achieve a deeper and more satisfying solution to some of the problems of lexical redundancy as well as to many other problems that we have skirted in the exposition.

Already we can see the structuralist bulkhead between form and substance beginning to weaken in SPE, as phonological computations are now allowed to be constrained by phonetic substance. Indeed, Chomsky and Halle's recognition of the intrinsic phonetic content of features meant that not only do features themselves contain the information needed for their externalization, but also that this content can now be utilized to restrain the ways in which rules manipulate symbols. Phonetic substance opens the way for phonetic motivations to either favor or inhibit certain patterns of changes. For instance, if it turns out to be the fact, as it indeed does (Maddieson, 1984; Halle, 2005; Wilson, 2006), that more languages tend to palatalize velars in the context of [i] than [e] and palatalization in the context of the latter asymmetrically implies palatalization in the context of the former then this asymmetry is directly explained by virtue of the gradient frontedness of the two vowels. The more fronted the vowel the more fronted the velar (Keating and Lahiri, 1993), and conversely more fronted vowels are more likely to induce velar palatalization. This idea that phonological processes are, in some sense, meant to reflect functional phonetic pressures is termed *"naturalness"*, and in virtually every post-SPE theory of phonology it is allowed to severely constrain the kinds of phonological processes in the theory's remit. Chabot, 2022b, p. 434 points out that the importing of functional motivations into the formal realm was facilitated by the re-interpretation of the Prague school's notion of markedness, but now in a more universal fashion. While for Trubetzkoy (Troubetzkoy, 1967) markedness was a language-specific phenomena, generative phonology typically assumes that markedness is universal and reflected in the fact that of the two values of SPE-styled binary features one always carry more weight, i.e. it is less "marked" (Dresher and Hall, 2022). The unmarked value, it is further assumed, is more likely emerge victorious as the result of assimilation (synchronically) and bias the direction of language change (diachronically), and it is often argued to bias perceptual asymmetries in psycholinguistic models of speech perception (Best et al., 1994; Best, McRoberts, and Goodell, 2001). In other words, the grammar prefers patterns that are phonetically natural and markedness ensures that this bias is encoded within the formal grammar itself.

The insistence on substantive features in SPE is not surprising given the position taken up by Chomsky, 2014[1965], p. 28 in *Aspects of the Theory of Syntax*. Here, Chomsky makes a distinction between *formal universals* – those that characterize the symbol manipulating rules of computations making up the transformational part of generative grammar – and *substantive universals*, or those that make up the descriptive vocabulary of the grammar.

> It is useful to classify linguistic universals as formal or substantive. A theory of substantive universals claims that items of a particular kind in any language must be drawn from a fixed class of items. For example, Jakobson's theory of distinctive features can be interpreted as making an assertion about substantive universals with respect to the phonological component of a generative grammar. It asserts that each output of this component consists of elements that are characterized in terms of some small number of fixed, universal, phonetic features (perhaps on the order of fifteen or twenty), each of which has a substantive acoustic-articulatory characterization independent of any particular language.
>
> [...] Substantive universals such as these concern the vocabulary for the description of language; formal universals involve rather the character of the rules that appear in grammars and the ways in which they can be interconnected.

Thus, in SPE, the repertoire of distinctive features that are available to the acquisition process of NL in general are fixed, finite and invariant. Moreover, these features encode substantive cues for the sensory-motor system that are language independent, indeed it is their presence that affords the child the ability to tell the [s] in, say, *sincere* from the hissing sound produced by pressing the tongue against the teeth ridge and forcing air out. The substantive featues constitute the vocabulary in terms of which NL sound patterns are described, and the intrinsic content of the features constitutes their identity. While this aspect of SPE has been overly influential in post-SPE phonology, it is also worth noting that Chomsky and Halle (1968) were not unaware of the fact that the formal universals cannot be beholden to susbtance in non-trivial ways. Thus, in the same chapter 9 of SPE Chomsky and Halle, 1968, p. 428 acknowledge:

> It does not seem likely that an elaboration of the theory along the lines just reviewed will allow us to dispense with phonological processes that change features fairly freely. The second stage of the Velar Softening Rule of English [...] and of the Second Velar Palatalization of Slavic strongly suggests that the phonological component requires wide latitude in the freedom to change features, along the lines of the rules discussed in the body of this book.

Chomsky and Halle are well-aware of the fact that there exist patterns of alternation that are surprising given their assumption that features have intrinsic phonetic substance. The phonological computational system, by their own admission, is fully capable of computing phonetically unnatural patterns. A rather well-known phenomena that runs against every notion of phonetic naturalness in phonology, for instance, concerns the so-called *crazy rules* (Hyman, 2013; Scheer, 2010a; Chabot, 2021). Briefly, a crazy rule is any pattern of alternation that makes little to no sense when viewed from the perspective of phonetic naturalness. For instance, Chabot, 2021, p. 87 cites that in Teton Dakota an underlying crazy input class, /p,l/, surface as /m,n/ when they occur immediately preceding a nasal vowel (/p,l/ → /m,n/ /_ [+Syl, -Cons, +Nasal]). This particular rule is crazy because there is no phonetic feature that can create a class out of /p/ and /l/, while excluding every other segment in the language. Yet, speakers of Teton Dakota clearly acquire this pattern just like any other pattern that may make more phonetic sense. The crucial aspect, as Reiss, 2017, p. 4 aptly points out, if the phonological system is at all capable of computing unnatural patterns, as it clearly is (Chomsky and Halle, 1968; Chabot, 2021), then it cannot be constrained by phonetic motivations in non-trivial ways. This is, at the very least, true of phonological computations, even if we are to debate whether features themselves are substance-free as well? Clearly, for Chomsky and Halle (1968) phonological features are substance-laden, but the computations are substance-free. In this canonical version, features served three related purposes. First, they mark binary contrastive oppositions within the inventory of a language, and in this aspect they much resemble the structuralist phoneme. Second, the generalized intersection of shared features by a group of segments creates natural classes – i.e. a natural class of segments is the set of all and only those segments in a language that share some set of valued features – in terms of which rules of alternations are defined. But perhaps most importantly, features serve as encoded instructions for the phonetic systems that are tasked with perception and production of utterances. They do this by virtue of the fact that, unlike the structuralist phoneme, SPE features are not just symbolic counters but also possess inherent phonetic content. SPE entertains, briefly, the possibility of utilizing this intrinsic content to restrain the formal computations, but quickly reneges on this for the reasons mentioned before. There is, thus, still a bulkhead of sorts maintained in SPE, not strictly between form and substance as in structuralism of old, but at least between formal (phonological rules) and substantive (phonological features) universals. While Chomsky and Halle (1968) were explicit in their admission that allowing phonetic motivations to constrain formal computations is unlikely to prove fruitful, in spite of the intrinsic content of features, the next section provides a brief illustration that post-SPE generative phonology has, by and large, not only relied upon substantive features but also increasingly dismantled the bulkhead between Chomsky's (1965) formal and substantive universals by attempting to maximize the reduction of phonological alternations to naturalness concerns.

2.2.2 Substance abuse in generative grammar

The notion of *"substance abuse"* (Hale and Reiss, 2000a) was sarcastically introduced by Hale and Reiss to refer to the overt reliance on functional substantive considerations in post-SPE phonological theory. The previous section discussed the role afforded to substance in SPE, where their function was limited to affording features with a phonetic identity. While Chomsky and Halle experiment with more substantive possibilities in chapter 9, they nonetheless make it abundantly clear that they do not believe that the substantive constraints will ultimately prove to be fruitful. In other words, chapter 9, in my opinion, is not so much bemoaning of the framework constructed earlier in the book but rather a practical illustration of what the authors saw as the limitations of substance in phonology. The primary criticism directed by proponents of substance-freedom (Fudge, 1967; Hale and Reiss, 2008; Chabot, 2021) at mainstream generative phonology concerns the fact that much has been made about the skepticism expressed by Chomsky and Halle, 1968, p. 400 in chapter 9 of SPE regarding the "overly formal" nature of grammar, while at the same time not much attention has been paid to the later sections of that same chapter where Chomsky and Halle, 1968, p. 428 admit the futility of allowing substance to constrain phonological rules. Put another way, SFP proposes that instead of importing functional motivations into the formal grammar one ought to ask whether observed typological and behavioral asymmetries reflections of constrains imposed by the grammar itself, or can these be explained by appeals to grammar-external factors?

Hale and Reiss, 2000a, 158ff. point to Beckman, 1997's discussion of positional faithfulness constraints as an example of substance-abuse in generative grammar. Briefly, Beckman's Optimality Theoretic (Prince and Smolensky, 2004) proposals are meant to allow faithfulness to [High] to be maintained only in root-initial syllables. Beckman proposes two constraints to this end:

Ex 2.2.1. Positional faithfulness constraints in Beckman, 1997

1. IDENT-σ_1[High]

A segment in the root-initial syllable in the output and its correspondent in the input must have identical values for the feature [High]

2. IDENT-[HIGH]

Corresponding segments in output and input must have identical values for the feature [High]

Hale and Reiss, 2000a term this formulation *substance abuse* because such formulations duplicate within phonological theory explanations that can be obtained elsewhere, i.e. from experience or third-factors (Samuels, 2009). They rightfully point out that Beckman's constraints need not be endowed by UG, because mere experience and asymmetric biases inherent in the perceptual channel (Moreton, 2007; Moreton, 2008) suffice to explain the effects intended by the constraints. For instance, Hale and Reiss suggest that we imagine a language, call it L_1, that allows high vowels in all syllabic positions, i.e. in initial, medial and final syllables, and also contains only initial stress. It is a well-established fact in psycholinguistic research that stress is realized primarily as increases in duration and intensity of vocalic signals, and that the same exhibit increased perceptual salience to the neonate learner from fairly on in life (Cutler, 2012). Given this it automatically follows that any child acquiring L_2 [1], based on exposure to L_1, is more likely to be able to succesfully distinguish high vowels from mid and low ones in the initial, stressed (and thus more perceptually prominent) syllables. Conversely, the child is more likely to neutralize the distinction in mid and final syllables. Over time such asymmetries lead to diachronic sound change, and achieve the same functional outcome as Beckman's constraints. Hale and Reiss, 2000a, thus, argue that since grammar external explanations are available that explain the positional distribution of high vowels, it is redundant to duplicate this substantive information in the formal grammar.

---

[1] We call this L_2 because within the I-Language perspective (Chomsky, 1956; Chomsky, 2002[1957]; Chomsky, [1966] 2009) every native-speaker's state of knowledge is unique and individual.

Beckman, 1997 is not the only example of substance-abuse in generative grammar, however. A more recent and well-known example comes from Colin Wilson's experimental investigation of acquisition of velar palatalization. Wilson observes that learners exposed to palatalization in the context of a less fronted vowel (e.g. [e]) are more prone to extend the generalization to more fronted vowels (e.g. [i]), and that this generalization is one-directional and asymmetric. Based on this Wilson concludes that phonology is substantively-biased, and proposes the implicational laws of palatalization – palatalization in the context of less fronted vowels implies palatalization in the context of more fronted vowels. There are two related ways that these laws can be interpreted. First, one could argue that across languages palatalization in the context of [e] asymmetrically implies palatalization in the context of [i]. That is, if a language contains both vowels and it palatalizes in the context of [e] it will also palatalize in the context of [i]. This, of course, does not say anything about whether such palatalization is phonological, or purely co-articulatory. The second claim made by Wilson is a clear illustration of substance-abuse – that the grammar itself is biased towards substantively motivated patterns, even though by Wilson, 2006, p. 84's own account the bias is "not so strong that it excludes unfavored patterns". The explanation for Wilson's implicational laws can, again, be found in channel concerns. It is a well-established fact that velars and palatals are more perceptually confusible in the context of front vowels (Keating and Lahiri, 1993; Guion, 1996). Indeed, Keating and Lahiri, 1993 observe that the more fronted the vowel the more palatalized the velar becomes. Given this, it is not surprising that learners are more likely to generalize palatalization in the direction of less to more fronted vowels. If a learner's perceptual processing yields a palatalized velar in the context of [e] that asymmetrically increases the probability of the same in the context of [i]. The reverse, of course, is not true. Cues for palatalization are stronger in the context of [i] than [e], and detection of the former does not guarantee detection of the latter. Like Beckman, 1997, then, Wilson, 2006 has imported performance factors into the realm of competence. This is vindicated by Wilson, 2006, p. 968's own words as he defends his view of a substantively-biased grammar,

> the substantively biased model yields detailed qualitative and quantitative fits to the pattern of behavioral data [...]

The concern of a formal computational-representational theory of grammar is a description of the implicit knowledge of a native speaker, knowledge that is only partially reflected in verbal behavior.(Chomsky and Halle, 1968; Bromberger and Halle, 2000; Samuels, 2015a; Chabot, 2021; Chabot, 2023). By assuming that observed trends in either typology, or verbal behavior, is a direct explanandum of phonological grammar Wilson, 2006 appears to have ignored the possibility that similar explanations might be found outside the grammar.

2.2.3 Substance-free phonology

Substance-free phonology, in essence, is not so much a novel idea as it is a re-emergent one. The structuralists, as discussed before, maintained a strict bulkhead between form and substance. Even before the publication of SPE, Fudge (1967) advocated for phonologists to burn their phonetic boats, observing that phonetically motivating feature theory encounters numerous problems. Fudge, 1967, p. 2 writes,

> The use of Jakobsonian distinctive features as primes on the systematic phonetic level rests therefore upon the assumption of a one- to-one correspondence between articulation and recognition. Although this correspondence frequently comes close to being one-to-one, there are points in most, if not all, languages at which certain articulatory details are irrelevant from the viewpoint of recognition[...] There are auditory (i.e. recognition-oriented - we reserve the term 'acoustic' for features on the physical-phonetic level) distinctive features which correlate with more than one articulatory feature: for example [+Flat] may indicate any one of lip-rounding, pharyngalization, or retroflexion [...] A generative phonology of a language based

> on auditory features must (if it is to achieve its aim) further specify which manifestation of Flatness appears in that language, which can only be done in articulatory terms.

Reviewing feature assignment in a wide range of languages, including Tswana and Hungarian, Fudge points out that this problem is not entirely resolved by assuming articulatory features, as this leads to redundancy problems in speech recognition/perception. Fudge's conclusion is remarkably similar to Chomsky and Halle, 1968, p. 428's, who also conclude that allowing phonetic substance to restrain phonological computations is unlikely to prove fruitful. Starting with Hale and Reiss, 2000a modern SFP seeks to return phonology to the purely formal domain. At the heart of SFP is the idea that phonology is not about action, but rather cognition. It is a species of representation and computation in the brain that determines what aspects of humanly producible and processable vocal sounds are linguistically relevant, both, across *all* human languages and in individual speakers' specific languages. The objective of phonological theory, in the substance-free view, is not to achieve a perfect fit with attested typological or verbal behavioral trends, but rather to make predictions about the humanly computable sound patterns and the syntax of the rules that operate over said patterns to produce alternations (Hale and Reiss, 2008; Reiss, 2007; Chabot, 2021). Swimming against the contemporary tide of functionally motivated theories, SFP argues that phonological properties are opaque to substantive motivations, and phonological rules make no reference to functional pressures. While SFP is by no means a monolithic theory, and there exist a wide variety of proposals regarding both the representational and computational components (cf. Blaho, 2008, 8ff. for a full discussion), it is characterized a set of overarching programmatic goals that defines the program. The rest of this section is dedicated to first elaborating on the core representational and computational assumptions of SFP, and then highlighting a central fissure that broadly divides the program into two opposing schools. I focus primarily on the representational aspect, however, as this where the fissure appears deepest and this also where phonology's interface with the substantive domain becomes most relevant.

SFP is built-around the idea of Fodorian modularity (Fodor, 1983) – the idea that the mind is comprised of independent and informationally encapsulated computationa modules, each specified for a particular kind of computation. One consequence of modularity is that every module is endowed with its own unique and proprietary representational format, or vocabulary. For phonology this vocabulary consists of distinctive features. The nature, and label, of features has been the subject of much debate in phonological theory (Jakobson, 1949; Chomsky and Halle, 1968; Goldsmith, 1976a; Sagey, 1986; Halle, Vaux, and Wolfe, 2000; Halle, 2002; Hale and Reiss, 2008; Chabot, 2021), but the general consensus within SFP appears to be that features are purely symbolic entities, abstractions that exist only in the mind of the speakers with no causal roots in the mind-external world. Put another way, SFP is a neo-Sausseurian approach to phonology, with strict delineations between form and substance, that seeks to uncover the mechanisms by which the linguistically relevant contrastive aspects of substantive information (Saussure, 1916[1967]; Sapir, 1925) are encoded in the mind/brain. This representational issue is also the first point of departure where disagreements within the program begin to appear. If phonology is substance-free then it begs the question how its representational primes connect with substantive phenomena, which constitute the primary data for phonological analyses? In SPE features were endowed with intrinsic substantive content, and despite Chomsky and Halle, 1968, p. 428's forewarning in chapter 9 post-SPE phonology increasingly made references to substance in order to fit phonological theory to typological trends. Within SFP, which eschews such functional substantive motivations in phonology, the link to substance is enabled through an intermodular interface, tasked with converting substance-free phonological forms to substantive information in perception and production. Two distinct approaches have emerged, broadly dividing the field into a raadical school (Blaho, 2008; Dresher, 2014; Chabot, 2021; Odden, 2022) and the Concordia school (Hale and Reiss, 2008; Reiss, 2007; Reiss, 2017), each advocating starkly different conceptualizations of what substance-free phonological primes are and how they relate to substance.

The radical school argues that the primes of phonological representation, distinctive features, are neither innate nor universal (Samuels, 2011; Dresher, 2014). Rather, they emerge out of domain-general innate abilities of categorical perception and symbol creation (Samuels, 2011; Samuels, 2015b). Humans, like many other mammals (Kuhl and Miller, 1978; Kuhl, 1991), are endowed with the ability for categorical perception of continuous sounds. The radical school holds that this ability is utilized during the critical period to create symbolic contrasts within a language’s inventory. Given that different languages exhibit different systems of contrasts, it automatically follows that they have different active features. Further, the precise features that are active in a language are a function solely of the phonological behavior of segments within that inventory, and never a reflection of its phonetic shape. For instance, Dresher, 2014 discusses the case of Manchu vowel harmony. In Manchu the vowels /ə/ and /u/ trigger ATR harmony within a word: in suffixes, /ə/ alternates with /a/ and /u/ alternates with /ʊ/.

Ex 2.2.2. ATR Harmony in Manchu: Stems with ATR and non-ATR vowels

ATR	xəxə	’woman’	xəxə-ŋgə	’female’
Non-ATR	aga	’rain’	aga-ŋga	’of rain’
ATR	xərə	’ladle out’	xərə-ku	’ladle’
Non-ATR	paqt’a	’contain’	paqt’a-qʊ	’internal organ’

Dresher further points out that the vowel /i/ is neutral, and can co-occur with both ATR and non-ATR vowels. In these examples, the suffix vowel /i/ resists alternation. In the roots, the vowel /i/ likewise are ignored by the harmony process which is affected by other root vowels.

Ex 2.2.3. ATR Harmony in Manchu: /i/ remains neutral

ATR	pəki	′firm′	pəki-lə	′make firm′
Non-ATR	paqtʂ′in	′opponent	paqtʂ′ i-la	′oppose′
ATR	sitərə	′hobble′	sitərə-sxun	′hobbled′
Non-ATR	panjin	′appearance′	panji-sχʊn	′having money′
ATR	əmt′ə	′one each′	əmt′ə-li	′alone′
Non-ATR	taχa	′follow′	taχa-li	′the second′

Dresher, 2014 argues that this pattern of activity suggests that the vowels // and /u/ must be specified for an active feature that /i/ lacks, namely [ATR]. Crucially, the vowel [i] on the surface displays all the phonetic properties of an ATR vowel, but lacks the specification for the feature. This observation lies at the core of the radical SFP framework – features are purely formal markers of phonological activity, and not sensory-motor commands for externalization. Consequently, the feature specifications of segments cannot be predicted purely from their phonetic shape. It is, thus, possible for a phonetically ATR vowel to lack [ATR] and phonetically velar consonant to lack [Back]. This renders the traditional feature-labels more or less useless, a sort of historical baggage borne out of methodological convenience at best. Indeed, Chabot, 2022b suggests that traditional feature-labels, such as [ATR] and [Round] can, and should, be replaced by simple variables, such as α, β, γ etc. Further, since features' phonetic interpretations are completely arbitrary, they cannot be learned, and must be acquired during the critical period.

The Concordia school of SFP (Hale and Reiss, 2000a; Hale and Reiss, 2008; Reiss, 2017; Volenec and Reiss, 2017), on the other hand, makes exactly the opposite assumption. Reiss, 2017 argues that features are both innate and universal. In this view, the child starts out with a complete set of features endowed by UG, and based on experience with ambient data unnecessary contrasts are deactivated during the critical period. Crucially, in their original works Mark Hale and Charles Reiss (Hale and Reiss, 2000b; Hale and Reiss, 2008; Reiss, 2007; Reiss, 2017) closely stick to the original SPE proposal, discussed in detail by Chomsky and Halle, 1968, 400ff in the now famous chapter 9, that features have intrinsice phonetic content. For the Concordia school’s original conception these intrinsic contents serve as sensory-motor commands that determine the phonetic shape of phonological forms, and are reflected in the usual feature labels, which are also granted ontological status. However, the phonological computations themselves are agnostic to this content. Consider, for instance, the pattern of velar palatalization in Malayalam. In Malayalam, velar consonants are palatalized when preceded by the vowels /i,e,a/.

Ex 2.2.4. Malayalam velar palatalization.

Nominative	**Dative**	**Gloss**
kutti	kuttikʲkʲə	’child’
makal	makalkkə	’daughter’
Base	**Causative**	**Gloss**
mara	marakʲkʲ-	’cover’
wilarppe	wilarkk-	’paleness’
UR	**SR**	**Gloss**
wekkə	wekʲkʲə	’cook-IMP’
tukkə	tukkə	’stand-IMP’

Mohanan and Mohanan, 1984, 584ff argue that the palatalized velars are derived from underlying velar consonants in the context of preceding [-Back] vowels. They acknowledge that the vowel [a] is phonetically a back vowel, but for purposes of palatalization pattern as [-Back](Mohanan and Mohanan, 1984, p. 586). Within the Concordia school’s framework, however, this is not a possible solution. This is because within this framework the feature specifications of segments are directly reflected in their phonetic shape, and as such [a] must [+Back]. Instead, Mandal, 2023 argues that the rule is triggered by the class of [-Round] vowels – [-Continuant, + High] → [- Back] / [- Consonantal, - Round] _ [- Consonantal]. Given the five vowel (/i,e,a,o,u/) inventory of Malayalam, it is argued that the proposed rules provides a natural class of triggers where there is no mismatch between the phonetic shape and phonological specifications of vowels. This approach successfully derives the required surface forms, however it might not be optimal given that it misses the generalization that across languages velar palatalization is triggered by front vowels (Halle, 2005; Wilson, 2006). The contentious issue here, obviously, concerns the nature of the relationship between phonological features and their substantive correlates. For the classical Concordia view features contain intrinsic content, and thus a vowel that is intrinsically a back vowel could not possibly be specified [-Back]. Otherwise, the intrinsic content loses all meaning. However, framing the rule in terms of [-Round] vowels makes immediate sense because, as Hale and Reiss, 2008 point out, rules do not take into consideration the content of features, manipulating them as arbitrary symbols. In a sense, then, the Concordia view of subtance-free phonology is only partially substance-free – computations are substance-free (recall, chapter 9 of SPE; Chomsky and Halle, 1968, p. 428), but representations are not.

The partial substance free formulation of the Concordia school has been subject to numerous criticisms by those who seek to pursue a fully substance-free phonology (Blaho, 2008; Samuels, 2009; Samuels, 2011; Chabot, 2021; Odden, 2022; Scheer, 2022). In an attempt to counter this criticism the Concordia school, in recent times, has clarified their position to claim that features too are substance-free (Volenec and Reiss, 2017; Volenec and Reiss, 2020; Reiss and Volenec, 2022a). However, proponents still maintain that features have deterministic substantive interpretations. Particularly, Volenec and Reiss, 2017 argue that substance-free features are mapped onto universal and invariant substantive correlates through a pair of transducers, a system they term *Cognitive Phonetics, or CP*. This system is discussed in greater detail below, where I argue that Volenec and Reiss not only fail to provide any empirical evidence for CP, but also misinterpret the neurobiological literature cited, fail to take into account the antagonistic architecture of the articulatory muscular system, and essentially retain the intrinsic content of SPE-style features, just without calling it the same.

Obviously, the point of contention in the substance-free program lies in how to relate a substance-free phonology to substance. At its heart it is the application to phonological theory the minimalist methodology (Chomsky, 2014) of positing only the bare minimum of representational primitives (distinctive features) and operational procedures (the syntax of phonological rules) that are required to provide an explanatorily adequate account of native speakers' knowledge of their language. The program is also strictly modular, from whence stems its rejection of functional substantive motivations for phonological processes. Such motivations need to be stated in substantive terms which, per modularity and proprietary vocabulary, are illegible to phonology. However, by the same token phonological forms are illegible to the sensory-motor systems tasked with externalization, which requires detailed spatio-temporal information. The point of contention, thus, lies in how to relate form to substance and do so in a manner that does not violate modularity.

2.3 The phonological grounding of phonetics

I have established, in the previous sections, my intent to adopt a Fodorian modularity driven approach to phonology, and its interfaces. This naturally raises the issue of what the architecture of those interfaces might be. As a disclaimer, for reasons of space I will be focusing only on the interface with phonetic substance, the necessary condition for the sensory-motor systems to be able to produce and comprehend speech[2]. In particular, in this section I elaborate on the Concordia school's (Volenec and Reiss, 2017) approach to the lower interface – the primary focus of my criticisms – which proposes a deterministic interface, driven by computational algorithms operating upon innate features. The radical school (Scheer, 2010a; Samuels, 2011; Dresher, 2014; Chabot, 2021; Odden, 2022), on the other hand, proposes a completely arbitrary

---

[2] This does not imply that the phonological module does not interface with other modules, for instance the visual module (Berent, 2013). Those, however, are beyond the scope of my dissertation.

interface with emergent features and learned phonetic interpretations. Proposals made within this school, for instance by Scheer, 2010a and Samuels, 2011, have been primarily of a theoretical nature, focussing mainly on architectural consistency between the upper and lower interfaces. This will be taken up in chapter 4, where I combine parts of this proposal with an analysis-by-synthesis model proposed by Bever and Poeppel, 2010 and Poeppel and Idsardi, 2011, and propose a hybrid framework that obeys strict modularity. My main purpose in this section will be to first outline a set of criticisms that the Concordia school has consistently failed to address, over at least three separate publications advocating Cognitive Phonetics, or CP (Volenec and Reiss, 2017; Volenec and Reiss, 2020; Reiss and Volenec, 2022a). I then discuss how this failure stems from logical and empirical inconsistencies inherent in CP, and point out instances where CP fails to adhere to even those sections of the neurobiological literature that it claims to build on.

SFP assumes that the input to and output of phonology are, both, composed of features drawn from the same set. They are vocabularies of the same *type* (Scheer, 2010a; Reiss, 2017). In other words, the output of phonology, the surface-representation (SR), is fully substance-free and algebraic. Production, and perception, of speech sounds, however, require access to substantive information (Chabot, 2021; Hickok, Houde, and Rong, 2011). However, the point of departure within the broader substance-free program begins with the details of how form should be related to substance. I will consider the Concordia school in greater detail in this section because (a) it will constitute the primary focus of my criticisms, and (b) it is the school that has made the most detailed neurobiological claims.

The Concordia School's proposal, Cognitive Phonetics (CP), assumes that a transductive interface makes use of a pair of syntagmatic and paradigmatic transduction algorithms to systematically, and deterministically, imbue phonological features with substantive-information (Volenec and Reiss, 2017). Volenec and Reiss (2017) sketch out a programmatic suggestion for

Cognitive Phonetics (CP), derived primarily from Lenneberg, 1967's discussion of the neurophysiology of articulation (Lenneberg, 1967), but make what I will argue to be conceptual errors in interpreting Lenneberg. I devote my attention to Lenneberg's original suggestions, the major motivations borrowed by CP, and raise a few foundational issues that are going to prove essential to the central claim of deterministic transduction, and thus by extension to the issue of labels as feature-identity.

Lenneberg's concern is much like Sapir's in that both point out the complexities involved in mapping between abstract atemporal mental representations of speech segments and the physiological outputs that constitute speech signals. Noting that there is a fundamental difference between the structure of the SR and the physical output – the former are discrete beads on a string, while the latter gradient acoustic signatures produced by continuous articulator movement – Lenneberg proceeds to provide an algorithmic overview of what this process might look like. The table below represents Lenneberge's (1967) illustration of how muscles are prepared for articulation of a successive segments. The columns (I-VI) represent segments, while the rows (a-f) represent muscles. Thus, for instance, for the articulation of segment I, muscles 'a', 'c', 'd', 'e' need contracting while muscles 'b' and 'f' are to be relaxed.

	I	**II**	**III**	**IV**	**V**	**VI**
a	+	+	+	0	+	0
b	0	+	+	+	0	+
c	+	+	0	+	+	0
d	+	0	0	+	+	+
e	+	0	+	+	0	+
f	0	+	0	0	+	+

TABLE 2.1: Muscle activity assignment schema.

Here, Lenneberg abstracts away from two substantive details. First, segments are assumed to be of uniform temporal duration, which is never the case in running speech. Second, muscles are grouped into classes based on activation latency, which also obscures gradient distinctions in latency. However, note that Lenneberg's schema much resembles Marr's (Marr, 1977) Type-1 theories – it is an idealized algorithmic representation of the hierarchical neuromuscular planning strategies involved in articulation. What it is not, and should not be treated as, is a real model of speech production. Marr (1977) points out that Type-1 Theories, or Clean Theories, decompose neatly into two conceptual tiers – a computational statement that decodes the problem to be solved (in this case truth-preserving mappings from the discrete units to the continuous movements) and an algorithm that implements a solution. The schema is Lenneberg's characterization of the computational problem itself. Now consider the tables below. In the figure on the left, adapted from Volenec and Reiss' (2017) discussion of Lenneberg, the columns reflect activation latency groups. The left most column, α, has four times the activation latency of the rightmost column δ.

α	β	γ	δ
			a
b			
	c		
		d	
e			
		f	

1	2	3	4	5	6	7	8	9
			+	+	+	0	+	0
0	+	+	+	0	+			
	+	+	0	+	+	0		
		+	0	0	+	+	+	
+	0	+	+	0	+			
		0	+	0	0	+	+	

FIGURE 2.1: Temporal co-ordination of muscle activity.

This table, thus, reflects a re-arranging of the previous table to capture the relative activation latencies of different muscles. The table on the right lists temporal events (1-9) that must occur in order to properly articulate the segments I to VI. Thus, the first temporal event concerns relaxing muscle "b" and contracting "e". Next, "b" and "c" contracts while "e" relaxes. Notice here that already the columns in this table cannot be put in a direct one-to-one correspondence with the columns in the table above due to temporal shifts required of the muscles relative to each successive segment. Next, Lenneberg (1967) proposes his famous neuromuscular schema that encodes the articulatory process as columns of temporal segments. Under each segment is listed the name of the muscles that need to contract, and in subscript the phone from table 1.5 that said muscle is associated with.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
e_I	b_{II}	b_{III}	a_I	a_{II}	a_{III}	d_{IV}	a_V
	c_I	c_{II}	b_{IV}	c_{IV}	b_{VI}	f_V	d_{VI}
		d_I	e_{IV}		c_V		f_{VI}
		e_{III}	f_{II}		d_{IV}		
					e_{VI}		

TABLE 2.2: Lenneberg's neuromsucular schema.

The astute reader would have noticed yet another curiosity in this schema. Notice that in column 3, for instance, the muscle 'e' is activated because it is needed for articulating segment number III. This activation happens *before* the activation of muscle 'a' for segment I in column 4. The activation of actions necessary for the third segment before those needed for the first segment, Lenneberg (1967) argues, reflects physiological constrains imposed by biology. For instance, notice that muscle 'e' is grouped under the highest latency α group in table 1.12, while 'a' is grouped under the lowest latency δ group. Such latencies arise from a variety of factors including cell-type, the distance the nerve has to cover in order to bridge the distance to the central nervous sytem etc. The crucial point, Lenneberg (1967) argues, is that because the order in which segments appear in a string is independent of the latency groups to which the muscles articulating said segments belong there is no necessary temporal correspondence between events at these two distinct event levels. Thus, even though segment I is temporally ordered before III, the muscles associated with latter has a much higher latency and thus needs to be activated before the muscle associated with I. The latency ensures, as it were, the segment itself is still actualized *after* segment I, in its proper position at the segmental level. As mentioned briefly before, Lenneberg's (1967) discussion is framed in structuralist terms, partly because he was writing well before the publication of SPE, while most of modern phonological theory has adopted a sub-segmental, feature-driven framework. This is where CP steps in, and attempts to extend Lenneberg's biolinguistic suggestions to the primal level of phonological theory (Reiss and Volenec, 2022b).

CP argues that the primary function of the phonology-phonetics interface (PPI) is to prepare the phonological output SRs for sensory-motor processing. This involves assigning to the constituents of phonological representations systematic cues for neuromuscular co-ordination of the type discussed by Lenneberg (1967). CP takes Lenneberg's insights and extends it to features. Instead of a schema for segments, CP conceives of this process as producing such neuromuscular schemas for the individual features that make up higher order structures, including segments and words. CP argues that this essentially involves a change in representational format, a *transduction* in Pylyshyn's (1980) terms, that produces what it calls a *phonetic representation* (PR). Crucially, PR is still not actual utterances but are still entirely brain-internal representations, just of a substance-laden form. To be specific, the output of CP are still brain-internal reprsentations, like the ouput of phonology. But unlike SR the PR can be read by the sensory-motor systems to activate the motor-effectors that will lead to phonation and utterances as actual behavioral outputs. Notice here that this eventual output will bear further effects of a multitude of factors. First, the transduction to PR involves anticipatory planning, which affects connected muscle strictures. But since this is introduced by transduction this variation can in theory be calculated, or so Volenec and Reiss, 2017 claim. It is systematic, an effect of physiological phonetics computing over PR. The second kind of variation is introduced by third factors not limited to cognition or action directly. This variation is unsystematic. If, for instance, I am speaking out the word *"Wait!"* in a certain context the way I speak will be determined, partly by the SR and PR of the form "wait". However, if the context involves an imminent danger I might scream it, while if it requires stealth I might whisper it, and if I am speaking to a child I might stretch out the vowel and modulate my pitch to attract more attention. Further, someone with a bigger or smaller oral cavity than me will produce perceptibly different tokens of the same word. Thus, effectively, speech tokens manifest a combination of two kinds of variations. CP argues that the deterministic nature of transduction to PR ensures that the sum total of possible phonetic variations always fall within some predictable range, such that truth-preserving perception of articulated forms is possible.

2.3.1 Things fall apart

So far the proposals made within the purview of CP has been of a programmatic nature. However, Volenec and Reiss, 2017 then attempt to provide some details on how CP, as a computational module, works. This where discrepancies begin to creep in. In this section I will identify, and deconstruct, three distinct problems with CP that casts serious doubts on its purportedly neurocomputational claims – (a) misinterpretation of Lenneberg's (1967) schema, (b) misrepresentation of the neurobiological literature that CP claims as its underpinnings, and (c) a lack of engagement the neurocomputational literature that have addressed the issue of mental representations. In other words, my purpose here is to show that the programmatic claims made in CP are based on literature that is, at least, half a century old (Chomsky and Halle, 1968; Chomsky, 2014[1965]) and are thus neurobiologically naive to an extent. This problem, I will argue, is exaggerated by the authors' lack of engagement with more contemporary literature that address the neurobiological issues concerning linguistic representations, both empirical (Eggermont, 1998; Poeppel and Idsardi, 2011; Lawyer and Corina, 2014; Brecht, 2017; Gwilliams et al., 2018; Monahan et al., 2022; Gwilliams et al., 2022) and theoretical (Pöppel, 1997; Watumull et al., 2014; Berent, 2021; Watumull and Chomsky, 2020; Chomsky, Gallego, and Ott, 2019; Yu and Lau, 2023).

CP conceptualizes a feature as a functionally connected activation pattern, or a neural spike. Crucially, however the unique location of the spike (place coding), and the rate of its repetition (rate coding), is claimed to be the key to *what* a feature does in being transduced. This line of argument is derived from recent neurobiological studies providing some evidence for neuronal specialization for distributions of frequency over time (Mesgarani et al., 2014; Hickok, 2022; Kemmerer, 2014), with specific sub-sets of these neurons encode acoustic patterns while others encode correlates for articulatory targets. Hickok, Houde, and Rong, 2011 argue that speech processing requires an integration of acoustic/perceptual and articulatory/motor information in the brain, primarily carried out by the Spt, a brain area conveniently situated in the Sylvian fissure at the parietal-temporal boundary. Crucially, Spt activates both during passive perception of speech sounds and during sub-vocal/covert articulation. Sub-vocal activation implies that Spt is not being driven by overt auditory feedback, suggesting rather that it is involved in sensorimotor integration (Hickok, Houde, and Rong, 2011). Further, Hickok and Poeppel (Hickok and Poeppel, 2007) report different sub-regional patterns of activity associated with the sensory and motor phases, suggesting distinct neuronal subpopulations for each phase. More broadly, while Spt is not speech-specific, activating reliably for perception of tonal melodies and tasks involving humming, speech induced activity in Spt is highly correlated with pars opercularis in Broca's region. Activity in Spt is also reported to be motor-effector selective (Pa and Hickok, 2008), with noticeably more robust activity when motor tasks involve the vocal apparatus as opposed to manual effectors. It is conveniently located in between networks of auditory (superior temporal sulcus) and motor (pars opercularis, premotor cortex) regions, and diffusion tensor imaging studies suggest that Spt and posterior sector of Broca's region are densely connected at the anatomic level. It is, thus, both anatomically and anatomically well position to support sensory-motor integration of the type required for linguistic processing of speech signals. CP posits that area Spt plays a crucial role in conjoining multiple activation patterns, each separately coding for low-level articulatory and acoustic correlates, to realize individual features (fig.2.2) below illustrates this schematically).

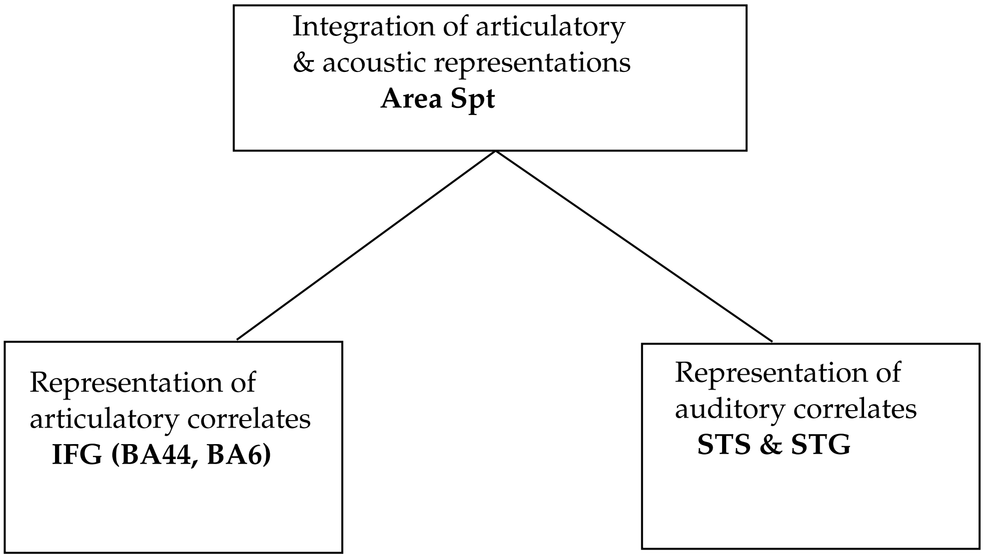


FIGURE 2.2: Wetware for features

The generative schema outlined above illustrates the basic tenets of the conjunctive neurons hypothesis (Stein and Meredith, 1993). Briefly, the conjunctive neurons hypothesis proposes that neurons in the brain combine information from different sensory modalities to create a unified perceptual experience. The hypothesis suggests that the process of multisensory integration occurs at all levels of the brain, from the primary sensory cortices to higher-order association areas. Neurons that combine information from different sensory modalities are referred to as conjunctive neurons. These neurons receive input from multiple sensory modalities and integrate this information to create a unified representation of a perceptual precept. Features are, thus, symbolic encodings of articulatory configurations and corresponding acoustic maps maintained in long-term memory (see figure below) (Volenec and Reiss, 2017, p. 270), enabling the bi-directional mapping from representations in memory to output action and the other way around from input to memory. The figure below is just a conceptual representation, but CP goes further in proposing a neuorcomputational account of how features are realized systematically across speakers and languages by appealing to a somewhat well-established hypothesis – the conjunctive neurons hypothesis (Stein and Meredith, 1993).

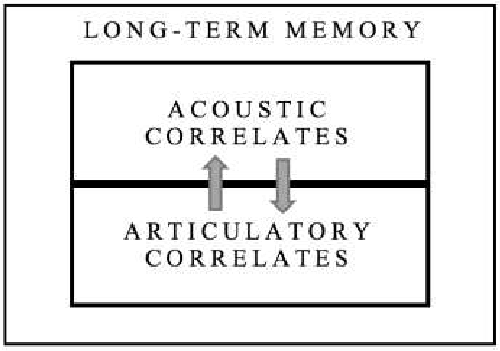


FIGURE 2.3: Phonological Feature

Unfortunately, here Volenec and Reiss, 2017 appear to make two critical mistakes. First, while liberal references are made to the works of Poeppel, Hickok and colleagues it is worth noting that the two major CP-claims that (a) the neuronal maps are specialized for features (qua features) and (b) that maps equal features (qua features) are just that, i.e. they are claims made within CP and not in the original literature cited therein(Poeppel, 2012; Hickok, Houde, and Rong, 2011; Guenther and Hickok, 2016). Indeed, the actual literature is quite clear about the position that such maps cannot be straightforwardly equated with phonological features (qua features in the theory). These preferences for spectro-temporal properties are indeed argued to be hardwired (Poeppel and Idsardi, 2011), though they are also assumed to be ecologically useful (Poeppel and Idsardi, 2011), shared across the mammalian brains and not specific to language (Fitch and Cutler, 2005). These neuronal preferences for STRFs afford a relatively high-resolution encoding of the signal, over which an initial estimate of (major-class feature) distributions can be made. This process is discussed in detail below, but note here that unlike Volenec and Reiss (2017), Poeppel and colleagues are categorical about the fact that these neural versions of spectograms (neurograms) encode substantial amounts of gradient phonetic information. For instance, Poeppel and Idsardi, 2011, p. 1073 write

> The properties of this putative intermediate representation are largely unknown, for

> the moment [...] For example, calling it 'phonological' implies a categorical representation, but the extent to which the information in each of the two windows is categorical is unclear because it is untested.
>
> [...] Such an intermediate (and fleeting) multi-time resolution representation will retain acoustic properties, but they will differ depending on whether one is looking at the shorter (segmental) or longer (syllabic) temporal primitive.

Volenec and Reiss, 2017 ignore this contentious representational issue in the section of the neurobiological literature they claim to build upon, among others to be discussed below, and simply assume the presence of hardwired preferences for spectro-temporal receptive fields (STRF) to be evidence for innate features *and* innate feature interpretations, or what they call *deterministic transduction*. What they borrow with unjustified enthusiasm, on the other hand, is the data pertaining to the functional anatomic background that has emerged from the last decade of research into speech processing (Poeppel, Idsardi, and Van Wassenhove, 2008; Guenther and Hickok, 2016; Poeppel and Embick, 2017; Hickok, 2022) and try to make a case for the type of neural representations they feel are necessary to support their representational biases. To be sure, Volenec and Reiss, 2017 provide an accurate summary of the cortical anatomy that has been repeatedly implicated in the processing of linguistic hierarchical structures. The issue I take here is not so much with the anatomical regions borrowed by CP, as it is with the granularity-blind [3] manner in which specific phonological properties are attributed to specific cortical anatomy. For instance, recall that CP proposes a set of algorithms that take phonological SRs as inputs and creates a hierarchical plan for feeding the sensory-motor system. A *paradigmatic transduction algorithm* (PTA) scans input forms and determines a muscle assignment for each feature in each matrice. This is claimed to be reflected in Lenneberg's (1967) schema discussed above. Likewise, *syntagmatic transduction algorithm* (STA) determines the temporal

---

[3] By this I mean that the straightforward way in which CP attributes linguistic functions to anatomical cortical regions appears to be indifferent to the granularity-mismatch problem discussed by Embick and Poeppel, 2015 and Poeppel, 2012, and appears naive to the issues inherent in interpreting functional anatomical and neurophysiological (Gwilliams et al., 2022) data in linguistic terms.

co-ordination of the neuromuscular commands for each feature. PTA and STA are assigned to specialized neural wetware (Volenec and Reiss, 2017; Reiss and Volenec, 2022a), and the output is integrated by specialized wetware to yield PR (see fig. 2.4 below).

Thus, for instance, the figure above (Volenec and Reiss, 2020, p. 55) attributes PTA and STA, unambiguously, to the Anterior insula and the basal ganglia and cerebellum respectively. Likewise the pre-SMA is claimed to be responsible for the integration of PTA and STA. The empirical literature, on the other hand, is much more nuanced and uncertain about such specific linguistic characterization of cortical areas. This is problematic because while most of the peri-Sylvian language areas of the brain are implicated in some aspect of speech processing, Poeppel et al., 2004 and Poeppel, Idsardi, and Van Wassenhove, 2008 rightly point out that the specific computational contribution of each cortical field remains to be discovered. Hackett, Preuss, and Kaas, 2001 point out that the first stages of speech processing occur bilaterally superior auditory areas, with subsequent lexical-level computations being carried out in a left-lateralized manner. Similarly, Binder et al., 2000, Belin, Fecteau, and Bedard, 2004 and Hickok and Poeppel, 2004 all report that the right STS and STG contribute significantly to a multitude of language processing functions, including but not limited to voice identification, dynamnic pitch analyses, identification of prosodic features. The part of the routine that permits lexical-access, on the other hand, is reported to be largely left-lateralized to the left temporal, parietal and frontal cortical regions (Binder et al., 2000; Hickok and Poeppel, 2004; Indefrey and Levelt, 2004; Poeppel and Idsardi, 2011).

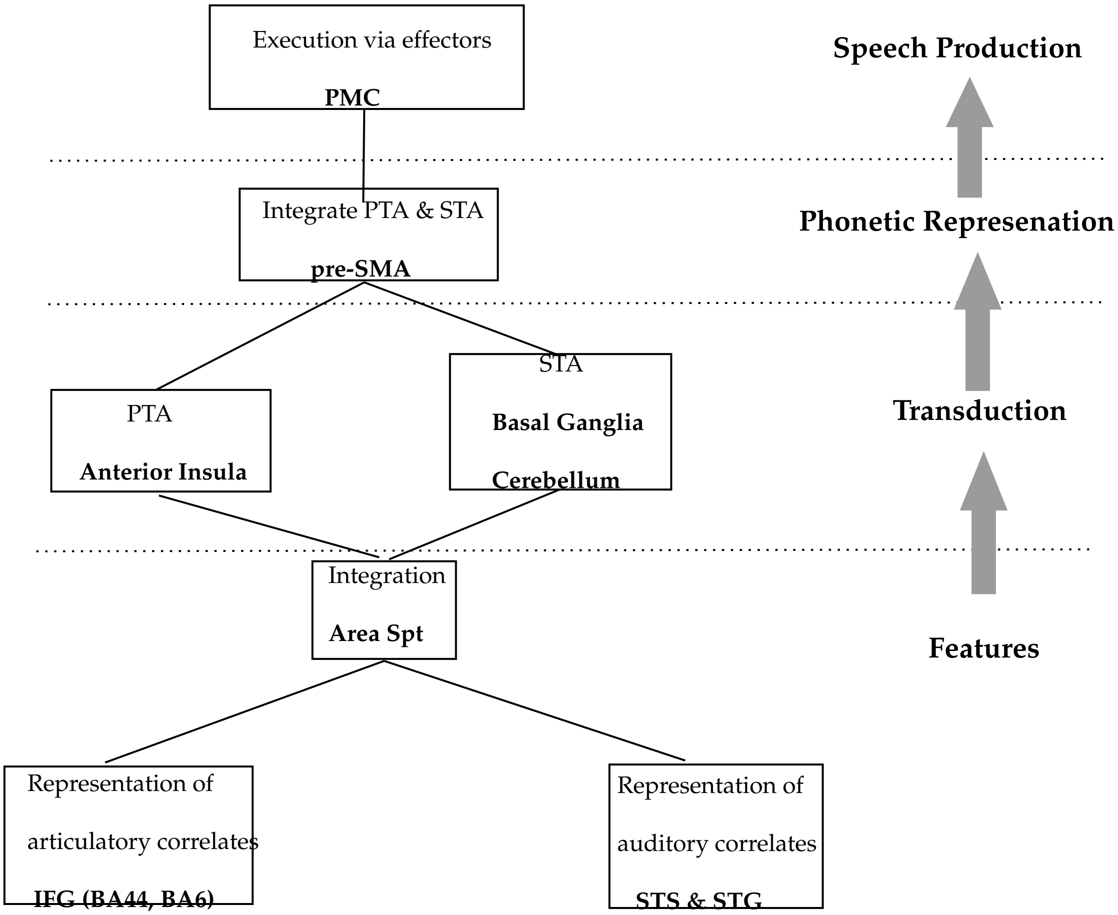


FIGURE 2.4: Transduction via CP

Not only do Volenec and Reiss not provide any discussion of these nuances, but over three different papers arguing in favor of CP-driven transduction (Volenec and Reiss, 2017; Volenec and Reiss, 2020; Reiss and Volenec, 2022b) the authors fail to provide any insights into how the different subsets of the cortical regions traditionally subsumed under the so-called "language network" come together to implement either spatio-temporal co-ordination of articulatory muscles (the workings of PTA and STA), or for that matter why they assume that any spatial-localization evidence implicating any of the regions they attribute to their models amount to a deterministic transduction. This particular issue is taken up in some detail below, but for the sake of clarity my objection here, again, is two-fold. First, the PTA and STA that Volenec and Reiss espouse are highly specialized algorithms that, by the authors' own account, operate over purely phonological information, transforming them into *true phonetic information*, the latter being purely "cognitive phonetic" and devoid of any third-factor information. This is problematic because the empirical literature cited above is very clear about the implicated cortical regions playing a significant role in systematically repackaging speech signals (unambiguously containing third-factor and physiological influences) until phonetically, and subsequently phonologically, relevant levels of granularities are attained. It is a huge jump in logic, therefore, to assume that localization of some aspects of phonetic or phonological processing to, say, the cerebellum, implies that the cerebellum is responsible for deterministically mapping phonological features to invariant phonetic correlates. Second, Volenec and Reiss repeatedly make the argument that PTA and STA are deterministic algorithms and that they localize to specific cortical regions, they provide no discussion of what the operational procedures of the algorithms are. What is claimed is that PTA and STA localize to distinct neural regions, based on evidence that may or may not favor such characterization, and also that such transductive algorithms are responsible for spatio-temporal co-ordination of articulatory muscles. The latter claim, on its own, is rather trivial. Speech is a continuous analog process, and obviously speaking requires spatio-temporal co-ordination of articulators (Cutler, 2012; Best et al., 2007; Levelt, 2013). But precisely why PTA and STA being able to be localized to their respective functional anatomies, assuming of course that they are, amount to their inner functioning being deterministic across all languages, especially given the highly antagonistic architecture of the articulatory system (Lenneberg, 1953; Lenneberg, 1967; Lenneberg, 1969), is never discussed. Rather, Volenec and Reiss combine partial evidence for localized speech processing (Binder et al., 2000; Belin, Fecteau, and Bedard, 2004; Hickok and Poeppel, 2004) with the SPE position that features have intrinsic content (Chomsky and Halle, 1968, p. 400) to weave together a story about deterministic algorithms and invariant transductions. In other words, based on partial evidence Volenec and Reiss have tried to settle the debate *a priori*. However, they are not, as

Jerry Fodor points out in *A Science of Tuesdays*[4], "the first to try to settle it a priori, but the a priori arguments don't convince". Here again one wishes that Volenec and Reiss were the realists they claim to be.

The criticisms so far have been directed, for the most part, at Volenec and Reiss' attempts to justify innateness of the transductive mechanism by referring to scattered evidence for neural specialization for aspects of speech signal processing (Volenec and Reiss, 2017; Volenec and Reiss, 2020). Specifically, I have objected to (a) the wetware referred to on the grounds that available evidence does not appear to allow the kind of inference drawn about their functions (Eggermont, 1998) that Volenec and Reiss, 2017 appear to do, and (b) on the ground that Volenec and Reiss, 2017 do not provide any discussion regarding the inner workings of PTA and STA beyond their reference to Lenneberg's neuromuscular schema (Lenneberg, 1967). It is crucial, however, to distinguish between, on the one hand, *innateness* (of primitive representations) from *localization* at the circuit level in the brain, and on the other hand between innateness of phonological primes and innateness of transduction between said primes and their substantive correlates. To begin with I want to stress that I do not object, and in fact admit readily, to the necessity of some notion of innate structures in phonology. I object, however, to Volenec and Reiss' claims regarding "neurobiological evidence" produced in support of innateness of both features and transduction. As Fodor[5] puts it (emphasis added),

> I want, to begin with, to distinguish between the question *whether* mental functions are neurally localised in the brain, and the question *where* they are neurally localised in the brain. *Though I find it hard to care about the second, the first clearly connects with deep issues about how the mind works*; ones that even us philosophers have heard of [...] I quite see why anyone who cares how the mind works might reasonably care about the argument between empiricism and rationalism; and why anyone who

---

[4] London Review of Books. Volume 22 No.14, 20 July, 2000

[5] London Review of Books,Vol. 21 No. 19 · 30 September 1999

> cares about the argument between empiricism and rationalism might reasonably care whether different areas of the brain differ in the mental functions they perform. Likewise for anyone who cares about how much of the mind's structure is innate.

Consider, for instance, that Samuels et al., 2022 in arguing for language-specific substantive correlates make it amply clear that their position do not entail a phonological tabula rasa. Rather, the authors (Samuels et al., 2022, p. 556) argue that phonological primes emerge out of a species-specific combination of shared pre-phonological primitives.

> We propose that phonological features build on innate infrastructure, but that their content is not itself innate. In other words, we agree with the position [...] that innate representational primitives are necessary for learners to begin learning a phonological system, but we contend that these primitives are not phonological fea- tures. The possibility that there are "more basic primitives at the initial state" is raised by Hale and Reiss, 2008, p. 37, who conclude that these initial primitives would still nevertheless be part of UG. Converging lines of evidence instead suggest to us that the initial primitives are (a) properties of the auditory system that are innate, but not exclusively phonological, and (b) an evolutionary inheritance of (at least) the mammalian lineage, shared with animals that do not have phonology

The proposal I have in mind is similar to what the authors allude to above. Specifically, I assume that the properties of both the auditory and articulatory channels are innate, and these provide the biological endowments that act as feature substrates. Consider the fact that neonates are born with a range of pre-phonological endowments that form the foundation for the later development of phonological features during the critical period of language acquisition. These endowments include fundamental perceptual and motor capabilities that are not yet fully formed phonological features but provide essential building blocks for their development. For instance, one precursor for the development of phonetic categories can be attested in the categorical perception ability that is attested not only in humans but also in finches and several mammalian species. Infants can distinguish between different speech sounds based on discrete categories, despite continuous variations in the acoustic signal. For instance, even before language exposure, neonates can differentiate between the voiced and voiceless sounds like /b/ in [ba] and /p/ in [pa], demonstrating a natural predisposition to parse the auditory landscape in linguistically relevant ways. This early sensitivity is supported by neurobiological structures, such as the spectrotemporal receptive fields (STRFs) in the auditory cortex, which encode the temporal and spectral aspects of sounds and provide a neural basis for processing complex auditory signals (Fitch and Cutler, 2005; Poeppel, Idsardi, and Van Wassenhove, 2008). Different sub-populations of neurons in the auditory cortex are widely known to display preferences for specific distributions of frequencies over time (STRF). These selective preferences are hardwired genetically, known to be ecologically usefull, and are highly unlikely to specific to the faculty of language in the narrow sense (Hauser, Chomsky, and Fitch, 2002). The ecological utility highlights a broader evolutionary function of STRFs: the ability to discern salient auditory features in the environment. In humans, the presence of innate STRF preferences facilitates the early detection and categorization of speech sounds, laying the groundwork for the eventual mapping of these sounds onto abstract phonological features. Thus, while the capacity for phonological feature development is unique to humans, the underlying neural mechanisms, such as STRFs, represent a shared evolutionary heritage that supports auditory perception across species. This shared mechanism underscores the notion that human neonates' pre-phonological endowments are both a product of evolutionary adaptation and a crucial substrate for the acquisition of language-specific phonological systems (Kuhl, 2004).

These innate auditory capacities are complemented by the articulatory potential inherent in the human vocal apparatus (Gick, Wilson, and Derrick, 2013). The physical properties of the articulatory channel, including the anatomy of the vocal cords, oral cavity, and nasal passages, define the range of possible sounds that humans can produce. In the framework I have in mind, these articulatory setups provide dimensions of possible phonetic contrast, such as voicing, aspiration, place, and manner of articulation, which are not themselves phonological features but (crucially) delimit the space within which features can emerge. For instance, the voicing distinction can be implemented either through VOT or presence/absence of aspiration in a language specific manner, but the crucial point is the possibilities of implementation are a priori limited by non-linguistic biological endowments. What goes for laryngeal phonetic phenomena, I assume, goes for other such possible contrasts, thus restricting the set of possible featural contrasts The critical period serves as the developmental window when these pre-phonological potentials are honed into the specific featural contrasts of a given language, shaped by the linguistic environment and guided by the brain's inherent plasticity (Werker and Tees, 1984). One benefit of this way of conceptualizing the innateness issue regarding phonological primes is that it nullifies the blank-slate argument familiar from emergentist theories, while simultaneously avoiding the temptation to justify innate features by referring to neural wetwares (cf. Volenec and Reiss, 2017). Note that the proposal presented here, not unlike Volenec and Reiss' framework, does allude to hardwired neural wetware. But unlike Cognitive Phonetics (Volenec and Reiss, 2017) it does not attempt to reduce (or localize) features to their maps. Such attempts, in my opinion, are more characteristic of embodied cognition and less of computational-representational frameworks. Symbols, which phonological primes are, are instantiated by neural wetware but are not the wetware itself.

Having attempted to clarify my position that the phonological tabula is, indeed, not rasa, I turn now to the distinction between innate features/primes and innate mappings between primes (substance-free in the overall framework I have adopted here) and their substantive correlates. Following the standard SPE approach (Chomsky and Halle, 1968) I take the primary concern of theories of the phonology-phonetics interface (the lower interface, henceforth) to be explicating the mechanisms by which discrete/algebraic values are mapped on to gradient/continuous information, and vice versa. The process of acquisition, then, involves the child figuring out exactly how this mapping is structured. Likewise, the task of the theorist is to propose a framework that provides a biologically realist model of the mental and neural events and processes that are called upon to implement this mapping. Dunbar, Dillon, and Idsardi, 2013, 21.3.1ff point out that this is a non-trivial matter given that the same, or similar, substantive values are mapped onto different phonological categories, as evidenced by both phonological processing and speech processing. Here, again, the issue of innateness and domain-specificity arises, though arguably the latter is more of a concern. No one, after all, believes that the child learns to process sounds, speech or otherwise. Neither the processes such as bone-conduction or excitation of cochlear hair-cells transforming vibrations in the air to electrical signals, nor the STRF preferences of cortical neurons that encode this signal as neurograms are "learned" in any meaningful sense. The crucial point of contention, then, concerns whether this mapping procedure is innate in the sense of being a property of UG, or whether this process co-opts non-language specific processing mechanisms. Volenec and Reiss, 2017 propose the existence of two specialized transduction algorithms, the PTA and the STA, that are tasked with systematically reading feature-matrices (i.e. segments) and calculating (a) the substantive correlates of individual features, and (b) spatio-temporal co-ordination of these values based on both intra- and inter-segmental configuration of features. Crucially, the crux of their claims is that once intrasegmental co-articulation (i.e. spill-over of feature interpretations based on intrasegmental specification of features) and intersegmental co-articulation (i.e. spill-over due to features present in adjacent segments) are computed out, any given feature (say, +Voice) will always map onto an invariant substantive correlate. The search for invariant correlates of features has a long and complicated history (Stevens and Blumstein, 1981; Stevens, 2000; Stevens, 2002; Poeppel and Idsardi, 2011), and I do not intend to even try to settle this matter here. Rather, my objective here is to simply draw attention to broad issues present in Volenec and Reiss' proposal – (a) their proposal is simply an extension of Lenneberg's neuromuscular schema (Lenneberg, 1967) from segments/phonemes (in Lenneberg, 1967, Ch.3 to features (in Volenec and Reiss, 2017, with the assumption that what goes segments goes for features, and (b) their proposal provides no discussion of the computational architecture of their algorithms nor appear to make use of the substantial literature available on feature-detection from speech signals. The second point appears especially surprising to me given that in the strong nativist position of Volenec and Reiss, 2017 such detections should be readily implementable via mechanisms of the sort discussed in the following sections. The proposal advanced in the following sections appeals to the analysis by synthesis (AxS) framework developed by Poeppel and Idsardi, 2011 and Bever and Poeppel, 2010, and adopts a largely *Bayesian* framework. This is in contrast to Volenec and Reiss' insistance on deterministic algorithms. A further point of difference, at least in my understanding of these two frameworks, is that unlike Volenec and Reiss and I make no claims to language-specificity of AxS as a mapping algorithm.

Typically in Bayesian models where a specific observed data point is a member of one of a finite number of categories is termed a *mixture model*. We can apply such a model to determine the mappings from phonological to phonetic information (Dunbar, Dillon, and Idsardi, 2013). Consider a phoneme C and some phonetic value x, and the task is to determine if they map onto each other. In other words, given C what is the probability of observing the phonetic value x?

Ex 2.3.1. $Pr(x) \; = \sum_{i=1}^{c} Pr(c_i) Pr(x|c_i)$

The equation above states that the probability of some phonetic value x is equal to the probability of the phoneme c_i times the within-phoneme conditional probability of the observed token given the assumption that x is an instance of c_i, summed over all instances of C. Crucially, however, Dunbar, Dillon, and Idsardi, 2013 point out that adopting this probabilistic framework *does not* necessarily entail a commitment to an overall probabilistic view of grammar itself. Rather, they argue that deterministic models are simply a special case of stochastic models. In terms familiar from the phonology-phonetics interface literature, then, we are simply asking, given the phonetic value x how do we determine which phonological category C

which generated x? In other words, the task at hand is to determine the value of C which maximizes Pr(c | x), expressed by *Bayes' Rule*. The following equation expresses this mathematically, with the denominator expanded using the law of total probability.

Ex 2.3.2.
$$Pr(c|x) = \frac{Pr(x|c)Pr(c)}{\sum_{i=1}^{c} Pr(x|c_i)Pr(c_i)}$$

If we now assume that there are no overlaps between phonetic categories, then given any possible value of x there could only ever be one category (c_i such that expansion of the denominator in 2.3.2 above is not zero. Dunbar, Dillon, and Idsardi, 2013, p. 13, thus, point out that specifically because the model is stochastic that

> it is capable of imputing detailed 'degrees of certainty' (probability) about various inputs (a probability distribution); nevertheless, probability distributions have as special cases both maximal certainty (determinism) and maximal uncertainty (uniform distributions). Because of this link, probability theory can be used as a way of formalizing reasoning in cases of high and low uncertainty alike. In the case of absolute certainty, it can be shown that it reduces to Aristotelian logic; when there is uncertainty, it can be shown to be reducible to a very small number of axioms of consistent reasoning.

It is, of course, possible that what Dunbar, Dillon, and Idsardi, 2013 call "maximal certainty" is what Volenec and Reiss, 2017 call "deterministic transduction". But if this is so the same is never made clear. On the other hand, this is precisely the interpretation of the word "probabilistic" (or "stochastic") that I have in mind in the discussions that follow as I attempt to sketch out an AxS driven transduction mechanism at the lower interface. To summarise, then, the proposal I have in my mind has two crucial distinctions from the one advanced under the rubric of "Cognitive Phonetics" (Volenec and Reiss, 2017). First, I admit an innate set of pre-phonological endowments that allow for an expectation of algebraic symbolic encoding of phonetic contrasts, with the possible dimensions of contrasts a priori limited by the properties of the articulatory and perceptual channels. Second, I propose that an innately available (though not necessarily FL-specific) AxS system, working largely along Bayesian lines, make use of pre-phonological endowments to categorize linguistic experience and encode them in a symbolic fashion. All subsequent discussions of the lower interface architecture advanced in this dissertation, specifically the appeals to it in chapters 3 and 5, are meant to be understood in this context.

2.3.2 The lower-interface: where we are (really) at, and where we can (realistically) go...

The main issue with CP stems not so much from the co-articulatory effects attributed to the process – for this is indubitably true (Lenneberg, 1953; Lenneberg, 1967; Lenneberg, 1969; Browman and Goldstein, 1989; Samuels, 2015a) – as it does from its insistence that CP first computes, and then occludes, some elusive invariant phonetic correlates for every feature. The human muscular system, including the subset that is tasked with vocal articulation, is essentially antagonistic in nature. That is, in the implementation of movement of any sort the degree of relaxation or contraction of any muscle is in a many-to-many correlation with the state of numerous other muscles (Lenneberg, 1967). Further, given variations in size of the oral cavity and articulatory muscles between individuals any articulatory configuration will inevitably produce varying acoustic maps. Since these maps form the input to acquisition, it follows that there exist no invariant correlate that listeners try to approximate as they discern the necessary articulator configurations. Yet, not only does CP insist on such invariance, but in proposing alleged algorithms that compute such correlates it consistently fails to elaborate on their nature. Volenec and Reiss, 2017; Reiss and Volenec, 2022a attempt to justify this by claiming that not much is known about how the brain computes anything. This is claim is exaggerated, as quite a bit is known about how the human brain processes continuous and analog signals and repackages them into discrete categories (Stevens, 2002; Poeppel and Idsardi, 2011; Poeppel, 2012), and how it utilizes somato-sensory maps to achieve muscle configurations (Brecht, 2017). CP ignores this entire body of literature, and I argue this is because much of the literature point to the varying and probablistic nature of signal processing (Poeppel, Idsardi, and Van Wassenhove, 2008), while CP insists on strict determinism. This section elaborates on this literature, and points to an alternative approach that retains an innate instinct for features (Dresher, 2014) and embraces probabilistic construction of feature-readings through experience during the critical period. In summary, the critical period is critical for a reason.

Clearly, there is a great deal of variability in speech. One of the goals of the distinctive feature theory is to recognize higher-order invariants that correspond with particular features, say ± Voice. Here again, note the importance of feature-based analysis becomes apparent. The required alternation between voiced and voiceless segments does not require replacing and entire segment, thereby requiring an entirely new speech motor plan. The effects can be achieved by changing a single feature specification. Not only can listeners recognize the same token of the same word spoken differently by the same person or even different people, but they are also capable of grouping completely contrasting sounds together as variations of the same form. A classic example comes from English listeners showing selective neural responses to /s/ vs. /z/ contrasts in a context-sensitive way. In words like "sue" vs. "zoo" listeners treat the initial sounds contrastively. However, when used in the context of plural markers listeners' neural signatures illustrate that the contrast is ignored, and both sounds grouped together under the same class (Poeppel and Idsardi, 2011), suggesting that in later stages such as morphological processing sounds are grouped into functional classes using distinctive features. Poeppel and Idsardi (2011) argue that such grouping of representations of speech sounds is attained via distinctive features acting as connectors between articulatory goals and auditory patterns. These encodings are efficient, discrete, modality-independent and task-neutral(Halle, 2002) [6]. The issue then arises as to what exactly features encode, articulatory targets, acoustic maps, or some combination of the two? Poeppel and Idsardi (2011) provide an account in terms of an interconnectred *perception-action-memory* loop that incorporates both types of information encoding.

---

[6] For a contradictory opinion see (Mielke, 2008)

The PAM loop proposes that the brain's processing of sensory information is not a one-way street but rather a dynamic feedback loop, with information flowing between different regions of the brain in a continuous and recursive manner. For example, when we see an object, visual information flows from our eyes to the primary visual cortex in the back of the brain. This information is then passed to other areas of the brain responsible for interpreting visual information, such as the parietal and temporal lobes. However, these regions are not simply passive recipients of information but instead actively contribute to the processing of sensory information, by generating predictions and expectations based on past experiences and current context. Consider the three components of the loop in turn, as it applies to natural language.

Perception: When we listen to someone speak or read a text, our brain receives a sequence of sounds or visual symbols that it needs to interpret as words and phrases. The PAM loop proposes that our brain uses past experiences and expectations to generate predictions about the words and phrases that are likely to occur based on the context and the speaker's intentions. These predictions help to guide the processing of incoming information, allowing us to quickly recognize words and anticipate what will come next.

Action: Our brain's processing of language is not just about perception but also involves motor representations of speech sounds and gestures. This means that when we hear someone speak, our brain generates motor commands that simulate the movements needed to produce the sounds and gestures of speech. These motor representations can then influence how we perceive speech, by modulating the sensitivity of auditory and visual receptors to incoming information.

Memory: Memory is a critical component in language processing. The brain uses past experiences and knowledge to help generate predictions about incoming stimuli, allowing us to quickly recognize words and comprehend meaning. For example, when we hear the word "apple," our brain draws on past experiences with apples to generate a mental representation of the word's meaning, which helps us to quickly identify the object being referred to. Likewie, in the continuous discourse listeners continuously make predictions about incoming words –

you hear *fantasti–* and our brains might predict either *fantastic* or *fantastically* depending on the context – suggesting that multiple candidates forms are activated in the memory as stimuli is processed incrementally.

Poeppel and Idsardi (2011) propose that specific features act as translation points between the articulatory space and acoustic space. For instance, the feature {+Round} is a cue for (a) the enervation of the articulatory muscles, in this case orbicularis orbis, for lip rounding and (b) downward formant transitions. The figure below illustrates the authors' proposal of how vibrations in the ear are turned into abstract representations in the brain.

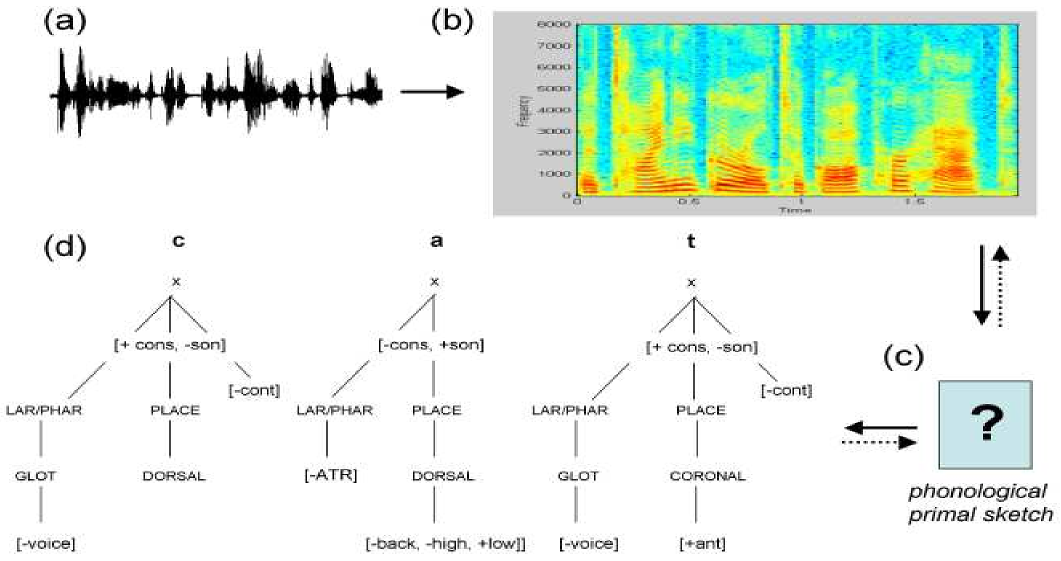


FIGURE 2.5: Transduction from speech signal to phonological representations

In the figure above, the waveform on the right (a) represents the waveform of the signal (call this R1, for representation format 1). R1 is analyzed by the perceptual channel and the patterns stored in the auditory cortex by neurons with sensitivities to specific spectro-temporal receptive fields. The authors propose that this "neural version of a spectogram" (R2 in b) as specialization for certain spectro-temporal properties such as velars demonstrating converging second and third formant transitions. They write.

> This type of representation (R2) is most likely a property of neurons in the auditory cortex, and it does not differentiate between speech and non-speech signals. Moreover, given the highly conserved nature of mammalian auditory cortex, these representations are very likely shared with other species, and consequently these representations can be investigated using animal models [...]"

Based on this representation (R2) the brain computes multiple representations in parallel at different temporal windows, utilizing multi-time resolution processing. The authors point out that one crucial aspect of the granularity of representations at this level – marked with ? (R3) in figure (c) above to indicate its rather poorly understood nature – is that it must be capable of mapping between simple acoustic encodings in early auditory cortical areas to speech primitives in the superior temporal gyrus (STG) and superior temporal sulcus (STS). Note here the reference to temporal levels. Poeppeal and Idsardi (2011) are referring to the fact that the brain simultaneously creates chunks of the stimuli that, on a linear beads-on-string structure, corresponds to linguistically relevant units. The authors propose the segment (10-80 ms) and the syllable (100-500 ms) as the two most relevant windows [7]. Poeppel and Idsardi (2011) refer to this as the "phonological primal sketch". The authors hypothesize that this might involve locating landmarks, such as those posited by the PFNA (Plosives-Fricatives-Nasals-Approximants) coarse-coading hypothesis. Briefly, this hypothesis claims that initially the signal is categorized into one of these four broad categories – plosives, fricatives, nasals and approximants. Crucially, these are all features that are articulator-free (Halle, 2005; Halle, Vaux, and Wolfe, 2000) – they are not bound to specific articulators but rather encode broad degrees of constrictions of the articulatory and phonation channels that give sound categories their distinctive signatures. For instance, plosives or stops are articulated with complete constriction that leads to an explosive ejection of air, fricatives with constrictions sufficiently narrow to generate turbulent noise,

---

[7] Note here that in this sense the word syllable refers to a flat sub-chunk of the stream that, post transduction, will correspond to a sub-string of the entire string. Whether this necessitates traditional hierarchical syllabic representations is a separate empirical issue.

nasals with airflow through the nasal cavity, and approximants with virtually no constriction to block airflow. Based on these major landmarks in the signal dedicated sub-routines specific to each of these major classes then determine the specific articulator bound features. The authors point out that the neurons equipped with these specific detectors are good-candidates for "hardwired circuits", and they are likely to be useful in other ecological contexts such as avoiding predators. Once the sub-routines have finished computing the bound features, we arrive at the final representation (R4 in d) which serves a bi-directional purpose. It is, both, the end point of perception and the input to future production.

Poeppel, Idsardi, and Van Wassenhove, 2008 note that the entire process is probabilistic, and the feature detectors are prone to returning mismatching values as the neural mechanisms attempt to resynthesize feature values from information in the signal. For instance, they point out that in creating the primal sketch the various "feature detectors" in the brain are not flawless and return probabilistic values.

Ex 2.3.3. H: Hypotheses, E:Evidence, p: Probability
$p(\mathrm{H}|\mathrm{E}) = p(\mathrm{E}|\mathrm{H})$ x $p(\mathrm{H})/p(\mathrm{E})$

In the equation above $p(\mathrm{H}|\mathrm{E})$ denotes the probability of the analysis concerned, and $p(\mathrm{E}|\mathrm{H})$ denotes the probability of the desired synthesis given the data. Based on this initial sketch, largely consisting of probabilistic guesses about the distribution of major class, articulator-free features in the signal, articulator-bound features (±Round, ±ATR etc.) are identified in a context-sensitive fashion. Mathematically, Poeppel et al. (2008) argue that this stage involves evaluating conditional probabilities of the sort $p([\text{+Nasal}]|[\text{Labial}])$, because detection of [Labial] and [Coronal] in [+Nasal] is different from, say, in stops in a languages that lacks velar nasals but contrasts all three positions with stops.

The mechanisms underlying the mapping of articulator configurations to acoustic maps, thus, involve a sophisticated interplay between sensory input, motor control, and learning processes.

At a fundamental level, the auditory system plays a crucial role in encoding acoustic signals, allowing individuals to perceive and analyze the shape of the received signal. Simultaneously, motor control circuits, particularly those involved in speech production, are responsible for orchestrating the precise movements of the articulators such as the tongue, lips, and vocal folds to produce speech sounds. Studies utilizing neuroimaging techniques have identified brain regions associated with vocal learning and imitation, such as the anterior forebrain pathway in songbirds and the mirror neuron system in primates (Petkov and Jarvis, 2012; Arbib, 2005). In humans, neuroimaging studies have implicated regions such as the inferior frontal gyrus and the superior temporal sulcus in vocal imitation tasks (Iacoboni and Dapretto, 2006). Research in developmental psychology and neuroscience, likewise, that during early infancy, there is a period of heightened plasticity in the brain, known as the *critical period*, during which these sensory-motor mappings are refined through experience and exposure to language input ((Kuhl, 2010). Furthermore, studies with animal models have identified specific neural circuits, such as the basal ganglia-thalamocortical loops, involved in sensorimotor integration and learning of vocalizations (Doupe and Kuhl, 1999). The innate mechanisms that govern the mapping of articulator configurations to acoustic maps involve a complex interplay of genetic predispositions, sensory experience, and intricate somato-sensory maps (Brecht, 2017), shaping individuals' ability to mimic and produce a diverse range of speech sounds essential for communication.

The human brain, like most other mammalian brains, is endowed with intricate somatosensory maps (Brecht, 2017). These maps, found in the brain's somatosensory cortex, represent the body's sensory surface, including the tongue, lips, and other articulators involved in speech production. When speakers articulate speech sounds, sensory feedback from these articulators is continuously monitored and processed by the brain to ensure accuracy and coordination in movements. This feedback loop, known as proprioception, allows the speaker to adjust and refine their articulatory gestures in real-time based on the sensations received from the articulators. Mammals, including humans, also possess a remarkable ability to instinctively configure these articulators to mimic perceived sounds, a phenomenon crucial for language acquisition and communication. This skill is evident early in human development, as infants begin to imitate the sounds they hear in their environment, laying the foundation for language acquisition. Research by Kuhl and Meltzoff, 1996 demonstrated infants' ability to mimic speech sounds. Studies with non-human mammals such as dolphins and parrots have shown similar abilities, indicating that this skill is not exclusive to humans. For instance, dolphins have been observed imitating human speech patterns, suggesting a capacity for vocal learning beyond primates (Reiss and McCowan, 1993). In the context of analysis by synthesis, as discussed by Poeppel, Idsardi, and Van Wassenhove, 2008 and Bever and Poeppel, 2010, it seems far more reasonable to assume that neural mechanisms for signal processing shared across mammalian and avian brains are utilized in combination with knowledge of the articulatory channel to create linguistically relevant symbolic markers. As discussed above, this process is not deterministic and returns probabilistic values, likely creating a limited range of acceptable values, a mass of probabilities of sorts that are linked to every symbol. Once established, a feat achieved during the critical period, these markers, now phonological features, can then be utilized by the phonological module in a manner that enables generative use.

2.4 Conclusion

The proposal presented in this chapter is not meant to be theory of feature implementation in a strict sense. Rather, my purpose here has been to show that claims to determinism made by CP appear to be far less supported in the empirical literature than a weak sense of feature construction. I wish to stress that by construction by no means do I mean that the neonate is phonological blank slate, starting out by making random predictions until a dart hits the bull's eye. Rather, what I have in mind is a proposal similar to that made by Dresher, 2014 and Samuels, 2015a, where the child is innately endowed with a sense of possible means to instantiate phonological features. Put another way, the child has its disposal a limited range of articulator configurations, a finite set by all means, and an instinct for the resulting acoustic maps. The role of experience, then, is to afford the child enough data points to narrow the mass of probabilities to small enough range such that Bayesian methods can reliably map systemic regularities in the speech signal to specific articulator configurations. In other words, there must exist, a priori, the necessary abilities for, say lip rounding, and an instinct that this results in a general downward sweep across the spectral range of the resulting acoustic map. The role of the critical period is merely to determine the range of variations that are permissible in any given inventory, thus establishing a partial veridicality between features and their interpretations. The approach to the lower interface advocated here flows naturally from the substance-free view of what phonology is – a purely algebraic, symbolic affair, devoid of all notions of substance.

Chapter 3

Partial Veridicality at the Lower Interface

3.1 Introduction

This chapter, based on a forthcoming article in the *special issue of Linguistic Society of India,* explores the architecture of the phonology-phonetics interface in light of the general nature of interface theories in an overarching modular approach to cognitive science. The resurgence of the substance-free program (SFP) in phonological theory (Fudge, 1967; Hale and Reiss, 2008; Chabot, 2022b; Chabot, 2023) has reinvigorated the old debate regarding the nature of the relationship that holds between the primes of phonological form (distinctive features) and the phonetic shape that obtains from them in the process of EXT(ernalization). Early structuralist thinking maintained a strong bulkhead between form – the abstract symbolic representations of linguistic knowledge in the mind – and substance – the physical shape and properties of speech sounds attested in individual tokens (Courtenay, 1870[1972], 211ff). Thus, for de Courtenay the phoneme was a symbolic marker, an abstract encoding of only those aspects of a speech sound that were linguistically relevant, while abtracting away from other perceptible physical qualities. This is an extension of the Sausseurian distinction between langue and parole (Saussure, 1916[1967]), where facts about the organization of langue are only partially reflected in the parole. The phoneme, for structuralists, belongs to the domain of langue, where it stood in a relation of oppositions with all the other phonemes in that system, and it is this network of oppositions that constitute the core concern of phonology, one that cannot be described in purely physical terms – dans la langue il n'y a que des différences (Sausseur,1916 [1967]:166). Roman Jakobson's decomposition of the phoneme into sub-phonemic distinctive features (Jakobson, 1949) marks a stark shift away from the structuralist zeitgeist as the content of features is now conceived of in terms of the acoustic and articulatory correlates of the linguistic system of oppositions. This position is taken up and further developed by Chomsky and Halle (1968) in The Sound Patterns of English, thus opening the gates for phonetic substance to be imbued into phonological forms, much of which was justified in terms of overt efforts to restrain the overgenerating capacities of a purely formal system. Chabot (2022) points out that this substantive trend has characterized virtually every theory of phonological representations post-SPE, steadily and gradually dismantling the structuralist distinction between form and substance and rendering dormant the many issues, often considered "settled" without much justification (Chabot, 2022b, 430ff.), that concern the status of features as algebraic computational symbols and their mappings onto continuously varying phonetic realizations. The core concern of the re-emerging substance-free program, explicitly neo-Sausseurian (Hale & Reiss, 2008) in its approach, is to re-examine the basic properties of phonological primes that set them apart from phonetic substance, and to offer up biolinguistically plausible explanatory accounts of, both, the mapping to phonetic substance and the role played by UG in this regard. In what follows I examine how this latter aspect has become the central fissure that divides the substance free program into two camps, one arguing for completely arbitrary relationship between primes and substance (radical-SFP) (Blaho, 2008; Scheer, 2010a; Dresher, 2014; Chabot, 2021) and the other maintaining strict deterministic veridicality (SFP) (typically known as *the Concordia school of SFP* (Volenec and Reiss, 2017; Reiss and Volenec, 2022a)) and attempt to sketch a hybrid framework that attempts a re-unification at the cost of partial veridicality.

3.2 Modularity in generative grammar

Since its inception the generative enterprise has taken a modular approach to theory building (Scheer, 2010a). Modularity, in its most widely discussed form, was formally introduced by Fodor, 1983 in his discussion of the architecture of the human mind. Fodor posits that the human mind is composed of specialized cognitive modules, each responsible for processing specific types of information or performing particular tasks. These modules operate independently of each other, with dedicated mechanisms for input, processing, and output, and they are characterized by their domain specificity, information encapsulation, automaticity, and fast processing speed. According to Fodor, modularity serves as an efficient and adaptive way for the mind to process information, allowing for rapid and reliable responses to environmental stimuli.

In the context of the phonology-phonetics interface, aspects of modular thought was visible in the early structuralist (Saussure, 1916[1967]; Sapir, 1925) work where a strict bulkhead was maintained between form and substance, between *langue* (abstract, mind-internal) and *parole* (substantive, external). The re-emerging substance-free program in phonology (Fudge, 1967; Hale and Reiss, 2008; Blaho, 2008; Reiss, 2017; Chabot, 2021; Scheer, 2022) explictily argues in favor of a modular approach. Proponents suggest that the interface between phonology (the abstract cognitive representation of speech sounds) and phonetics (the physical realization of speech sounds) operates through distinct modular mechanisms. The central idea in SFP is that the phonological module processes linguistic structure independently of phonetic detail, relying on abstract representations to encode language-specific patterns and rules. In contrast, the phonetic module handles the articulatory and acoustic properties of speech sounds, translating phonological representations into physical speech signals.

By adopting a modular perspective, SFP proposes that the phonology-phonetics interface can be understood as a system of interaction between specialized cognitive modules, each with its own processing principles and constraints. This view allows for a more nuanced understanding of how linguistic structure interfaces with the physical properties of speech, highlighting the role of domain-specific mechanisms in mediating between abstract phonological representations and their phonetic realizations. This section, and subsections herein, provides an overview of the modular approach as it has been discussed within SFP, and subsequent sections then elaborate on a novel framework arguing that partial-veridicality, rather than complete arbitrariness or determinism, holds between phonological primes and phonetic substance.

3.2.1 Primes in substance-free phonology

Phonological primes, more commonly called distinctive features, have been the staple of phonological theory ever since (Jakobson, 1949) decomposition of the structuralist notion of the phoneme into sub-phonemic units. As mentioned before, for the structuralists the phoneme was not just a speech sound, for this implies a physical entity with a variety of analog properties many of which do not play any linguistically significant role. Rather, a phoneme was solely defined in terms of the relationship in which it stood with respect to all the other phonemes in a given language. It was a symbolic counter of contrast in a system of contrastive oppositions (Courtenay, 1870[1972], 211ff.). Jakobson's work was the first to mark a stark shift away from this purely formal conceptualization while simultaneously decomposing the phoneme into sub-phonemic features. Crucially, for Jakobson (1949) the binary oppositions marked by the distinctive features were not derived solely from a system of contrastive relationships like the structuralist phoneme was, rather they were based on the specific articulatory and acoustic correlates of those contrasts. The Sound Patterns of English (SPE; Chomsky and Halle, 1968) builds on and extends this Jakobsonian approach to phonology, adopting both the sub-phonemic zeitgeist and the substantively laden features. While SPE attempted to replace many of Jakobson's acoustic features with the now familiar articulatory features, the reliance on substance remained intact. Indeed, Halle, 2005, p. 28 writes:

> A major consequence of Jakobson's 1928 proposals was that phonetics and phonology were seen as a single field of study: that is, it was recognized that not only do phonetic facts shed important light on the phonology of a language, but also facts of phonology provide insights into the nature of phonetics.

In SPE these phonetically motivated features plays three important roles. First, they mark binary contrastive oppositions within the inventory of a language. Second, the generalized intersection of shared features – i.e. the set of feature shared by certain segments to the exclusion of all others in the same inventory – by a group of segments creates natural classes, in terms of which rules of alternations are defined. But perhaps most importantly, features serve as encoded instructions for the phonetic systems that are tasked with perception and production of utterances. They do this by virtue of the fact that, unlike the structuralist phoneme, SPE features are not just symbolic counters but also possess inherent phonetic content. In the famous Chapter 9 of SPE, Chomsky and Halle, 1968, p. 400 bemoan the fact that the system of computations they have developed have, perhaps, become too powerful, generating patterns that lack any basis in reality.

> "The problem is that our approach to features, to rules and to evaluation has been overly formal. Suppose, for example, that we were systematically to interchange features or to replace [αF] by [$-\alpha$F] (where $\alpha = +$, and F is a feature) throughout our description of English structure. There is nothing in our account of linguistic theory to indicate that the result would be the description of a system that violates certain principles governing human languages. To the extent that this is true, we have failed to formulate the principles of linguistic theory, of universal grammar, in a satisfactory manner. In particular, we have not made use of the fact that the features have intrinsic content."

The imbuing of features with substantive information has the direct consequence of allowing phonological computations direct access to phonetic motivations, which can now be utilized to restrain their generative capacity. Chabot, 2021 and Scheer, 2010a provide book-length evaluations pointing out that this trend came to characterize virtually every post-SPE theory of phonology both within (Archangeli and Pulleyblank, 1994; Prince and Smolensky, 2004) and outside(Donegan and Stampe, 1979) the generative program. Much of this, Chabot, 2022b points out, is done through a re-interpretation of the notion of markedness inherited from the Prague school. The original notion of markedness reflected facts about individual languages, while in SPE terms it was recast in a more universal fashion – of the two values of features, ±F, one value is considered marked. Grammars are, then, sensitive to markedness preferring to maximize unmarked patterns. The resulting system, then, builds in facts about human articulatory and perceptual biology directly into the phonological module. This is, perhaps, most iconically reflected in Sagey, 1986 feature geometric representations of feature hierarchies that directly reflects the organization of the articulatory channel.

The importing of substance into the formal grammar undoubtedly leads to a more tighter theory, of sorts, since markedness and phonetic concerns can now rule out patterns that are otherwise computable by (universal) grammar alone (Hale and Reiss, 2008). However, whether this also amounts to a more explanatory theory is contested by those that demand a substance-free phonology (SFP). While Hale and Reiss, 2000a provide the first of the more recent attempts to reinstate the structuralist bulkhead between phonetics and phonology, as arguments for substance-freedom can be traced as far back as Fudge, 1967 proclamation that phonologists, “and above all generative phonologists”, ought to burn their phonetic boats and turn to a truly abstract framework. Fudge, 1967, p. 1, reviewing alternations from Tswana and attempting a featural analyses, writes:

> The facts of Tswana phonology [...] proved especially awkward to handle; when the original twelve features were taken and assigned in a ’classical’ manner (i.e.

> [+Vocalic, +Consonantal] for all Liquids, [-Vocalic, -Consonantal] for all Glides, etc.), there were found to be many different ways of characterizing the various segment-types. What was disturbing was that every one of these ways entailed writing phonological rules which failed to highlight the underlying structure, or even obscured it. Conversely, if assignments other than the 'classical' were permitted, any attempt to group the segment-types into an arrangement which faithfully reflected the relationships between them [...] left one with insuperable phonetic problems of feature-assignment

While Fudge, 1967 criticism is squarely aimed at the substantive content of features, Chomsky and Halle, 1968, p. 428 note yet another issue with substance in phonology, again in the now famous Chapter 9. Reviewing the process of velar softening in English, particularly the second stages of the rule, they note a problem with allowing substance to dictate the structure of rule syntax:

> It does not seem likely that an elaboration of the theory along the lines just reviewed will allow us to dispense with phonological rules that change features fairly freely [...]

A well-known phenomena that creates similar problems for phonetically driven notions of natural classes, and phonetically motivated rules, is found in what has come to be known as crazy rules (Hyman, 1985; Scheer, 2009). Briefly, a crazy rule [1] is any pattern of alternation that makes little to no sense when viewed from the perspective of phonetic naturalness. Chabot, 2021, p. 85 aptly points out that what constitutes a crazy rule depends greatly on what one allows within the scope of phonological theory, especially since phonetic naturalness itself defies any objective quantification. What constitutes natural is typically in the eye of the beholder,

---

[1] One crucial distinction with the Concordia school of SFP needs to be noted here. Bale and Reiss, 2018 argue that a rule must be stated in terms of natural classes. This being the definition of what makes an alternation a rule-driven phenomena, in the Concordia view so-called *crazy rules* are not rules at all. In this view alternations classified as crazy cannot be viewed as coherent rule-driven phonological processes.

although an argument can be made that within a phonetically motivated phonology a rule can be crazy by at least one of four criteria: (a) a rule that accepts a crazy class of inputs, (b) a rule defined around a crazy set of triggers, (c) a rule that implements a phonetically nonsensical structural change, or (d) a rule with no phonetically sensible relationship between its structural description and the mapping from the input to the output. For instance, Chabot, 2021, p. 87 cites that in Teton Dakota an underlying crazy input class, /p,l/, surface as /m,n/ when they occur immediately preceding a nasal vowel (/p,l/ → /m,n/ /_ [+Syl, -Cons, +Nasal]). This particular rule is crazy because there is no phonetic feature that can create a class out of /p/ and /l/, while excluding every other segment in the language. Crucial for the present discussion, however, is the fact that crazy rules are attested across languages (Chabot, 2021). If phonology is at all capable of representing and computing crazy rules then clearly phonetic motivations cannot be a core guiding principle of its architecture. In such a case, explanations for the fact that such rules are relatively rare, as also the fact that a large number of phonological patterns do make phonetic sense, must be sought elsewhere, outside of phonology proper.

SFP, at its core, seeks to apply the minimalist principle (Chomsky, 2014) of attributing to the phonological grammar only those aspects of the sound patterns of languages that cannot be accounted for elsewhere. It posits the minimum number of primitives required to account for the knowledge that native speakers have of their language(s), seeks to expunge from phonology all facts about typology and functional considerations, which are ascribed to extra-phonological domains. Proponents of substance-freedom are strictly neo-Saussurean in viewing the primes of phonological representation as mere algebraic counters of contrast, and rules as purely formal symbol manipulating mapping processes that are completely blind to substantive considerations. While there exists a general consensus among proponents that functional phonetic facts are not within the purview of phonological theory, there is broad disagreement about how substance-freedom is to be attained. The program can be broadly divided into two schools of thought. On one hand the so-called radical school (Blaho, 2008; Scheer, 2010a; Dresher, 2014;

Chabot, 2021; Chabot, 2022b) insist that both features and rules are completely blind to substance. Particularly, the radical school is generally of the belief that features and their phonetic readings exhibit the same arbitrariness that is found between morphosyntactic terminal nodes and the phonological shape of the vocabulary items that are inserted into them during spell-out (Scheer, 2010a; Chabot, 2021). A direct consequence of this view is that features are constructed, as opposed to being innate, from experience during the critical period and traditional feature-labels, a marker of their intrinsic substantive content in SPE, is relegated to the status of tools of methodological convenience, devoid of any specific ontological status. On the other hand, the Concordia school (Hale and Reiss, 2008; Volenec and Reiss, 2017) maintains similar commitment to substance free rules, but have been less consistent in their view of features. In the original conception of Hale and Reiss (2008), features retain their "intrinsic content", just like SPE, but rules are claimed to be completely blind to this content which only becomes visible in the phonetics. This position is later changed by Volenec and Reiss, (2017) who claim that features, too, are substance-free, however still insisting that the phonetic readings of features is deterministic and universal. That is, both features and their readings are innate, as evidenced by the retaining of traditional feature labels. Volenec and Reiss (2017) propose a set of algorithms – syntagmatic transduction algorithm (STA) and paradigmatic transduction algorithm (PTA) – that map each feature onto fixed, predetermined substantive correlates, generating a phonetic representation that is fed into the sensory and motor systems. Performative factors are then claimed to occlude these deterministic readings by introducing co-articulatory variations. This transductive algorithm is claimed to work along Lenneberg, 1967's neuromuscular schema for spatio-temporal co-ordination of articulators, but the authors provide very little discussion of their inner workings. The end-effect, I will argue below, is that Volenec and Reiss' (2017) algorithms provide very little in the way of explanatory value, and simply attempts to push the "intrinsic content" of SPE features one level outward. In other words, Volenec and Reiss (2017), in my opinion, retain the SPE notion of "intrinsic content" of features, but simply avoid calling it that.

3.2.2 Substance in substance-free phonology

Any theory of phonology, no matter how substance-free, must eventually make explicit statements about substance. There are, at least, two reasons why this must be so. First, phonological forms are eventually fed into the sensory-motor systems for EXT, without which communication is impossible. While it is, in principle, possible to argue that phonology itself is not meant for communication (Hale and Reiss, 2008; Reiss, 2017) it remains indubitable that phonological grammars have non-trivial impact on communication through both vocal (Cutler, 1994) and visual (Stokoe, 1980; Sandler, 2014) mediums. If communication is to be systematic then there must be systemic ways of mapping from phonological form to phonetic substance. If there is no substance in form, then form must be mapped onto substance somehow, and this mapping must be systematic at least on a language-specific basis. Second, any survey of the sound patterns of the vocal languages would illustrate that a great many phonological patterns hew rather close to the substantive bone, and are amenable to acoustic and/or articulatory explanations. Within the broader umbrella of the substance-free program there is generally a consensus that the latter aspect falls outside the purview of phonological grammar, and are better explained through third factors (Chomsky, 2005; Reiss, 2017; Chabot, 2023). Briefly, Chomsky (2005) argues that a general application of Occam's Razor to develop a parsimonious and minimal account of universal grammar (UG) leads to the differentiation between three different factors in the overall architecture of natural language – (1) factors that are specific to language, and are part of a genetically endowed UG, (2) experience with a natural language during the critical period that helps the transformation of the initial state of UG (So) to an individual's specific state of I-language, and (3) factors deriving from the general rules of physics, physiology etc. that are not specific to language. The general substance-free approach to the so-called prevalence of phonetically natural processes in phonology involves appeals to these third-factors. Moreton, 2008 provides a well articulated example of third-factor explanations in the form of channel bias – systematic phonetically motivated errors that derive from human perceptual and articulatory biology and serve as precursors to phonologization (Hyman, 2013). For instance, in Hindi during the production of a distinctively nasalized vowel immediately followed by a voiced stop, the stop is pre-nasalized. This effect is only observed with voiced stops, and Ohala and Ohala, 1991 argue that this epenthetic nasal has a purely phonetic origin since:

> [. . .] among the auditory cues for a voiced stop there must be a spectral and amplitude discontinuity with respect to neighboring sonorants (if any), low amplitude voicing during its closure, and termination is burst; these requirements are still met even with velic leakage during the first part of the stop as long as the velic valve is closed just before the release and pressure is allowed to build up behind the closure. However, voiceless stops have less tolerance for such leakage because any nasal sound – voiced or voiceless – would undercut either their stop or their voiceless character.

Within Moreton's (2008) channel bias based account such automatic phonetic reflexes are often exaggerated and misinterpreted by speakers as being phonological distinctive. Over generations such processes are more likely to be phonologized, thus impacting the lexical representations of speakers diachronically. Such accounts provide an explanation of the abundance of phonetically sensible patterns in phonology without importing functional and phonetic facts into the grammar itself. In other words, there is nothing in phonology itself that prefers an epenthetic nasal in between a nasalized vowel and voiced stop in Hindi. Chabot (2021: 159ff.) argues that once such patterns pass the phonetic veil and becomes phonologized the phonetic precursors are then reanalyzed by the learners as contextual allophonic variation, stated explicitly in terms of distinctive features and implemented by extrinsically ordered rules. However, since the intrinsic biological aspects of perception and articulation are inherently phonetically natural the rules themselves reflect the phonetic motivations from their points of inception. The fact that the grammar itself is not natural in any coherent sense is reinforced by the existence of crazy rules. In fact, within a purely substance-free view crazy rules are not so much crazy, as they are rare. (Chabot, 2021, p. 161), thus, argues that crazy rules are not born crazy, but become so through similar diachronic processes (emphasis added):

> Crazy rules are derived from phonetically plausible rules by a sequence of individual diachronic steps, each of which may be phonetically natural but together can result in unnatural patterns in the grammar. Since phonology is isolated from phonetic influence, once phonologized, the structural description of a phonological rule is no longer sensitive to the phonetic context at its origin—rules are no longer yoked to phonetic exigencies. If a further change to the target, trigger, or conditioning environment occurs, the rule may loose its natural description while continuing to operate.

Crucial to this view of crazy rules and diachronic impact of substance is a return to the structuralist separation form and substance in contemporary SFP, in the form of Fodorian modularity. This issue is directly related to the mapping problem discussed before, and also appears to be the primary point of disagreement among proponents of substance-freedom. Since its inception, the generative program has been modular in nature (Scheer, 2010a). This is reflected, for instance, in the familiar inverted-T schema from the early days of generative grammar, where the surface structure splits into two modules, the Phonological/Phonetic Form (PF) and the Logical Form (LF), each of which is opaque to the other. The mutual opacity of modules is a characteristic feature of the modularity of mind hypothesis (Fodor, 1983), the idea that the cognitive capacities of the human mind are made of independent modules, each specialized for a particular type of task and incapable of everything else. Fodor's modularity hypothesis proposes a number of identifying qualities of a cognitive module, but for the present purposes the most important ones are domain specificity/proprietary vocabulary, informational encapsulation and inter-modular communication through interfaces. Briefly, domain specificity/proprietary vocabulary implies that a module is designed for a very specific computational task and can compute over information stated explicitly in terms of its own proprietary representational vocabulary. This idea is not new to linguistic theory where generally it is accepted that morphosyntax and phonology are distinct modules, and information in each is represented in terms of their respective vocabulary. There is no [+Voice] in morphosyntax, and syntactic structures are only useful to phonology post-spell out. These systems are mutually unintelligible due to the distinct vocabularies they process. The substance-free program in general extends this modular approach to phonology and phonetics based on the observation that just like syntax and phonology, phonetics and phonology also make use of different conceptual vocabularies. Namely, phonetics is concerned with the mind-external world and physiological phenomena that are continuous and analog in nature, while the vocabulary of phonology is completely discrete – a segment is either [+Continuant] or it is not, but never [2.573 Continuant], for instance. Modules are, thus, informationally encapsulated and mutually incommunicado, because of their proprietary vocabulary.

Module	Vocabulary	Type
Morphosyntax	NP, VP, Aspect, Tense, SG/PL etc.	Algebraic/Discrete/Digital
Phonology	Voice, Labial, Coronal, Continuant, Dorsal, High, Low etc.	Algebraic/Discrete/Digital
Phonetics	VOT, Harmonics-to-Noise Ratio, Amplitude, Duration etc.	Conitnuous/Analog

TABLE 3.1: Modules and vocabularies

One consequence of modules being mutually incommunicado is the need for interface driven transductions, where a transductive intermodular interface is tasked with format-conversion, converting information encoded in the vocabulary of one module – the output of this module and the input to the transduction process – into the vocabulary of another module – the output of the transduction process and the input to the second module. Thus, for instance, morphosyntax generates a tree stated explicitly in its own vocabulary consisting of non-terminal nodes (DP, TP, VP, V, v etc.) and terminal nodes where vocabulary items are inserted. These vocabulary items, as opposed to syntactic constituents (VP, DP etc.), are endowed with a phonological form (the upper-interface of phonology with morphosyntax). The upper interface of phonology with morphosyntax has two distinct qualities. First, its operation is not computational in any meaningful sense, but rather it makes use of a list-based look-up system where each morphosyntactic feature is listed in a spell-out lexicon, paired with a vocabulary item. Second, the spell-out process is completely arbitrary (Scheer, 2010a), at least from a synchronic point of view, and morphosyntactic features do not exhibit any systematic relationship with the shape of the vocabulary items that are inserted into the terminal nodes of syntactic trees.

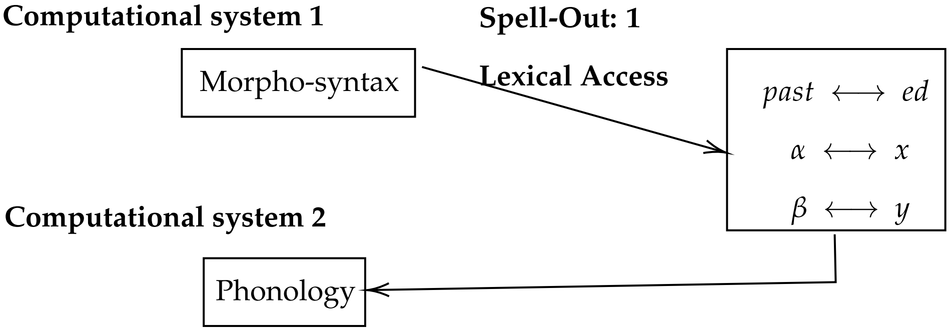


FIGURE 3.1: The upper-interface. From Scheer, 2010a

If phonological forms are substance-free then these forms, by themselves, are equally illegible to the sensory and motor systems that are tasked with production and perception of speech signals. The interface with phonetics (the lower-interface of phonology) is responsible for mapping these phonological forms onto their corresponding phonetic shapes (in the direction of production), and the other way round (for perception). Within the substance-free program there have been two broad proposals for handling the lower-interface. For the Concordia school (Hale and Reiss, 2008; Volenec and Reiss, 2017) the interface consists of a set of transducers. The paradigmatic transduction algorithm (PTA) is responsible for assigning a set of muscles to each feature which in turn are responsible for articulator configuration. The syntagmatic transduction algorithm (STA) is claimed to be responsible for temporal co-ordination of muscular activity. Crucially, Volenec and Reiss (2017) insist that the functioning of the PTA and STA are completely deterministic. For them, thus, the output of this transduction process is predictable, given a specific feature identity. In other words, while phonological rules enjoy a wide latitude in manipulating symbols arbitrarily, the transduction to phonetic representations is fixed and invariant.

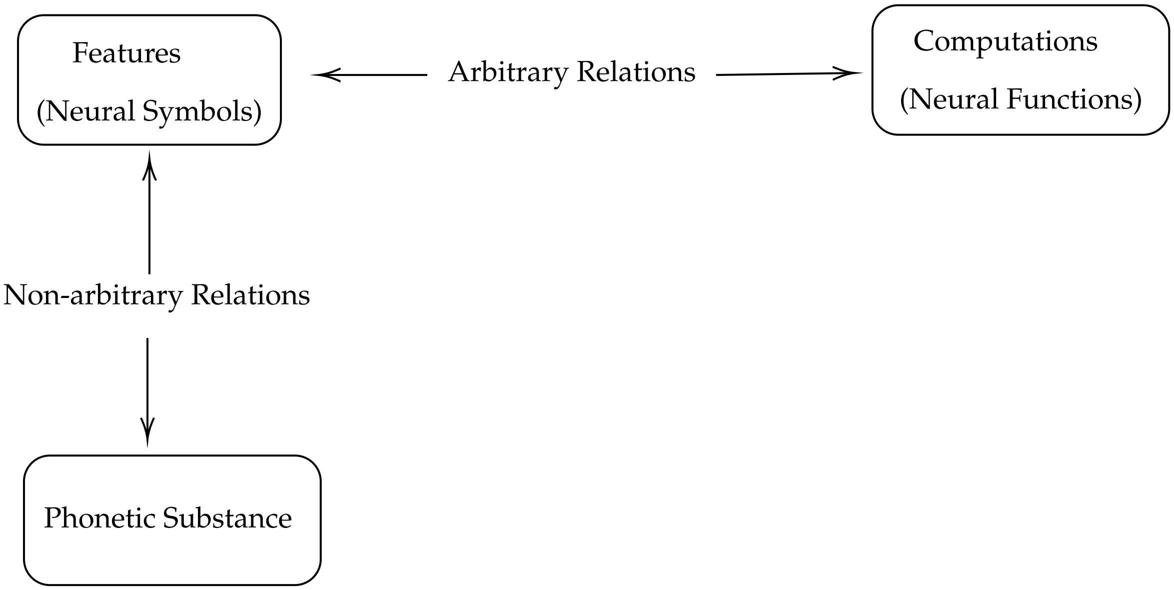


FIGURE 3.2: Arbitrary computations, deterministic transductions

The lack of invariance problem – the fact that phonetic realizations of specific phonological forms are rarely (if ever) identical – is then accounted for by what the authors call intersegmental and intrasegmental co-articulation. Briefly, intersegmental co-articulation refers to the fact that features in surrounding segments affect the phonetic realization of any given feature in any given segment. Likewise, intrasegmental co-articulation refers to the fact that lip rounding, for instance, implemented by the feature [+Round], varies depending on whether the segment concerned is specified for, say, [+High] or [-High]. In Volenec and Reiss, 2017, 257ff.'s rather sparse description of this process, "transduced features often 'spill over' these temporal borders, crossing segmental and sometimes even syllabic boundaries in both directions, thus leading to coarticulation". Volenec and Reiss repeatedly refer to Lenneberg, 1967, Ch.3's neuromuscular schema as the neurobiological grounding for their transductive algorithms. However, the motivations behind this reference is unclear to me, particularly because Lenneberg (1967) is quite clear about the fact that the largely antagonistic design of the muscular system implies that the spatio-temporal co-ordination of muscle activations during articulation take the form of automatic synergisms that are largely probabilistic in nature. Volenec and Reiss (2017), on the other hand, are quite insistent on their algorithms being deterministic. It is not hard to see, of course, that the hypothetical PTA should reliably activate the same set of muscles for, say, lip rounding. However, the spatio-temporal co-ordination that the STA is tasked with is highly unlikely to be deterministic in its functioning. By Volenec and Reiss' (2017) own account, the intra-segmental specification of all the other features that any given feature, say [+Round], is packed with in a segment determines the extent and nature of lip rounding that is implemented.

Volenec and Reiss (2017) further argue that their proposal adheres to strict modularity as neither the transducers nor the grammar can look into each other. This, however, leaves aside two issues. First, there is the issue of whether these transduction algorithms are modular, and if they are whether PTA and STA are distinct modules? The authors provide no discussion of this issue. Generally, in modular architectures computations are performed by modules which are informationally encapsulated and operate over proprietary vocabularies. Second, this conception of transduction appears to violate at least one condition of Fodorian modularity – namely, proprietary vocabulary. Volenec and Reiss' transducers, by their own account, accepts a digital input and produces an analog output, and vice versa in the direction of perception. This is either an oversight that consists of gross violation of modularity – given that the transducers appear to be compatible with different vocabulary items (substance-free phonological primes and substantive phonetic representations) or there is an implicit assumption that transducers are not modular. This latter case, if intended, will constitute a rather unique claim, though not necessarily an implausible one. To my knowledge there are only two broad approaches to modularity – the standard Fodorian modularity holds that only input systems (which transducers should count as, and thus be subjected to the proprietary vocabulary requirement) and the evolutionary psychology position of massive modularity (Plotkin, 1997; Pinker, 2005) that views all aspects of the mind being modular. Further, the requirement of proprietary vocabulary holds even if one adopts massive modularity. It is, of course, possible that these two phonology-phonetic transducers are conceived of in the second sense, and as modules whose vocabulary consists of both phonological and phonetic primes, but no such claims are made explicit by the authors.

In contrast, (Scheer, 2010a) argues that if the mind is modular then all intermodular interfaces should adhere to a similar architecture. The similarity stems from the fact that they are sub-components of a larger module (the faculty of language), but they are not quite identical as different sub-modules compute over different types of primal vocabulary. In this view the architecture of the intermodular interface remains the same at both the upper and lower interfaces of phonology. Just as a lexical spell-out inserts vocabulary items into the terminal nodes of the trees that morphosyntactic computations generate, a post-phonological spell-out system maps phonological features onto language-specific substantive correlates by looking up associated values. Since this process is not computational, there is no violation of modularity. In this latter view features and their phonetic realizations are completely arbitrary, and there is nothing inherently lips-related about the feature [+Round] for instance. (Chabot, 2021) further points out that in this radically substance-free view feature-labels (e.g. Labial, Coronal etc.) are redundant, and can be replaced with random variables, such as α or β. The figure below illustrates the radical approach, where the acquisition process involves the creation of a post-phonological spell-out lexicon where phonological primes are represented by random variables associated with language-specific substantive correlates. (Dresher, 2014), for instance, proposes the successive division algorithm, where the learner successively divides the inventory, based solely on phonological activity of features, until all available contrasts have been exhausted. The outcome is an arbitrary list of symbol-substance pairs that are emergent and language-specific. Since not all phonological contrasts are active in all languages in this system there is no systematic correspondence between the phonetic qualities of a segment, and the phonological features it bears.

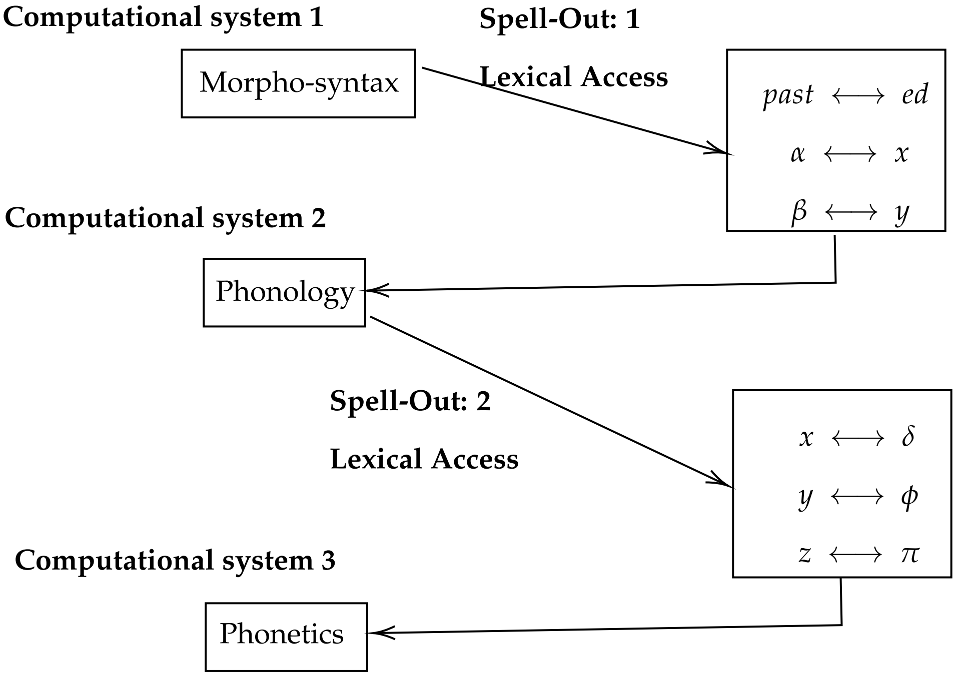


FIGURE 3.3: Upper and lower interfaces of phonology. From Scheer, 2010a

In the next section I attempt to sketch a hybrid approach that is largely derived by retaining what I take to be the core arguments of each school. My primary concern is with the Concordia school's notion of deterministic transduction (Volenec and Reiss, 2017), which appears to lack any empirical support, besides being ambiguous. However, I will also attempt to retain a weaker notion of a priori available sense of the set of possible features, and insist on at least partial veridicality holding between features and their phonetic functions. I borrow primarily from Poeppel and Idsardi, 2011 perception-action-memory loop based internal forward model for probabilistic feature-detection in perception, and suggest that the same mechanism can be reliably used during the critical period to combine an innate knowledge of the human articulatory system and an innate analyses-by-synthesis (AxS) processing of speech signals to create a language-specific post-phonological spell-out lexicon that associates every active feature in an inventory with a probabilistic mass of phonetic values.

3.3 The lower interface

This section proposes a hybrid framework for the lower interface. It is hybrid in the sense that it retains a weak notion of an innate *sense for* features, though to what extent this can be called features *qua* features remains dubious, and advocates for partial veridicality between phonological features and phonetic substance. Unlike the radical school, the framework proposed argues that there exist, a priori, a sense for a feature that, say, causes the rounding of lips, and thus a general downward sweep across the spectral range. But also unlike the Concordia school this framework eschews the notion that such a feature is innately linked to some elusive, never-to-be-captured, point-value substantive reading. Rather, what I have in mind is an instinct for articulator configurations, accompanying acoustic targets, and the combination of the same with categorical perception (Kuhl and Miller, 1978; Kuhl, 1991) and innate neural mechanisms (Poeppel, Idsardi, and Van Wassenhove, 2008) to create symbolic linguistic contrasts during the critical period. I first introduce the two main concepts that I appeal to – analysis by synthesis (AxS) and multiplexing, and then proceed to discuss the details of my proposal.

3.3.1 Multiplexing

Multi-time resolution processing refers to the idea that complex forms are processed by the brain at different temporal window sizes in parallel. While this idea is usually discussed in the context of neural oscillations (Ward, 2003), Poeppel et al. (2008) point out that similar observations can be made at the spectrogram level of phonetic analyses. For instance, at roughly the 20 to 80ms mark the segmental cues are observable, while at roughly the 115-350ms window syllabic phenomena are reflected. A limitation of looking at these two windows in the spectrogram derives from the serial nature of our conscious attention mechanisms. One could, for instance, examine the syllabic window first and then consider the information in the segmental window, or vice versa. In neural terms, however, Poeppel and colleagues argue for parallel processing in the two windows. This process, Poeppel et al. (2008, p.1077) argue, is based on the idea of temporal integration windows in the brain. Physically continuous signals are fragmented and discretized for parallel analyses in different temporal windows. Both left and right auditory cortices have populations of neurons (wired as neuronal ensembles) that exhibit preferred integration constants of two types– one set of neurons prefers approximately 25 ms integration, another 250 ms. In electrophysiological studies, such integration windows are often reflected as activity in the gamma and theta bands, respectively. The spectro-temporal receptive fields(STRFs; neural recordings of distribution of frequency over time) in the auditory cortex then construct high-fidelity representation of frequency over time, which are then sampled by the two aforementioned integration windows to construct segments and syllables (or words) respectively. Poeppel, Idsardi, and Van Wassenhove, 2008, p. 1077 write,

> What is the purpose of such a proposed temporal quantization of the input waveform? We hypothesize that this sampling serves as a logistical or administrative device to generate **auditory representations of the appropriate granularity to interface with higher-order, abstract representations.**

In the view adopted here features, qua features, are represented at the intersection of stable patterns of neural activations entraining with oscillations at the theta and gamma bands. It is important to note that in my view neither the stable activation patterns, nor the oscillations themselves, can be reasonably equated with features (cf. Poeppel & Assaneo, 2020 for the view that patterns of activation *are* features). It is the process of parallel entrainment at the two relevant temporal windows that instantiates features. Conseqquently, the precise interpretation of a feature is contingent on three factors: (a) the segment and syllable the feature appears in, (b) the temporal window one is considering. For instance, consider the +ATR high vowel /i/ in a /CVC/ syllable/word. The +ATR is bundled in a segment, represented by the IPA symbol /i/, alongside other features such as +High and -Round. Given the antagonistic and connected architecture of the muscular system the way +ATR is implemented by the articulatory muscles in a high unrounded segment like /i/ will be different from the implementation of +ATR in a [+ATR, -High, +Round] segment. Likewise, at the syllabic level the precise identity of the consonants on either side of the /i/ will exert distinct co-articulatory effects. The figure (3.4) below illustrates the two parallel interpretation a feature receives in any given output of the phonological module.

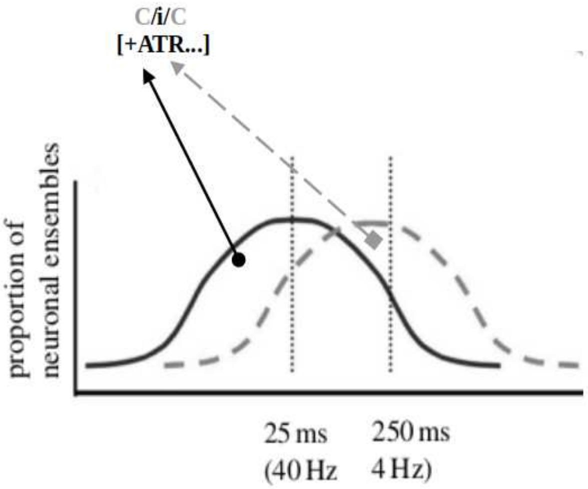


FIGURE 3.4: Features as Neural Entrainments

3.3.2 Analysis-by-Synthesis

Analysis-by-Synthesis (AxS) (Bever and Poeppel, 2010; Poeppel, Idsardi, and Van Wassenhove, 2008) is a computational strategy to incorporate both top-down and bottom-up processes in order to attain predictive coding of perceptual stimuli that Poeppel et al. (2008) characterize as a *hypothesize-and-test* approach. In fig. 2.4 above, the neuronal STRF codings yield a high-fidelity representation of the stimuli that can be thought of as the neural version of spectrogram. These representations then yield rough estimates of major class features in the signal, in a manner that is rather similar to Kenneth Steven's *landmarks* approach (Stevens, 2002; Stevens, 2000) to speech processing. In other words, at a first pass the acoustic information encoded in STRFs yield a rough estimate of the presence of the articulator-free major class features (Halle, 2005; Halle, Vaux, and Wolfe, 2000) in the signal. These usually correspond to the substantive correlates of major classes of speech sounds like plosives, nasals, approximants etc. (Mesgarani et al., 2014; Halle, 2005; Poeppel, Idsardi, and Van Wassenhove, 2008). Once a specific distribution of these major class features have been detected in a sample signal this has the effect of a reduction of the search space in the lexicon for forms that the signal must be mapped onto. This stage of representations corresponds, roughly, to what has been called the *(phonological) primal sketch*. The parenthesis indicates that information at this stage of the transduction process could, possibly, retain substantive information. In this sketch we are talking about *phonetic primal sketch*. Crucially, however, whether the information at this stage remains graded or is completely digital what is important for our purposes is that we now have a format that mediates between STRFs on the one end and lexical representations on the other. Poeppel et al.'s (2008, p.1073) suggest that this primal sketch is substantive and retains acoustic information, but the exact information varies between the two temporal integration windows, corresponding roughly to the segmental and syllabic levels. Once such top-down processes yield a primal sketch, it is used to generate a rough neighbourhood of likely lexical candidates. This process is akin to the cohort-type selection proposed by Marslen-Wilson (Marslen-Wilson, 1987) to explain how listeners or readers recognize and comprehend words during language processing. This theory focuses on the initial stages of word recognition and proposes a model that involves the activation and selection of potential word candidates from a cohort of similar-sounding words. The basic idea behind the cohort-type selection theory is that when we hear or see a word, our brain activates a set of word candidates that share similar sounds or visual features. This set of candidates is known as the "cohort." For example, when presented with the sound /kæt/, the cohort may include words like "cat," "can," "catch," and so on. As the word unfolds, additional information is processed, such as the context in which the word appears, the word's syntactic category, and the overall meaning of the sentence. This additional information allows the brain to narrow down the cohort and select the most likely word candidate. In the example above, if the word appears in a sentence like "I saw a cute ...," the context and meaning would guide the selection process, leading to the activation and recognition of the word "cat" from the cohort. The cohort-type selection theory also accounts for the fact that during word recognition, there can be competition among word candidates within the cohort. For example, if the sentence is "I saw a fast ...," the words "cat" and "car" from the cohort would both be plausible candidates. In such cases, further linguistic and contextual cues are employed to disambiguate the options and select the appropriate word. Crucial, for the current discussion, is the fact that cohort-type selection is employed in order to generate a neighborhood of likely lexical entries that the signal could be mapped onto.

Poeppel et al. (2008) further note that these AbS type mechanisms are broadly mathematically similar to Bayesian models. For instance, they point out that in creating the primal sketch the various "feature detectors" in the brain are never flawless and return probabilistic values.

Ex 3.3.1. H: Hypotheses, E:Evidence, p: Probability
$p(\mathrm{H}|\mathrm{E}) = p(\mathrm{E}|\mathrm{H})$ x $p(\mathrm{H})/p(\mathrm{E})$

$p(\mathrm{H}|\mathrm{E})$ denotes the probability of the analysis concerned, and $p(\mathrm{E}|\mathrm{H})$ denotes the probability of the desired synthesis given the data. Based on this initial sketch, largely comprising of probabilistic guesses about the distribution of major class, articulator-free features in the signal, articulator-bound features (±Round, ±ATR etc.) are identified in a context-sensitive fashion. Mathematically, Poeppel et al. (2008) argue that this stage involves evaluating conditional probabilities of the sort $p([\text{+Nasal}]|[\text{Labial}])$, because detection of [Labial] and [Coronal] in [+Nasal] is different from, say, in stops in a language which lacks velar nasals but contrasts all three positions with stops. This yields a series of temporally synchronized feature-trees, where each tree can be thought of as (roughly) a segment.

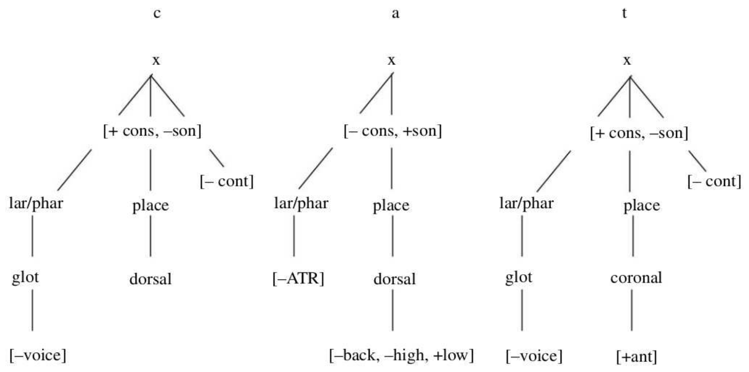


FIGURE 3.5: Temporally synchronized features in the signal.

3.4 Partial-veridicality

As a starting point I will adopt the position, emerging out of the last half a century's worth of phonological research, that phonological primes, or distinctive features, form the basis of lexical representation of the words that are conveyed through an analog and varying signal. Emerging neurobiological evidence suggests that specialized sub-populations of neurons in the auditory-cortext display selective preferences for spectro-temporal properties (STRFs) in the signal, in other words distributions of frequency over time (Poeppel and Idsardi, 2011; Hickok, 2022). These properties are argued to be hardwired (Poeppel, Idsardi, and Van Wassenhove, 2008), though they are usually assumed to be ecologically useful, shared across the mammalian brains and not specific to language (Fitch and Cutler, 2005). These neuronal preferences for STRFs afford a rather high-fidelity encoding of the signal, over which the initial estimate of major-class feature distributions can be made. The proposal I have in mind is based on Stevens, 2000's landmark hypothesis (LH), and has been argued for by Poeppel and colleagues in conjunction with an analysis by synthesis (AxS) framework (Bever and Poeppel, 2010) and multi-time resolution processing, or multiplexing (Poeppel and Idsardi, 2011). The underpinning idea behind LH is that the speech signal is interspersed with – locations in the stream that are relatively rich in information regarding the configurations of the larynx and the vocal tract articulators. For instance, Stevens (2000) points out that landmarks are often expressed in terms of acoustic cues for consonant closures, consonant releases, glide minima and vowel maxima, and their distribution offers cues to the relative positions and distribution of the major class features in the signal. Poeppel and Idsardi suggest that the human auditory system utilizes multiplexing specifically to encode and process speech signals across different timescales simultaneously. This multiplexing allows for the integration of information from various temporal scales, ranging from rapid changes in phonetic features to slower prosodic patterns. By incorporating multiple time resolutions, the internal forward model can effectively capture the dynamic and hierarchical structure of speech. In other words, the neuronal representation of the acoustic cues in the signal triggers a hypothesize-and-test process that yields a probabilistic estimate of the distribution of major class features. Poeppel and Idsardi, 2011, p. 1079 argue:

> Frequency information is encoded throughout the system, so that the search can proceed on a 'best-first' basis, with more probable parses assigned greater weight in the system. More specifically, we see the landmarks of Stevens, 2002, which correspond to the articulator-free features of Halle, 2002, and which define the 'major' classes of phonemes (stop, fricative, nasal and approximant), as defining a PPS [phonological primal sketch] of the segmental time scale. This primal sketch gives a neighbourhood of words matching the detected landmark sequence.

Once this distribution of major class features is attained, what the authors above term the primal sketch, conditional probabilities are then evaluated around these landmarks to obtain the articulator-bound features.

> That is, we see this layer of analysis as evaluating conditional probabilities of the sort p([labial] | [+nasal]). The hypothesized temporally synchronized feature sequences are then matched against the main lexicon and a list of candidates is generated, and then the rules of the phonology of the language are used to resynthesize the predictable features, again using Bayes's rule [...] **(Poeppel Idsardi, 2011; 1080)**

Poeppel and Idsardi, 2011, p. 1080 provide the figure below to illustrate the general idea of feature detection using a conjunction of AxS and landmark detection. Here, the boxes at the lowest level represent the distinct levels of mapping from "vibrations in the ar" to "abstractions in the brain". The topmost tier represents the AxS mechanism, while the ones in between represent the hypothesized computational processes that interact with the top and the bottom. Crucially, note that both while the neurogram is used to generate a PPS by the internal forward model, which in turn yields a cohort of possible lexical matches, the lexical hypotheses talk back to the internal model, which is in turn used to re-evaluate the neurogram. In this way, both bottom-up and top-down processes mechanically interact until a best match is found.

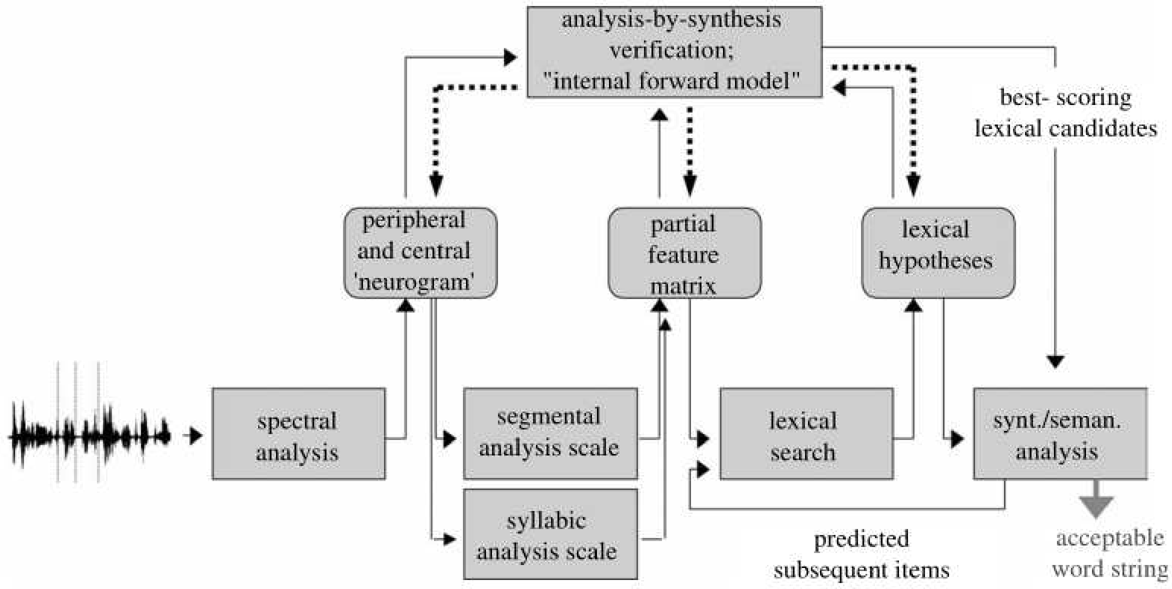


FIGURE 3.6: From vibrations in the ears to abstractions in the brain

In the view adopted here this process plays a crucial role during the critical period in the construction of the Scheer, 2010b's post-phonological spell-out lexicon (Lexicon 2 in figure 4.3 above). I propose that the contributions of phonological UG to the adult synchronic grammar take the form of (a) an instinct for the set of possible feature-based contrasts in any natural language (distinguished into major class features that correspond to broad acoustic effects of the configuration of the articulators, and articulator-bound features that correlate with anchor points that link possible configurations/states of specific articulators with expected acoustic targets), and (b) a *a priori* limited set of computational operations available to manipulate these symbols algebraically. The computations themselves are substance-free, as are the features. Exposure during the critical period, likewise, plays two distinct roles. Within phonology it allows the learner to detect the active contrasts in the language being acquired. While at the lower interface the AxS process outlined above is utilized to create a mass of probabilities that is associated with each active feature in the post-phonological lexicon. Crucially, note that in order for this model to work it is essential to have available, at birth, both an idea of the possible contrasts and the ability to segment and analyze substantive cues in order to assign readings to features. The features themselves are substance-free, in that they are neither articulatory nor acoustic in any meaningful sense. Rather, they are conceptualized as symbolic markers that serve as anchor-points for co-ordinate transformation between articulatory and acoustic spaces. The specific transformations necessary are contextual, language-specific and emerge during the critical period. In the figure below, the oval shape on the right represents the set of possible feature-based contrasts allowed for any natural language. I take this set, an expectation of what is at all possible, to be present at birth. Although what I have in mind is the rather trivial claim that there exist clear correlations between features, such as [+Round] (or [Round] if one prefers a privative system) and [-Back] and the resulting substantive cues – in the case of the former rounding of the lips (articulation) and a decrease in frequencies across the spectral range, and for the latter the pushing forward of the tongue body and an increase in first format values. Likewise, the neuromuscular mechanism utilized to control the articulatory organs' configurations, such that acoustic targets can be reliably hit, are trivially innate. This is also a rather trivial claim, as one can no more "learn" neuromuscular co-ordination than one can "learn" to have opposable thumbs. The genetically predetermined nature of the articulatory and perceptual channels *a priori* limits the range of possible contrasts, and the articulatory tools available to implement them in natural languages. This is what is meant by partial veridicality – the neuromuscular mechanisms for articulator control are rather invariant across the species, even though the acoustic and articulatory phonetic implementations show contextual variation within a range. Likewise, the neural coding, at least for articulator bound features, could not possibly be completely arbitrary. [+Round], for instance, emerges out of an innate ability to control the lips, and an innate instinct for hitting specific acoustic targets by specific configurations of the lips as active articulators. This feature, consequently, always rounds the lips, and never lowers the tongue body, for example. The specific implementation of lip rounding, however, are completely contextual and probabilistic. In other words, the ability to configure the articulators to hit an acoustic target, and the expected acoustic signature of a particular configuration, are both innate. Though it is highly unlikely that such innate predispositions are Language-specific (Fitch and Cutler, 2005; Samuels, 2011). This line of reasoning is derived from the FLN-FLB distinction made by (Hauser, Chomsky, and Fitch, 2002). The process of acquisition is indicated by the dotted arrows, makes use of AxS in order to determine (a) which features are active in the language being acquired, and (b) what is the range of possible of values that correspond to this feature (i.e. its acoustic signatures, and necessary articulator configurations). The former helps create a language-specific set of active features, while the latter creates a language-specific post-phonological spell-out lexicon. It is important to note that the values acquired are not specific point-values. Rather, keeping in track with the probabilistic nature of the internal feedforward process, I hypothesize that the learner constructs probabilistic ranges for each feature-value which allows for contextual as well as inter-token variations in the values attained on each externalization.

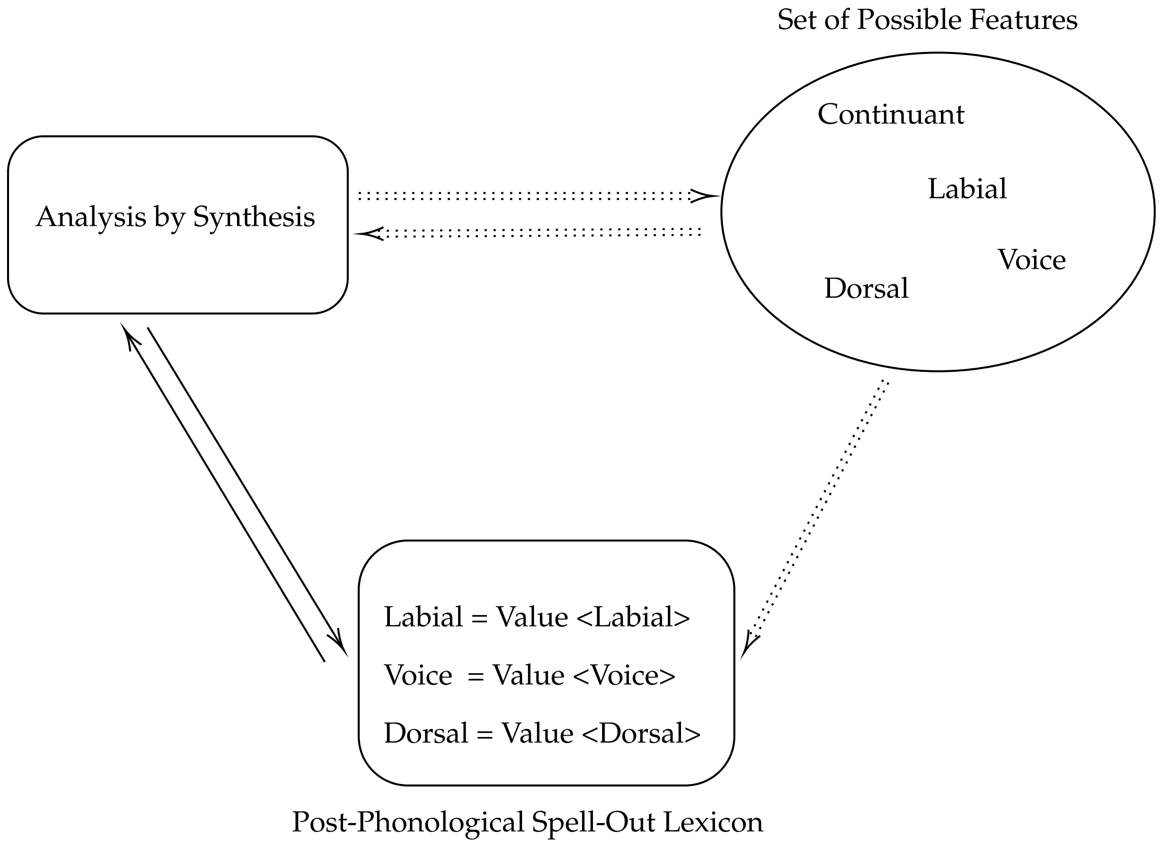


FIGURE 3.7: Post-phonological lexicon

The one way dotted arrow between the oval set of possible features and the lexicon indicates the critical-ness of the critical period, suggesting that once non-L1 contrasts have been lost the post-phonological lexicon remains (relatively) frozen in time.On the other hand, the interaction between the AxS module and the post-phonological lexicon is an active process in all synchronic I-languages, indicated by solid two-way interaction arrows. The combination of top-down and bottom-up processes try to match observed values in the signal to a possible feature-value combination in the lexicon, until the best match is found. Given the probabilistic nature of this process, of course, this process is liable to performative failures, leading to misperception, and ultimately diachronic change.

Finally, note that adopting the framework proposed here also allows us to make a testable prediction regarding the bi-directional processing of features, i.e. in production and perception of speech. Poeppel, Idsardi, and Van Wassenhove, 2008 and Poeppel and Idsardi, 2011 discuss the detection of features in the direction of perception, where *landmark detection* implies that the major class, articulator free features are detected first. This is because these features correlate with broader substantive properties of phonetic classes, like plosives or approximants. The values of articulator bound features tend to vary depending on the surrounding landmark, and are thus detected next by evaluating conditional probabilities. In my view it is reasonable to assume that in the direction of production exactly the opposite holds. Articulator bound features are interpreted first, and muscle configurations assigned to the relevant articulators. The resulting configurations have a restrictive effect on the interpretation of the major class features. For instance, Halle, 2005 discusses the issue of *designated articulators* (discussed in detail in chapter 6) – articulators that are assigned on a case-by-case basis to execute the major class features in different segments. I hypothesize that designated articulator assignment takes place after the interpretation of articulator-bound features, specifically because the latter both restricts the available options due to pre-configuration of a subset of articulatory muscles, and also lends to the overall phonetic qualities that characterize the major class features as landmarks.

3.5 Conclusion

The proposal outlined here attempts an unification of the core concerns of the two major schools of substance-free phonology. The primary criticisms are addressed squarely at the Concordia school's recent (Volenec and Reiss, 2017) insistence on deterministic transductions from phonological primes to phonetic readings. I argued that the notion of determinism is, at best, incoherent as no account is provided of what the deterministic values are of features, and how such determinism could be implemented in a biologically plausible fashion within a largely antagonistically designed articulatory system. The framework also adopts the radical school's idea that architectural similarities between the upper and lower interfaces of phonology is to be expected, and implements Scheer, 2010a's post-phonological spell-out system as the underpinning of the lower-interface. However, unlike the upper interface where spell-out is completely arbitrary, I argued that at the lower-interface at least partial veridicality must hold between primes and their phonetic functions. This is to be expected because all available evidence suggest that features are implemented in the brain at the conjunction of innate, though not necessarily language-specific, neural substrates and operations (Fitch, 2000; Fitch and Cutler, 2005; Poeppel and Idsardi, 2011; Samuels, 2011). Research into the human somato-sensory system and maps in the brain (Brecht, 2017) provide good reason to assume that humans have an innate understanding of their neuromuscular capacities, and an ability to internally simulate muscle movement and configurations. Samuels, 2011, likewise, provides a good overview of a variety of evidence that points to the fact that humans, like many species of birds, have an innate ability to vocalize and mimic perceived sounds, configuring articulators to hit perceived necessary acoustic targets. If we adopt the view that phonological representations and computations are implemented by a species-specific combination of largely shared, but still innate, neural wetware and operations, then it seems reasonable to assume the form of partial veridicality argued for here.

Chapter 4

At The Lower Interface-I: Dichotic Listening in Bilingual Phonology

4.1 Introduction

Listening to speech in the real world involves continuous detection and processing of speech signals in a situation where conditions are not ideal. Rarely, if ever, do listeners receive speech signals in isolation, and the environment almost inevitably contains noise as well as speech from other people, both of which are possibly irrelevant to the attended speech stream. Competing speech signals vie for the listener's attention, so that the listener's task includes isolating a single target utterance as well as locating meaningful linguistic units, processing them for a meaningful message and interpreting the total message to arrive at its semantic content. In that process, both linguistic knowledge, and the representations of that subset of linguistic knowledge that help listeners abstract semantic concepts from speech signals, i.e. phonological knowledge, are called upon. For bilingual listeners, the task is further complicated by the availability of more than one set of linguistic/phonological knowledge, especially given the evidence that speech input activates for such listeners the linguistic units of each language (Grosjean, 1988; Spivey and Marian, 1999; Weber and Cutler, 2004). For all listeners, processing speech requires a combination of both linguistic and non-linguistic mental/cognitive and motor functions; for bilingual listeners the linguistic functions are potentially increased to encompass the phonological grammars of two languages. It is only by a combination of all these functions that active speech perception, and then comprehension, is made possible.

One function that listeners must use, in order to limit the amount of data forwarded for processing, is selective attention, i.e., focusing on particular stimuli while ignoring others. Cherry, 1953's landmark study of selective attention in speech examined how listeners track certain conversations while tuning others out (now known as the "cocktail party-effect"). In his experiments, two auditory messages were presented simultaneously, one to each ear (i.e., dichotically), and participants were asked to attend to and repeat back one of them. The monolingual English participants were able to do this easily, but most interestingly, when asked about the content of the other message, they were unable to say anything about it. Cherry found that even when contents of the unattended message were suddenly switched (such as changing from English to German mid- message, or suddenly playing backwards) very few participants noticed. If the speaker of the unattended mes-sage switched from male to female (or vice versa), however, or if the unattended message was swapped with a 400-Hz tone, the change was always noticed. Cherry's findings have been often replicated, with similar results holding for lists of words and musical melodies. The task, now called dichotic listening, was widely adopted in studies of brain lateralization in language processing. A right ear advantage is typically found for dichotic perception of consonants, which is interpreted as reflecting a left brain hemisphere superiority in phonological processing (Kimura, 1961a; Kimura, 1961b; Liberman et al., 1967). That right ear advantage is more reliable in right than left handers.

In this study, we investigated the interaction among the general cognitive ability of selective attention, physical properties of the speech signal, and the linguistic functions involved in phonological decisions. Our participants were bilinguals with two phonologically quite different languages, allowing us to examine the ability to selectively attend to one set of phonological knowledge rather than another, while our task was modeled on Cherry's classic dichotic selective attention paradigm and its use in studies of hemispheric asymmetries in speech perception. Specifically, we explored how competing speech signals are mapped onto constitutive linguistic knowledge, using dichotic listening. Our experiments were conducted using second language (L2) dominant early sequential bilinguals whose first language (L1) is Malayalam and whose L2 is Australian English, and we tested their perception of consonants carefully selected for their phonological properties. In what follows, we describe this study and its outcome. Section 2 provides background information on the Malayalam and English consonants of interest and their respective phonetic and phonological properties. This section outlines a feature-based proposal for the consonants. Sections 3 and 4 report the experimental design and the results, respectively. Section 5 provides a discussion of the results and Section 6 briefly concludes.

4.2 English and Malayalam target consonants

Our experimental design focused on the labial stops and fricatives (obstruents) of Malay-alam and English. Malayalam, sometimes referred to as Kairali, is an Indic language of the Dravidian family, speculated to be philologically related to 6th Century Sen-Tamil (Middle-Tamil) (Asher 1985). For labial obstruents, Malayalam phonology employs a four-way voicing-aspiration contrast with stops, but it lacks fricatives at the labial place of articulation. English is a Germanic language that has a two-way laryngeal contrast between its two labial stops `/p, b/` and its two labio-dental fricatives `/f, v/`. The chart in figure 5.1 below illustrates the complementary gaps in the phonological inventories of English and Malayalam labials. English has fricatives, Malayalam lacks them; Malayalam makes full use of a 4-way stop voicing × aspiration contrast, English uses only a 2-way laryngeal contrast.

Labial Obstruents			**English**	**Malayalam**
STOPS	**Voiceless**	**Unaspirated**	/p/	/p/
		Aspirated	*	/pʰ/
	Voiced	**Unaspirated**	/b/	/b/
		Aspirated	▨	/bʰ/
FRICATIVES	**Voiceless**	■	/f/	▨
	Voiced	■	/v/	▨

FIGURE 4.1: Distribution of stops and fricatives in English and Malayalam. Gray cells in a given language indicate phonological gaps in that language

For the purposes of description, we have characterized English in figure 5.1 according to what Hall (2001) calls the "standard approach" to laryngeal features. This approach maintains that the phonological features capturing two-way laryngeal contrasts are the same in languages that realize the contrast primarily in terms of voicing, e.g., /p/ vs. /b/, and those that realize it primarily in terms of aspiration, e.g., /p/ vs. /pʰ/. The standard approach goes back to at least to Lisker and Abramson, 1964, is adopted by Chomsky and Halle, 1968, and is explicitly argued for by Keating, 1984 and Lombardi, 2018. There is also a substantial phonological tradition that treats voicing distinctions in aspiration languages as featurally distinct from voicing distinctions in true voicing languages. Hall, 2001 dates this tradition back to Jakobson, 1949. In more contemporary studies, it has been referred to as "laryngeal realism" (Honeybone, 2005) and has been considered justifiable on both phonetic and phonological grounds by a number of researchers (Iverson and Salmons, 1995; Iverson, Salmons, et al., 1999; Iverson and Salmons, 2003; Jessen and Ringen, 2002; Kehrein and Golston, 2004; Petrova et al., 2006; Iverson and Ahn, 2007; Kager et al., 2007; Beckman, Jessen, and Ringen, 2013). A key aspect of our interest in Malayalam-English bilinguals is that in contrast to the substantial debate over the voicing feature in Germanic languages, such as English, as opposed to Romance languages with true voicing contrasts, laryngeal specification in Malayalam is rather uncontroversial, since features for voicing and aspiration ([spread glottis]) are fully crossed. We return to the issue of laryngeal phonology in our discussion, in light of our dichotic listening results.

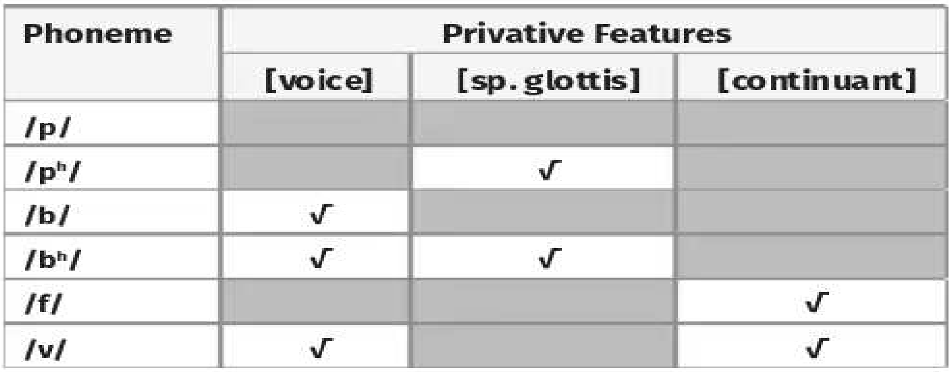
One motivation for investigating the interaction of phonetic features of stops and fricatives by this population is prompted by the fact that native Malayalam speakers often replace the labiodental fricatives (which don't exist in Malayalam) of English loan-words with one of the aspirated Malayalam bilabial stops. Adaptations such as pronouncing the English word 'Venice' [vɛnɪs] as [b^henis] are quite common for these speakers. This suggests that at some level of representation /v/ and /b^h/ are in correspondence. Our experiment investigates whether bilingual listeners of Malayalam (native language, or L1) and English (second language, or L2) are able to switch attention from one language to the other in a dichotic task, ignoring stimuli presented in the unattended language (opposite ear). To the extent that they are unable to ignore the distractor stimuli, we ask how dichotic presentations of stimuli will be mapped onto phonological representations in each language.

We anticipate that both phonetic and phonological similarity of the consonants in figure 5.1 may play a role in intrusions. Acoustic distinctions between voiced and voiceless stops and fricatives can be captured by the measure we focus on in the present study, the ratio of periodic (harmonic) to aperiodic noise, also referred to as harmonics to noise ratio (HNR), as measured during the consonantal period. Voiced fricatives have a higher HNR value (expressed in dB) than voiceless fricatives. For stops, the ratio should again be higher for voiced than voiceless ones. However, because stops involve much more rapidly changing acoustic properties within a shorter time window than fricatives (brief release burst followed by pre-vocalic voicing, silence or aspiration) the magnitude of the HNR difference between voiced and voiceless stops is smaller than seen in fricative voicing contrasts. Moreover, aspiration, which involves turbulent airflow, would also be expected to affect the ratio, thus lowering HNR scores for aspirated relative to unaspirated stops. In the following section, we present HNR comparisons for the stops and fricatives used in our experiment; the comparisons confirm the coarse-grained expectations stated here.

To make concrete how phonological factors may also structure patterns of intrusions, we have constructed a feature-based proposal for our bilingual population. The proposal makes use of privative feature theory and underspecification (Archangeli, 1988; Steriade and Goldsmith, 1995; Golston, 1996; Lahiri and Reetz, 2010). This approach aids us in relating the phonological notion of markedness to our experimental data, as we can define markedness as the number of phonological features required to represent a speech sound. Unmarked sounds tend to be typologically more frequent in the world's languages. Specifically, the presence of a marked sound in a language tends to imply the presence of an unmarked counterpart. This is true of voiced and voiceless stops. The presence of voiced stops in the inventory of a language almost always implies the presence of voiceless stops, but not necessarily the other way around (Maddieson, 1984). Unmarked sounds are also sometimes argued to be either perceptually salient and/or articulatorily simple (Kenstowicz, 1994), although empirical evidence supporting this claim is incomplete. If unmarked sounds are percep-tually more salient and thus easier to process, then one would expect that such ease of processing would make the unmarked voiceless stops in our experiments harder to ignore, leading to more intrusions when they occur in the unattended language (and ear).

On the other hand, unmarked sounds have, on our proposal, fewer phonological features, as we have assumed that predictable information is unspecified. Information can be considered predictable in one of two ways. First, information is predictable if it is redundant or allophonic: the aspiration of the bilabial stop at onset of the word pan /pæn/ in English is predictable (because in stress-initial syllable position, voiceless stops are always aspirated in English). Second, this predictability means it is possible to leave one value of the feature blank in underlying representations. If the feature is not specified, e.g. [+F], then it must be [–F], by default, i.e., a privative feature. The approach that defines predictability in this latter fashion is known as radical underspecification, and it makes a different prediction for our experiment than that made by the assumption that unmarked features are perceptually salient. Specifically, it could be that we observe more intrusions from segments that have a greater number of feature specifications

Our feature-based proposal, shown in Table 2, makes use of three features to differentiate the labial stops and fricatives relevant to this study. The Malayalam voiceless stop /p/ is underspecified for all three features, which is indicated by gray shading. In our experimental design this Malayalam consonant is paired with its dichotic competitor, the English voiceless fricative /f/, which is specified for one feature ([continuant]). In a similar fashion, in the voiced stimulus pairing, we find that the Malayalam voiced stop /b/ bears a single specification ([voice]), whereas its competitor, the English voiced fricative /v/, bears two specifications ([voice] and [continuant]), the latter again being the feature that differentiates the English from the Malayalam item.



Phoneme	Privative Features		
	[voice]	[sp. glottis]	[continuant]
/p/	▨	▨	▨
/pʰ/	▨	√	▨
/b/	√	▨	▨
/bʰ/	√	√	▨
/f/	▨	▨	√
/v/	√	▨	√

FIGURE 4.2: Privative feature matrix contrasting the stimuli-initial consonants.

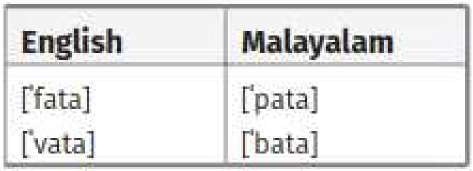
The proposal follows from privative feature matrices and radical underspecification in the following way. The voiceless stops lack voicing and are assumed to be the default specification for stops; hence they do not require an underlying specification for [voice]. The feature [continuant] does not apply to stops, and [spread glottis] is, likewise, not an essential for unaspirated stop articulation, but is only required to express an aspiration contrast (where applicable) between voice-matched stops. Further, the voiceless fricative /f/ lacks specifications for [voice] and [spread glottis], and only requires [continuant], which captures the turbulence of continuous air-flow through a narrow constriction some -where in the vocal tract that characterizes a fricative. This phoneme contrasts with the voiced fricative only in respect of the latter's specification for the feature [voice]. By making explicit our phonological proposal for contrasts within our bilingual population, we can evaluate how phonological specification may relate to patterns of intrusion in the dichotic selective attention task. Specifically, we investigate bilinguals' ability to attend selectively to one of their languages, while attempting to tune out the other, asking whether such selective attention tasks are marked by significant numbers of intrusions of phonetic properties from the simultaneously presented item in the unattended language. If intrusions are indeed observed, we ask whether they can be explained by acoustic prop-erties, e.g., HNR, and/or how phonological features are distributed across the bilingual representational space.

4.3 Experimental Design

We configured our methodological choices so as to enable us to create an environment where listeners receive simultaneous auditory input from both of their languages, in separate ears, but are required to attend to only one language and attempt to ignore the other language.

4.3.1 Stimuli

The audio target and response choice stimuli, displayed in figure 5.3, were CVCV (initial stress) nonce words with the phonological properties of Malayalam and English but meaningless inboth languages. They were designed with attention to syllable structure and consonant realizations in stress-initial positions of the two languages (see Table 1 above). For English there were two labio-dental fricatives contrasting in [voice], /f/ and /v/; for Malayalam,there were two bilabial stops contrasting in [voice], /p/ and /b/. In the dichotic trials,nonce words from English always began with one of the two fricatives /f, v/ in the initial stressed syllable, while nonce words from Malayalam always began with one of the two unaspirated stops /p, b/ in the initial stressed syllable. While the participants in the study only ever heard Malayalam unaspirated stops /p, b/, and English fricatives /f, v/, we also recorded and calculated the HNR of every labial phoneme from Malayalam and English that were provided as a target response choice (as we explain below, there were also coronal response choices included as distractors; these however are irrelevant for the HNR analysis).



English	Malayalam
[ˈfata] [ˈvata]	[ˈpata] [ˈbata]

FIGURE 4.3: The dichotic nonce stimuli in each language, in which the stress-initial consonants are the target and distractor items in each respective language-attend condition.

To record our stimulus materials, we recruited a Malayali-Australian male bilingual(age: 27 years) from the Sydney Malayali-Australian community, who produced all stimuli for the experiment including those with the English labial stops and the Malayalam aspirated stops, which were not used in the dichotic selective attention experiment but were measured for HNR comparisons. He was born in Australia to Malayalam-speaking parents, and acquired Malayalam as his L1 in the home within the first few years of life. However, all of his formal education was in Australia in English, thus he was a fluent speaker of Australian English, which had become his dominant language (L2-dominant), although he used Malayalam regularly and remained fluent in his L1 as well. We recorded him at MARCS Institute, Western Sydney University in the anechoic chamber using a Roland UA 25-EX sound card on a Lenovo Thinkpad laptop running Windows 7. He produced 10 or more tokens of each of the eight nonce stimulus types (the four Malayalam stops; the two English stops and two English fricatives) in citation form, with a constant intonation contour. We selected the 8 tokens of each category that were best-matched in duration, loudness, and pitch to use as the audio stimuli in the dichotic task (described in Procedure), as well as in the HNR analyses we conducted (described next).

In order to ascertain the acoustic phonetic nature of the phonological differences represented in our stimulus materials, we derived HNR scores for all tokens of each labial phoneme presented either as auditory stimuli in the dichotic task (English /f, v/; Malayalam/p, b/) or as response choices in the task (English /p, b, f, v/; Malayalam /p, p^h, b, b^h/). For the stops, we calculated HNR over the temporal window starting with stop-release and ending at vowel onset. For the pre-voiced Malayalam /b/, we excluded the pre-voicing temporal window from the measure. For the fricatives we measured the entire duration, starting with beginning of frication and ending at vowel onset.

A 2 × 2 × 2 repeated-measures ANOVA with Language (English and Malayalam), Voicing (voiced and voiceless) and Turbulence was conducted on the HNR scores. The Turbulence factor refers to whether there is a narrow articulatory constriction at some location in the vocal tract that results in airflow turbulence and acoustic noisiness (i.e., fricatives and aspirated stops), or whether vocal tract lacks such a constriction and the articulation thus lacks turbulence (unaspirated stops). The HNR ANOVA revealed significant effects of Language (higher HNR in English than Malayalam), $F(1, 7) = 5182.55$, $p < 0.01$; Turbulence (higher HNR for fricatives and aspirated stops than for unaspirated stops), $F(1,7) = 2937.52$, $p < 0.01$; and Voicing (higher HNR scores for voiced than voiceless items overall), $F(1,7) = 5148.75$, $p < 0.01$). Interactions also appeared between Language and Turbulence, $F(1, 7) = 38.72$, $p < 0.01$, Language and Voice,$F(1, 7) = 7027.91$, $p < 0.01$, Turbulence and Voice, $F(1, 7) = 8391.52$, $p < 0.01$, as well as Language, Voice and Turbulence, $F(1, 7) = 720.16$, $p < 0.01$. Overall, we found that while English phonemes display higher HNR values on average, across languages the more turbulent phonemes (English fricatives and Malayalam phonologically aspirated stops) have much higher HNRs for voiced than voiceless consonants, whereas phonemes with low turbulence (phonologically unaspirated stops in both languages) instead shows lightly higher HNRs for voiceless items than voiced ones. However, average HNR values are much lower for voiceless phonemes that are low in turbulence than for the turbulent voiced phonemes. Accordingly, the audio stimuli used in the dichotic selective attention perceptual experiment (Malayalam /p, b/, English /f, v/) can be arranged in the following hierarchy, ranging from highest to lowest HNR values: English /v/ > Malayalam /b/> English /f/ > Malayalam /b/. These data will inform our discussion of the dichotic listening results in the General Discussion section of the paper.

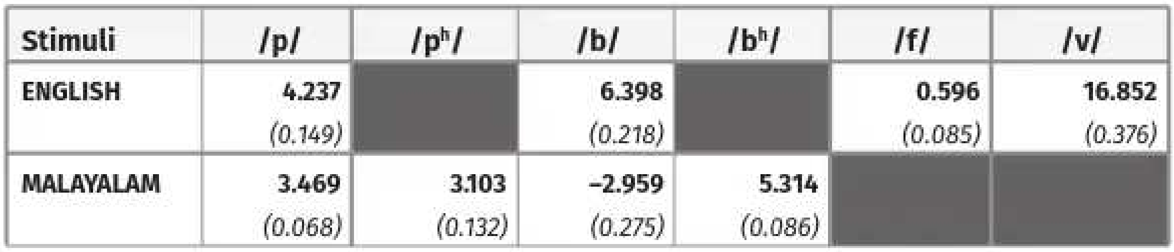
In the dichotic task (see Procedure), the Malayalam-English (or English-Malayalam) pairs of audio stimuli were always matched for voicing, and contrasted only in terms of aperiodicity in the signal (as reflected by HNR measures of the consonant and vowel portions of the initial consonants of the nonce words – see figure 5.4). In phonological terms,this reduces to a contrast simply between presence/absence of the feature [continuant] in each trial (see figure 5.2). In the Malayalam-attend condition, the Malayalam unaspirated stops were the target items and the simultaneously-presented English initial fricatives in the opposing ear were the distractors. The converse was true for the English-attend condition.

4.3.2 Participants

Thirteen participants (seven female; age range 18–45 years) took part in the dichotic listening study, recruited from the Malayali-Australian community via flyers posted in churches, schools, and other community activity locations. Like the stimulus speaker, all were adult Australians with a Malayali heritage, and were born in Australia, but acquired Malayalam as their L1 from their family in the home within the first few years of life.However, their formal education being in Australia, all were fluent speakers of Australian English, with all segmental and suprasegmental qualities relevant to the current design. All participants completed a language-background questionnaire, which confirmed their fluent bilingual Malayalam-L1/ English-L2 dominant language status. None had any auditory or speech impairment, and all were right-handed, as handedness is known to correlate with the lateralization of phonological processing. All gave voluntary consent for participation. One interesting observation early on in participant recruitment was that while all our participants were verbally fluent in both English and Malayalam, they were literate only in English. This was taken as a confirmation that the participants were L2-dominant in English, while remaining fluent bilinguals.

4.3.3 Procedure

Stimuli were presented to the bilingual listeners via Sennheiser M2200X isolated head-phones, using a Lenovo Thinkpad laptop computer and a Roland UA 25-EX audio card.They were presented dichotically, with the listener being instructed to attend to a given language in a given ear in 4 blocks of trials, one each for Malayalam-attend right ear and Malayalam-attend left ear, and the corresponding two blocks for English-attend. Within each block, there were 128 trials. These consisted of 8 tokens of each of the four stimulus categories (shown in figure 5.3) combined with one of two distractors. On each trial, one 'CVCV item from each language set (matched for voice quality, speaking rate and pitch) was presented to the two ears simultaneously. The two items presented on any trial differed in manner but matched in the voicing of the initial consonant in the first (stressed) syllable (see figure 5.3). All participants completed four blocks in which they attended to L1 and to L2, each in the right ear vs. the left ear, with order of blocks counter balanced across participants



Stimuli	**/p/**	**/pʰ/**	**/b/**	**/bʰ/**	**/f/**	**/v/**
ENGLISH	**4.237** *(0.149)*		**6.398** *(0.218)*		**0.596** *(0.085)*	**16.852** *(0.376)*
MALAYALAM	**3.469** *(0.068)*	**3.103** *(0.132)*	**−2.959** *(0.275)*	**5.314** *(0.086)*		

FIGURE 4.4: Mean HNR scores in dB for English and Malayalam stimulus consonants. Standard deviations are provided in parentheses.

For the task, the participants were required to listen to the simultaneous stimuli presented dichotically. A set of pictorial representations of actual Malayalam and English content words were displayed on a monitor for the participant to select from. The pictures included words beginning with the target consonants, which were all labial, e.g. pot for /p/, ball for /b/, etc., as well as coronal consonants, which were never the correct response choices and served as foils. The total set of 14 consonant response choices per language were given, represented by the pictures: for English /p, b, f, v, s, ʃ, t, d, m, v, w,z, tʃ, ʤ/; for Malayalam /p, p^{h}, b, b^{h}, s, ʃ, t, d, m, z, c, c^{h}, ɟ, ɟ^{h}/). An example of a picture response wheel for each language is provided in Figure 1 for each language. The orthographic labels in the picture are presented here for the readers' aid and were not included in the actual task. The circular arrangement of the items was selected in order to not bias responses towards any particular option. In order to accommodate all target consonants (labials) and distractor consonants (coronals), two response wheels per language were required (see fig. 5.5 below). Pictorial representations were selected based on the target consonants in the language that the participants were required to track and identify. Thus, the initial consonant in the name of the picture, for example Malayalam /p/ in /paava/ "doll" in the picture of a rag-doll, was required to match the target consonant /p/ in the attended Malayalam nonce word /pala/. The participants were instructed to click on the image whose name began with the same consonant as the word they heard in their attended ear.We used pictures as response choices instead of the orthographic form of the phones that formed the response choice set because the participants were not literate in Malayalam.

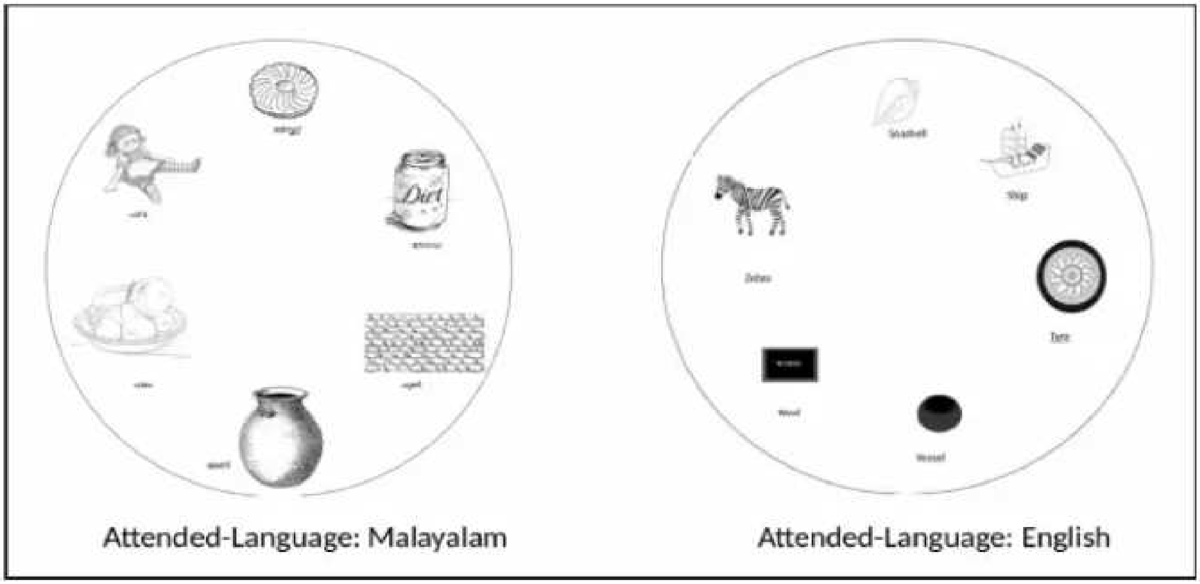


FIGURE 4.5: Examples of response-choice wheels.

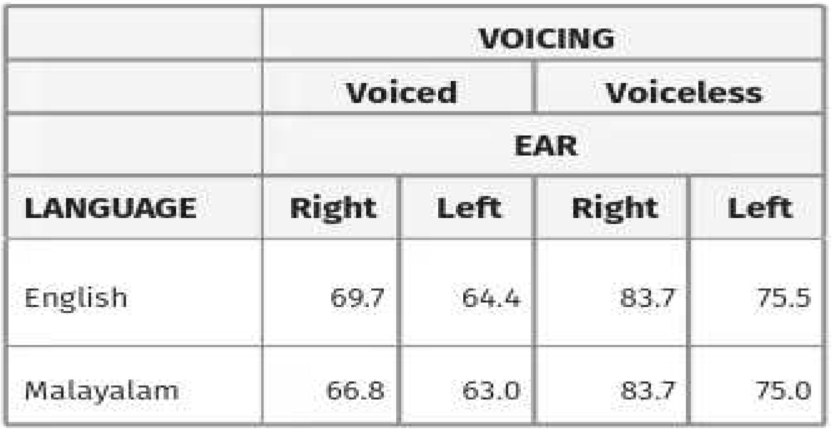
We provided the listeners with a whole range of coronal consonants as possible response choices in order to ensure that participants have the phonological freedom to select an emergent perceptually assimilated form without being restricted a priori by the available response choices.

4.3.4 Results

We report our results in terms of accuracy (figure 5.6) and in terms of intrusions (figure 5.7).A response was counted as accurate when the picture corresponding to the initial consonant in the attended ear was selected. For example, in the Malayalam-attend condition,responses of /p/ to [pata] (with [fata] in the unattended ear) and of /b/ to /bata/ (with [vata] in the unattended ear) ear were counted as correct responses. An intrusion was defined as an incorrect response influenced by the stimulus played in the unattended ear. To continue with the same example of the Malayalam-attend condition, responses of /p^h/(to [pata] with [fata] in the unattended ear) and /b^h/ (to /bata/ with [vata] in the unattended ear) were counted as intrusions.

Our design also allowed for other incorrect responses. For example, a participant couldhave, in principle, responded /p^h/ to [bata] with [vata] in the unattended ear or could instead have chosen one of the coronal response options. These types of errors were extremely rare. Incorrect responses in which a listener chose a consonant that differed invoicing from the target stimulus were non-existent. As a consequence, the percentage of intrusions (figure 5.7) is very nearly the complement of the percentage of accurate response(figure 5.6).

Figure 5.6 provides the mean percent correct for each cell in the design. Accuracy rangedfrom 63.0% to 83.7%. Consistent with past work using the dichotic listening paradigm (Kimura, 1961a; Kimura, 1967; Liberman et al., 1967), the means suggest a tendency for accuracy to be higher when the target was presented to the right ear than when it was presented to the left ear. There was also a tendency for accuracy to be higher for voiceless targets than for voiced targets. The error patterns by voicing and ear (left/right) were similar regardless of the target language.



	VOICING			
	Voiced		Voiceless	
	EAR			
LANGUAGE	**Right**	**Left**	**Right**	**Left**
English	69.7	64.4	83.7	75.5
Malayalam	66.8	63.0	83.7	75.0

FIGURE 4.6: Mean percent correct responses for English and Malayalam..

Since there were no incorrect responses involving a mismatch in voicing, e.g., selecting/p^h/ to [bata] or /b^h/ for [pata] in the case of the Malayalam-attend condition or selecting /p/ for [vata] or /b/ for [fata] in the English-attend condition, the pattern of intrusions largely mirrors the accuracy patterns in figure 5.6. The intrusions are the issue of core interest to the present study, and are presented by condition in figure 5.5. Visual inspection indicates that intrusion rates were lower when the target stimulus was presented in the right ear and lower for voiceless stimuli than for voiced stimuli. This pattern was consistent across language conditions.

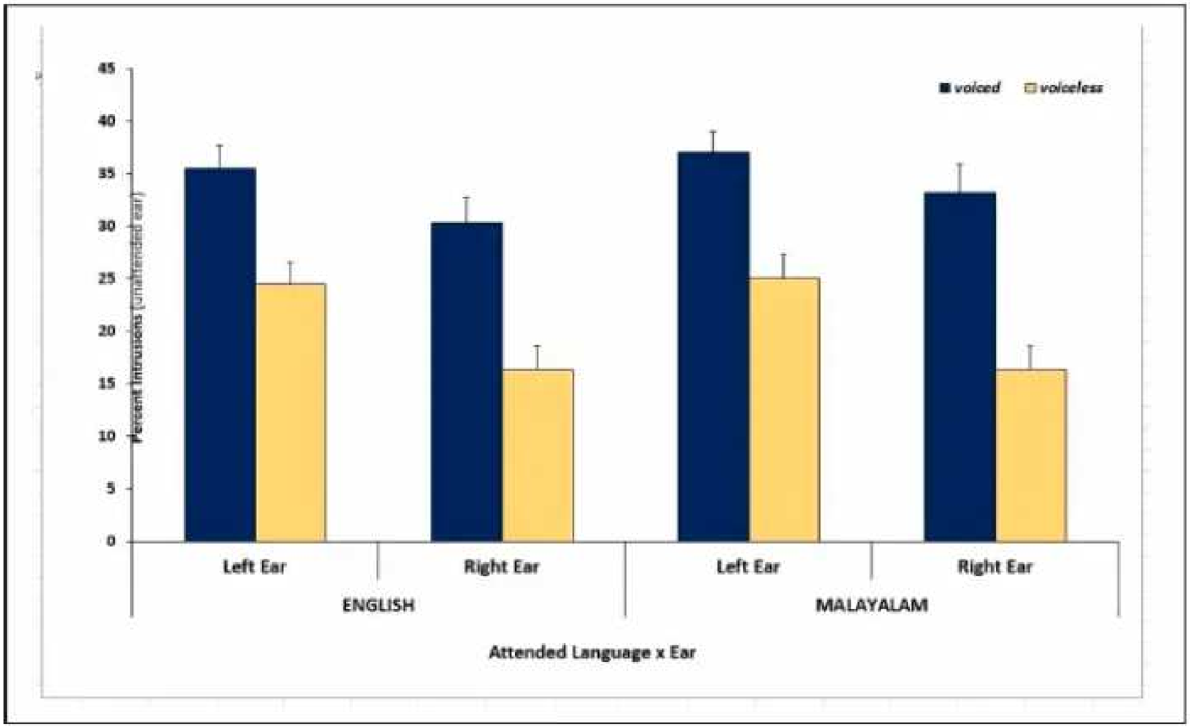


FIGURE 4.7: Intrusions from unattended language/ear by language, ear and target consonant type.

To evaluate the statistical significance of these observations, we ran a 2 × 2 × 2 repeated-measures ANOVA on the rate of intrusions, with the factors attended language(English, Malayalam), ear of attended language (right, left) and target voicing (voiced, voiceless). The ANOVA revealed statistically significant intrusions from the phone in the unattended language/ear, with significantly more intrusions when the attended language was in the left ear (ear of attended language main effect), $F(1, 12) = 8.58$, $p < 0.05$; and significantly more intrusions for voiced targets (target voicing main effect), $F(1, 12)= 52.38$, $p < 0.01$; but no significant effect of attended language, $F(1, 12) = 2.69$, $p >0.05$. We found no significant interactions (all $p > 0.05$).

In sum, our results indicate that participants are able to maintain attention and successfully track the target phone in the attended language/ear the majority of the time, independently of which language was being attended. There were, however, a significant number of intrusions, indicating that the stimulus in the unattended ear influenced responses, and this occurred more often for voiced stimuli than for voiceless stimuli.

4.4 Discussion

The patterns of intrusion attested in our results are as follows. First of all, in line with other work on dichotic listening, our results showed a right ear advantage. There were fewer intrusions from the unattended ear when participants focused on target stimuli presented to the right ear than when they focused on the left ear. This likely reflects a bias towards processing phonology in the left hemisphere, according to the widely-accepted interpretation that stronger contralateral than ipsilateral ear to cortex connections reflect a right-ear advantage and hence left hemisphere superiority in speech perception, at least for right-handed listeners (Kimura, 1961a; Kimura, 1967; Liberman et al., 1967). Second, we found that there were no voicing errors in this dichotic perception task. When listeners selected incorrect responses, which was not rare, the incorrect response was always one that matched the correct response in the voicing feature, e.g., /p/ was sometime selected for [fata] in English listening mode but /b/ was never selected. Third, we found that there were more intrusions in manner for voiced stimuli than for voiceless stimuli. Fourth, we found that the attended language had no significant effect on the rate of intrusions. That is, the right ear advantage as well as the effect of voicing on manner intrusions (more intrusions for voiced stimuli than for voiceless stimuli) was the same regardless of the attended language.

With respect to our feature-based analysis of Malayalam-English bilinguals (Table 2), the effect of voicing on manner intrusions is of particular interest. Voiced stimuli in both languages/ears made it more difficult for the participants to maintain attention to the target language/ear (mean intrusions 34.04%) than did voiceless stimuli in each ear/language (mean intrusions 20.55%). On our account, voiceless phones are the unmarked specification, which is consistent with the tendency for languages to develop voiceless obstruents before voiced ones. However, our results indicate that markedness cannot be equated with the psycholinguistic notion of salience, since in our study it is the marked segments (voiced) that presented greater perceptual salience by permitting intrusions upon attentional mechanisms more often than the unmarked (voiceless) segments do. That is, participants found it harder to tune out unattended voiced segments than voiceless ones.

This result speaks to the possibility raised in the introductory section that the perceptual salience of a segment is related to the number of phonological features needed to specify it. Consider, for instance, a trial in the current design where a listener is presented with two dichotic stimuli of the pattern [vata] – [bata], and thus faces one of two possible tasks – to attend to the English item [vata] (if the trial is English-Attend), or to attend to the Malayalam item [bata] (if the trial is Malayalam-Attend). In an English-Attend trial,the listener's unattended ear is receiving a signal (Malayalam /b/) that, on our analysis, can only trigger the feature [voice], while the attended signal (English /v/) triggers[voice] and [continuant]. The listener's rate of correct detection of [voice] should be high, given that evidence for this feature is present in both ears. This is attested in our results. However, intrusions do also occur to a substantial extent, as our results also attest,and this also demands an explanation. When listeners have to recognize both [voice] and an additional feature [continuant], intrusions occur to a greater degree than when the listener must recognize the absence of [voice] and the presence/absence of [continuant]. Given our feature-based account, it is possible that there is a general principle at play in these results—selective perception is more difficult when there are a greater number of features to recognize.

The voiced stimuli, which involved a greater number of features, on our account, also showed larger acoustic differences from each other in harmonic-to-noise ratio (HNR) than did the voiceless stimulus pairs (see figure 5.4). English /v/ and Malayalam /b/ have a large HNR difference (19.859 dB); the difference between English /f/ and Malayalam /p/ is notably smaller (3.555 dB). This large difference appears to make it less likely that listeners will be able to completely tune out the unattended ear. Recall that Cherry, 1953's original selective attention experiment showed that a large spectral change in the unattended ear was always noticed. The attended English /v/, with the highest HNR value of all the phones in English, is influenced by the Malayalam /b/, with the lowest HNR value of all the Malayalam phones. The perceived phoneme is the one with the second highest HNR value in English, a /b/.

In the Malayalam-Attend condition, however, the listener faces the opposite challenge. It is now necessary to tune out a signal that has greater triggering capacities(unattended English /v/ has two features, [voice] and [continuant]), while attending to the Malayalam target /b/ that triggers only one feature ([voice]). The more marked signal in the unattended ear in this case causes intrusions into the attended ear. When our participants' reports indicated a misperceived phoneme, it was always an aspirated voiced stop, Malayalam /b^h/. This is easily accounted for under our current set of assumptions. The signal phonetics will trigger features that help maintain contrast between competing sounds within a given language's phonological system. The competing signal, in this Malayalam-Attend case, differs from the attended signal in terms of the feature [continuant]. Given the higher feature specification of the competing signal and large periodicity difference, it will intrude, but the listener's lexical access will be faced with a gap in Malayalam, in that the intrusion would now trigger a feature that the lexicon does not utilize at this place of articulation (labial): [continuant]. The consequence is that the listener selects /b^h/, which has an increased HNR higher than Malayalam /b/

We argue that this situation is best explained by referring to the notion of perceptual assimilation of closest resembling phonological category as elucidated in the Perceptual Assimilation Model, or PAM (Best, 1995b). In this framework, the greater the phonetic-articulatory similarity between two unrelated consonants, the more likely they are to be perceptually assimilated to one another. With respect to determining what counts assimilar, for the present paradigm we use the articulatory phonetic correlates of features.The feature [continuant] in fricatives correlates (roughly) to a narrow constriction that produces continuous turbulent airflow at some location in the vocal tract. This is captured, as described in Section 5.3.1 (Stimuli), by the Harmonics-to-Noise Ratio (HNR) of thesignal, which relates its periodic and aperiodic components. Similarly, the feature [spread glottis] in stops indicates aspiration – the opening gesture of spreading the vocal cords to allow turbulent airflow through the glottis. This also increases aperiodicity in the signal. Given that the glottal vs. oral turbulence distinction, in featural terms a [spread glottis] versus [continuant] distinction, is not present in the Malayalam phonology for labial consonants, the intruding phonetic information, in this case the [continuant] feature of the English fricative, triggers the closest available featural correlate in Malayalam, which is [spread glottis]. Since the feature [voice] is already triggered by the attended signal (and is not contradicted by the competitor), the combination of this with the intrusion from the unattended signal leads to the detection of a voiced, aspirated segment. A (highly harmonic) /v/ in the unattended ear affects the perceived noisiness of the attended phone /b/, and the resultant percept is thus a more highly harmonic Malayalam phoneme, namely /b^h/.

When target and competitor are both voiceless segments, fewer features are involved in the contrast (on our account, the feature [voice] is absent from voiceless stops and voice-less fricatives). We suggested above that this feature sparsity accounts for the relatively lower intrusion rates. However, the nature of intrusions found in voiced and voiceless trials has a unified explanation. Both can be explained by how variation in periodicity (as measured by HNR) across ears can either trigger the feature [continuant] when that feature is relevant for the attended phonology or be assimilated to the closest phonological match in the case when the feature [continuant] is absent. We can identify the closest phonological match with reference to the HNR hierarchy we can extract from figure 5.4 for English and Malayalam, as indicated below:

Ex 4.4.1. HNR hierarchy for English and Malayalam.

1. **English HNRs:** /v/ > /b/ > /p/ > /f/
2. M**alayalam HNRs:** /b^h/ > /p/ > /p^h/ > /b/

For the voiceless target-competitor pair, effects of harmonic pull follow the HNR hierarchy, as they do for voiceless pairs. Thus, for English /f/ and Malayalam /p/, intrusions effect a shift to the adjacent step in the harmonicity hierarchy. In the English-attend mode, this results in /p/ responses to [fata]; in the Malayalam-attend mode, this results in /p^h/ responses to [pata], by which [fata] in the unattended ear triggers activation of [spread glottis] as the closest matching feature in the attended language phonological system. In this way, the phonological inventory constrains participants' available options in choosing a matching phoneme, limiting them to a particular set of grammatical representations.

If an unattended phoneme with a higher harmonic ratio pulls a less harmonic attended phone up the scale of harmonicity, the larger HNR distinction between the voiced pairs(/v/-/b/) will increase the perceived harmonicity of less harmonic Malayalam /b/ to the more highly harmonic /b^h/ when an English /v/ is presented to the unattended ear. On the other hand, the very high harmonicity of the target /v/ is only slightly affected by low HNR of /b/, and thus the perceptual HNR is reduced to the second most harmonic phoneme in English, also a /b/. In this way, intrusion patterns for both voiced and voicelesss timuli can be understood phonetically in terms of harmonic pull.

On the premise that harmonic properties of the signal trigger featural distinctions in the mind, the featural account provided above complements the phonetic analyses. At least for the current case of voicing, a greater number of features in the representation corresponds to larger acoustic differences and greater harmonic pull. There is a greater harmonic difference for voiced stimulus pairs than for the voiceless stimulus pairs. Thus, the sparsity in feature specications for the voiceless stimulus pairs corresponds to smaller differences in HNR. As can be seen in figure5.4, the competing sounds in a voiced target vs. voiced competitor pairing differ greatly in HNR (nearly 20 dB); but in the voiceless trials the HNR difference is much smaller, about 3dB. Thus, while an unattended English /f/, with the lowest harmonicity of the English consonants, causes the highly harmonic Malayalam /p/ to drop to a slightly less harmonic /p^h/, the unattended Malayalam /p/, with a high HNR, increases the perceptual harmonicity of the English /f/ to that of the English stop /p/. These intrusions are in the expected direction but occurred less often than for voiced stimuli.

The feature-based approach to representing voicing and manner contrasts that we have pursued (see figure 5.2) assumes privative features and underspecification. Additionally, in the Introduction, we took the “standard approach” (Lisker and Abramson, 1964; Keating, 1984; Lombardi, 2018) to voicing features as opposed to “laryngeal realism” (Jakobson, 1949; Honeybone, 2005). Thus, we represent the laryngeal distinctions between stops in English with the feature [voice], the same feature that distinguishes /f/ and /v/ in English fricatives and /ph/ and /bh/ in Malayalam. As reviewed in Section 5.2, there is an on-going debate over the proper characterization of laryngeal features in English, centered on whether the contrast between stops is better captured by the feature [spread glottis], the “laryngeal realism approach” or by the feature [voice], the “standard approach” (Hall, 2001). For Malayalam, it is clear that both features are required. The “standard approach” to feature representation was crucial to our interpretation of the intrusion patterns in Malayalam-English bilinguals. Specifically, it allowed us to maintain that phonological features are triggered in perception when the phonetic signal is consistent with those features. When properties of the phonetic signal do not map directly to features in the listener’s attended phonology, they map instead to the nearest phonological category. This mechanism of perceptual assimilation to the nearest phonological category is well-established from research on cross-language speech perception (Best, 1995a; Faris, Best, and Tyler, 2016), L2 speech perception (Best et al., 2007; Bundgaard-Nielsen, Best, and Tyler, 2011), bilingual speech perception (Antoniou, Tyler, and Best, 2012; Antoniou, Best, and Tyler, 2013), and cross-accent speech perception (Best et al., 2015a; Best et al., 2015b; Shaw et al., 2018). Here, we observed its effects in bilinguals during dichotic listening.

While additional research is needed to evaluate the generality of our account, we have presented here an explicit phonological proposal for a population of speakers, Malayalam-English bilinguals, and explained how that proposal dictates behavior in a controlled perception experiment. The dichotic listening paradigm was motivated in part by the connection between this experiment and the real life task faced by bilinguals of balancing selective attention and two rather different sets of phonological contrasts. We are optimistic about the prospect of future work linking phonological theory to a broader range of speech behaviors, both for the potential of theory to guide our understanding of speech behavior and for experimental results to constrain the development of phonological theory.

4.5 Conclusion

In this study we tested adult L2 dominant bilinguals using a task that combines the involvement of selective-attention mechanisms alongside normal signal processing, in a manner that listeners – especially bilingual listeners – are often faced with in situations akin to cocktail-party paradigms. Our tasks presented stimuli from listeners’ L1 and L2, dichotically, and required the participants to undertake a phonemic categorization task for one of the two languages. Consistent with past work using this paradigm, we found that listeners showed a right ear advantage, suggesting left hemisphere superiority in phonological processing. When the stimulus presented in the unattended ear contained a phonological feature absent in the language of the attended ear, that feature was perceptually assimilated to the closest matching feature. Intrusions were more likely for segments containing more distinctive features. These accounts of the results hinge crucially on our particular feature-based analysis of the languages involved, which underscores the potential for experiments such as this to provide new lines of evidence for phonological representations in bilinguals.

4.6 Special acknowledgment

This research was published as a co-authored study by Sayantan Mandal, Catherine T. Best, Jason Shaw and Anne Cutler (Mandal et al., 2020), and was supported by a MARCS postgraduate scholarship to the first author. I will like to earnestly thank Prof. Catherine T. Best for guiding me through my very first experimental study, for holding my hand throughout the design and analyses sections of the study, and also for introducing me to the intersection of phonological theory and speech perception studies. I am also deeply indebted to Prof. Jason Shaw for his immense patience, and the hours he took out of his time to discuss phonological theory with me. Jason was also instrumental in helping me learn the acoustic and statistical analyses skills that turned out to be crucial for this project.

I need to mention, also, the profound impact made by the late Prof. Anne Cutler, both in the pursuing of this project and on my general academic interests. I cannot count the number of hours and days I spent Anne's office, never to be sent away, talking about any number things, academic and otherwise. Her works exploring the role of abstractions in psycholinguistic research were among the very first empirical literature I ever read, and they continue to be the cornerstone of how I approach *biolinguistics* today. *Miss you, Anne.*

I will also like to thank all my colleagues at the MARCS Institutes – you were all way too kind to me, and without your generous support this project would never have been completed. Gratitude is also owed to the participants at the Australian Linguistics Society meeting, where parts of this work were first presented, for their comments, as well as the *Glossa* Associate Editor and three anonymous reviewers, whose comments greatly improved the paper.

Chapter 5

At the Lower Interface-II: Velar Palatalization

5.1 Introduction

We discuss phonological theory (Sapir, 1925; Jakobson, 1973; Chomsky and Halle, 1968; Hale and Reiss, 2008) and the types of psycholinguistic evidence that are usually brought to bear upon it (Best et al., 1994; Cutler and Fay, 1982; Mandal et al., 2020), and attempt to explore the extent to which the relationship has been symbiotic (Cutler, 1992). The input to phonological reasoning is alternations, and here we focus on a particularly contentious one known as *velar palatalization* or *velar softening*. Here we focus on Malayalam, where velar consonants take on a secondary, palatal, articulation, as seen in *makalkkə* (daughter.DAT) ∼ *kuttik*^{j}*k*^{j}*ə* (child.DAT), given certain well-known peculiarities of the process in Malayalam (Mohanan, 1982; Mohanan and Mohanan, 1984). First, velar palatalization in Malayalam is driven by a set of preceding vowels while typically the context is provided by the following vowel (Halle, Vaux, and Wolfe, 2000; Halle, 2005). Second, the set of triggering vowels in Malayalam include two front vowels [i,e] and a (phonetically) low back vowel [a]. The phonological identity of this vowel is dubious. While Warrier (1976) claims that [a] is a low, back vowel, Mohanan and Mohanan (1984, p. 586) argue that for purposes of palatalization this vowel patterns as [-Back]. A further complication stems from the fact the process is claimed to exhibit a "a number of idyosyncratic dialectal and ideolectal variations" (Mohanan and Mohanan, 1984, p. 585), with speakers exhibiting seemingly arbitrary exceptions to palatalization.

From the perspective of phonological theory there are at least two points of contention involved. First, one could ask, given the numerous exceptions noted by Mohanan and Mohanan, 1984, whether this is a productive rule of synchronic Malayalam phonology? We will adopt the I-language perspective (Isac and Reiss, 2013) (see section 6.1.2) and argue that this question reflects a conceptual fallacy about the word "phonology" (or more generally, "grammar")[1]. Second, for phonologists of all varieties the phonological identity of Malayalam [a] represents a core concern regarding the nature of the relationship between *form* and *substance* in cognition (Reiss, 2007; Scheer, 2022; Wilson, 2006; Archangeli and Pulleyblank, 1994). The former relates to the issue of substantive vs. formal motivations for observed surface alternations, while the latter concerns the nature of phonological primes and the extent to which they are determined by substantive information. Here we will adopt a modified flavor of what has been called *substance-free phonology* (SFP) (Fudge, 1967; Hale and Reiss, 2008; Reiss, 2017; Chabot, 2021; Samuels, 2011; Boersma, Chládková, and Benders, 2022), rejecting however the assumptions made by most practitioners regarding the interface with phonetics.

For those wearing a psycholinguistic hat, the phonological issues provide at least three distinct empirical opportunities. First, established psycholinguistic research methodologies (Best et al., 1994; Cutler, 2012; Gleason, 2019) can be applied to probe the productivity of any process

---

[1] To be specific, the notion of 'dialectal' or 'idiolectal' *exceptions* imply that there is some, mind-external, notion of 'grammar' from which either certain dialects, or some idiolects, are exceptions. In contrast, the I-language perspective assumes that each speaker is a unique model of grammar, and one that needs to be studied in its own right. It is important to note that such perspectives also surface in the psycholinguistic literature (Levelt, 2013; Cutler, 1992). For a recent discussion on converging perspectives between the generative program and psycholinguistics the reader is referred to the discussion by Scheer and Mathy (2021)

using nonce words. Second, asymmetries observed in such behavioral tasks involving phonological alternations are to be explained through the division of labor between postulated phonological entities and psycholinguistic processes involved in parsing of speech stimuli from raw signal (Lewis and Phillips, 2015; Cutler, 2017; Best, McRoberts, and Goodell, 2001). Finally, and perhaps most crucial to our purposes in this article, linguistically motivated empirical studies provide an opportunity to probe the relationship between two distinct objects of study, those belonging to Linguistics and Psycholinguistics respectively (Cutler, 1992; Lewis and Phillips, 2015) – Are they models of different phenomena, or different models of the same phenomena? Given the differences stemming from conceptual granularity mismatch and ontological incommensurability (Poeppel and Embick, 2017; Poeppel, 2012; Poeppel and Assaneo, 2020) how can theories from these distinct domains be best aligned?

The following sections elaborate on these issues in the following manner. Section 5.1 discuss the formal issues concerning phonological theory and its architecture. We elaborate on certain conceptual issues stemming from the cognitive revolution in the 1950s and 1960s (Chomsky, 1959; Mandal, 2021), and provide some clarifications regarding theoretical and meta-theoretical terminologies and how they are used in the formal literature (Reiss, To Appear; Samuels, 2011; Dresher, 2014; Chabot, 2021). This section establishes the basic framework within which we discuss velar palatalization, both, in Malayalam and as a general phenomena. Section 5.2 provides a brief overview of the psycholinguistic and laboratory phonology studies of velar consonants and cross-linguistic palatalization phenomena. We focus on how the conceptual terms discussed in section 5.1 are appealed to in this literature, and suggest some revisions. Section 5.3 introduces our experimental design, describes the stimuli materials and methodological choices, and provides a detailed description of the phonetic and statistical behavioral analyses performed. Section 5.4 provides a general discussion of the results, with a focus on our primary motivation of attempting to align the psycholinguistic accounts of perception and production in Malayalam with a phonological account of representation and rules in Malayalam. Particularly, it argues for a modified version of Substance-Free Phonology coupled with a Bayesian interface with phonetics, in a strictly modular (Fodor, 1983; Pylyshyn, 1980) framework that maintains partial veridicality between features and their phonetic interpretations.

5.1.1 Language and linguistic theory: the "sound" domain

We pointed out in the previous section that this section will discuss a process seen in many *languages*, called velar palatalization, and that we will discuss this issue in light of data from one specific *language* called Malayalam. This raises two imminent questions; What is Language (as a natural object), or conversely what could never be a (specific instantiation of) natural language (Chabot, 2021)? What can some phenomena in Malayalam tell us about other languages, and if anything at all then why (Reiss, To Appear)? In series of publications Reiss (Reiss, 2003; Reiss, 2007; Hale and Reiss, 2008; Isac and Reiss, 2013; Reiss, To Appear) has attempted to clarify this issue. Reiss (To Appear) points out that in order to study linguistic processes we first need to know what makes something, to the exception of everything else, count as a language? That is, ontology aside, anyone interested in understanding Language needs a parts-list, of sorts. This parts-list will contain the necessary properties and principles of organization necessary to talk about languages in general, to the exception of things that are not languages.

> The suggestion to treat UG as a background assumption, a postulate, is partly a rhetorical device designed to flip the null hypothesis. Instead of feeling pressure to prove the hypothesis of UG, linguists can show that it is only by assuming UG that we have achieved the sophisticated results of recent decades. We can again appreciate the parallel to Newton. The assumption that the same force was at work on apples and planets led Newton to the inverse square law (the gravitational force between two objects varies inversely with the square of the distance between them) and the notion of centripetal force. Having worked out such details with mathematical rigor, Newton could see that his postulate led somewhere good, and he was

> then able to formulate an articulated theory of Universal Gravitation. The proof of the value of the postulate was in the proverbial empirical pudding.
> **(Reiss, To Appear, p. 3)**

This background assumption, Reiss terms UG-object, a universal hypothetical parts-list that forms an equivalence class[2], Natural Language. The term UG, or Universal Grammar, is both ambiguous and controversial, and assumes the existence of entities or categories that specific to language.

> We want to make a theory over a coherent domain of phenomena. Newton postulated that apples and planets fall into a single domain for his purposes, and linguists think that French and Swahili, but not chess, belong together. As discussed above this process of defining domains involves postulating an abstract set of properties that, for mechanics, say, includes mass, but not the Granny Smith vs. Golden Delicious distinction. [...] linguists obviously need a UG-theory containing a universal set of linguistic entities and operations of various sorts because they are trying to model human languages which are built from a universal set of linguistic entities and operations of various sorts. These are the things in the world that make languages and only languages form a natural class of entities. In other words, the theory and the object should be isomorphic. **(Reiss, To Appear, p. 7)**

In the generative(Chomsky and Halle, 1968; Halle, Vaux, and Wolfe, 2000; Reiss, 2017) and biolinguistic literature (Samuels, 2015a; Samuels, 2009; Hornstein and Boeckx, 2009) the axiomatic assumption is that UG-object, also called *Faculty of Language*(FL), is genetically endowed (Hauser, Chomsky, and Fitch, 2002; Fitch, Hauser, and Chomsky, 2005) and present at birth in an *initial state*, or S_0. Given this state the child is capable of parsing any natural language, and in fact it enables her to tell linguistic stimuli apart in the first place. This line of reasoning is

---

[2] In the mainstream literature the term *natural class* is more commonly used. We will stick to mathematical term given the controversial status of the word "natural" in phonological theory

derived from the information theoretic hypothesis that any information processing system can only parse informtation in terms of a representational language inherent to itself.

> Given that an *Information Processing System* (IPS) cannot process data except in terms of whatever representational language is inherent to it, data could not even be apprehended by an IPS without becoming representational in nature, and thus losing their status of being raw, brute, facts. The representational language of the IPS provides the categories in terms of which the IPS 'views' reality[...] **(Hammarberg, 1981, p. 261)**

In this sense, then, S_0 provides the child with the representational language with which to parse linguistic experience. As the child develops the ambient data gradually narrows down the range of possibilities until FL has reached a certain mature stage, call it S_n. This system of rules is implicit in the native speakers' mind/brains, and are often called *mental grammars*, or in Chomskyan terms, I-language. This notion of mental faculties interacting with environment to instantiate grammars will be relevant to our discussion of Malayalam, so we provide some brief clarifications below.

5.1.2 I-Language

The term I-language is often misunderstood by researchers, including by philosophers and psychologists. Reiss (To Appear) aptly points out the error in Green and Vervaeke's (1997) attempted elaboration of their understanding of the notion.

> There is a systematic ambiguity in the term "grammar," an ambiguity that has caused much confusion among Chomsky's critics. It refers both to the knowledge that is hypothesized to be "in the head" of the human, and to the linguist's theory of that knowledge. In an effort to stem the confusion somewhat, Chomsky has recently adopted the term "I-language" to refer to the knowledge of language the human is

> thought innately to have, the "I" standing for individual, internal, and intensional. Once again, the point here is not to reject the possibility that there are, by contrast, social, external, and extensional aspects to language. It is only to identify those features of language that Chomsky is interested in addressing.
> **(Green and Vervaeke, 1997)**

Reiss' (To Appear) criticism is targeted at the authors' equating I-language to the hypothesized innate component leading, perhaps, the reader to construe I-language as *I(nnate)-language*. This is a commonplace confusion, albeit it is all the more interesting given that the excerpt cited above does contain the technical definition of what the *I* in I-language is intended to imply – individual, internal, and intensional. *Not*, however, innate, even though the generative program is inherently nativist (Chomsky, 1956; Chomsky, 2002; Chomsky, 2005). Rather, I-language is simply a convenient shorthand for a full description of the nature of linguistic knowledge as instantiated in the mind/brain of a naive speaker. It is individual because specific languages, such English or Bangla, must be learned through experience, of which each individual person must undergo their own. If we admit that experience is necessarily individual, then so is the knowledge of language that is shaped by experience. Such knowledge is also *internal*, since they exist in the mind/brain of the individual. Finally, such knowledge is *intensional*[3] because it is comprised of primes and rules (or constraints) governing their organization. Consider, for instance, that Sam's favorite days of the week are *Thursday* and *Sunday.* We could simply capture this fact as a list – Sam's Favorite Days: Thursday, Sunday. This is called an extensional characterization, and contrasts with a set-theoretic intensional characterization – F is the set of days of the week, such that x is in F if and only if x is one of Sam's favorite days. The extensional works for smaller lists, but with increasing number of items in the list such formulations become unwieldly. Further, suppose Sam later changes his mind and decides he likes Fridays

---

[3] This term is to be understood purely in the mathematical sense. This is not the same kind of intension as when Chomsky opines, *"Intensions are irrelevant, what matters is the effect of our actions and inactions."*

better than Thursdays, perhaps because the weekends are soon to follow. In this case, an extensional architecture will require creating a new list where 'Thursday' is replaced with 'Friday'. In contrast, the intensional characterization captures this change by virtue of its structural description. While the concept is usually associated with the generative program the underpinning idea itself does not necessarily have to entail nativism. As Reiss (To Appear) points out (emphasis added):

> **One could, in principle, be a psychologist who believes that humans have mental grammars, without believing in UG-object.** One could potentially believe that mental grammars arise through "general learning mechanisms" without any language-specific categories. So belief in I-language does not entail belief in UG-object. On the other hand, nativist linguists must believe in the I-language perspective, because UG-object, and all subsequent states of the language faculty must be understood as individual—they are part of each person; internal—they are represented in the mind/brain; and intensional in the set theoretic sense—they consist of rules, patterns or functions, rather than lists or look-up tables, since the number of sentences of a language is unbounded; the grammar cannot extensionally characterize the set of sentences of the language

Indeed, connectionism (Rumelhart, Hinton, McClelland, et al., 1986; Buckner and Garson, 1997) is claimed to be one such framework where no domain-specific categories or prior principles are allegedly necessary. In such frameworks one can still have an I-language comprising of, for instance, arbitrary variables, randomly set connection strengths and activation values. As Buckner and Garson (1997) point out in such systems, while they contain built-in priors, the learning capacity of the system is not constrained in any particular sense. We will refrain from taking a position on the *innateness* debate in this section, but we will reiterate the usefulness of adopting the I-language perspective – the assumption that in the mind/brain of a native speaker inhabits a kind of implicit knowledge of language that is productive and acquired rapidly over a critical period with minimal amount of exposure. Ontological committements notwithstanding this much should be acceptable to any linguist or psychologist interested in the general process of language acquisition, perception and production.

Once we adopt such a perspective the answer to the question of what any language, say Malayalam, is essentially reduces to *'a set of mutually intelligible I-languages'*. This perspective, too, is not very controversial. Most of us already recognize dialectal differences, and assume they form the "same" language because they are usually mutually intelligible to various degrees. The I-language perspective simply pushes this gradient scale of variations further inward, and stresses for an appreciation of idiolectal differences even when they are superficially invisible, or does not reach some notion of statistical significance. In this case, then, there really is no coherent notion of *the* Malayalam grammar, just a set of multiple Malayalam-type I-languages. This perspective offers our psycholinguistic investigation two benefits. First, instead of worrying about the dialectal and idiolectal exceptions to palatalization noted by (Mohanan and Mohanan, 1984) we treat each participant as data point instantiating a Malayalam-type I-language, and ask to what extent their individual grammars (I-language, state S_n) employ palatalization of velar consonants as a productive strategy? Second, having established an empirical coverage of velar palatalization over a controlled population of speakers, we next ask the question of interest to any working psycholinguist – how do we characterize Malayalam-type I-languages such that these idiolectal variations fall out of the general architecture applied to these I-languages that lends itself to the notion of *'Malayalam'*. In other words, while notions like *'psychological reality'* and *'I-languages'* are of ontological and methodological significance to (psycho)linguists, affording fine-grained appreciation of idiolectal variations, a complete psychological account, in our view, also must elaborate on the unifying qualities that lend itself to the notion of a language, here Malayalam, as a broad socio-cultural umbrella [4]

---

[4] See Wilson, 2000 on psychology, sociology and culture being derived, in that order, from basic biology.

5.1.3 Malayalam

The empirical coverage of the theoretical postulations discussed in this paper comes, primarily, from Malayalam – a Dravidian language spoken in the Indian state of Kerala, and the union territories of Lakshadweep and Puducherry (or, Pondicherry) by an L1 population of around thirty-seven million 'Malayali' people [5]. The language has generally attracted considerable attention from a wide variety of linguists for both its historical roots and elaborate phonological and phonetic systems.

As is typical of almost all Dravidian languages, Malayalam boasts a long tradition of philological and linguistic inquiry (Ramaswami, 1936; Shanmugam, 1976; Caldwell, 1998). Given our focus on the psycholinguistic aspects of speaking Malayalam, however, we will restrict our discussion to the generative (Mohanan, 1982; Mohanan and Mohanan, 1984; Schein and Steriade, 1986) and psycholinguistic (Mohanan and Mohanan, 1986; Van Oostendorp, 2006; Wilson, 2006) literature that focuses on the nature and acquisition of linguistic knowledge. Mohanan, 1982 provides the most systematic survey of Malayalam, pointing out that the "consonant system of Malayalam is of particular interest to phonologists" given its "rare seven-place articulation contrast in stops and nasals" (p. 575). Mohanan (Mohanan, 1982; Mohanan and Mohanan, 1984) reports that the seven places of articulation are, respectively, bilabial, dental, alveolar, retroflex, palato-alveolar, palatal, and velar. The forms in a-e below illustrate the surface contrasts. Given such a system Mohanan and Mohanan, 1984 rightly points out that the pertinent issue is just how many of these surface contrasts are present in the underlying lexicon?

---

[5] cf. *Malayalam* at Ethnologue.

a.	kuppi ’bottle’; kut̪t̪i ’stabbed’; kutti ’stump’; kuʈʈi ’child’; mucci ’face’; mukki ’sank’
b.	appam ’bread’; at̪t̪am ’constellation’; attam ’end’; at̪tam ’attic’, akkam ’number’
c.	maraccu ’covered’, marakʲkʲu ’cover-imp’
d.	kammi ’scarce’; pan̪n̪i ’pig’; kanni ’a mouth’; kan̑n̑i ’rice soup’; kaɳɳi ’link’; maŋŋi ’dulled’
e.	man̑n̑al ’turmeric’; mattaŋʲŋʲa ’pumpkin’; maŋŋal ’dullness’

TABLE 5.1: Surface place contrasts in Malayalam

Mohanan and Mohanan, 1984 argue that reduction in underlying representational complexity can be achieved by focusing on the distribution of velar and palatal consonants that suggest that the latter be derived from the former by derivational rules. Consider the distribution of the dative, causative and verbalizing forms in the table below. Based on this Mohanan (1984; 586) proposes that the palatal consonants — kʲ, kʰʲ, gʲ, gʰʲ, ŋ ʲ – are derived from the respective velars by the application of a fronting rule.

Ex 5.1.1. Malayalam velar palatalization rule.

[-Cont,+High] → [-Back] / [+Syl,-Back] __

Dative		
Nominative	**Dative**	**Gloss**
makal	makalkkə	’daughter’
kutti	kuttik^{j}k^{j}ə	’child’
Causative		
Base	**Causative**	**Gloss**
pilar	pilarkk-	’split’
mara	marak^{j}k^{j}-	’cover’
Verbalizing		
Noun	**Verb**	**Gloss**
wilarppə	wilarkk-	’paleness’
wira	wirak^{j}k^{j}-	’trembling’

TABLE 5.2: Distribution of Velar and Palatals

Consider, for instance, the base and causative forms of ’cover’ – mara and marak^{j}k^{j}-. The preceding stem-final /a/ causes the geminate velar (kk) to undergo palatalization (k^{j}k^{j}). In contrast, the stem of ’split’ ends in a consonant (pilar) and thus the suffix initial velars surface unpalatalized (pilarkk-). However, such a formulation encounters three major problems by the authors’ (Mohanan and Mohanan, 1984, pp. 586–587) own account. For one thing the process has numerous lexical exceptions at both dialectal and idiolectal levels. Consider the forms in the table below. Mohanan (1984; 586) points out that there is no systematic distinction between the forms in the left and right columns, with the authors’ own speech manifesting these contrasts.

a. Velar	**b. Palatalized Velar**
wikalam ′broken′	mik^{j}acca ′excellent′
wikkan ′stammerer′	mik^{j}k^{j}a ′most′
tikkə ′crowd′	atik^{j}k^{j}ə ′beat-imp′
nekkal ′squeezing′	wek^{j}k^{j}a ′cooking′

TABLE 5.3: Exceptions in Velar Palatalization

Given the lack of any systematicity the forms on the left leave little recourse to the analyst but to list them as ′lexical exceptions′. Next, the authors note that velars in non-geminate clusters never palatalize.

Ex 5.1.2. wikramam ′brave deed′
tiktam ′bitter′
kiŋaram ′follower′

Given this constraint the authors suggest revising the rule to apply to single-melody velars only.

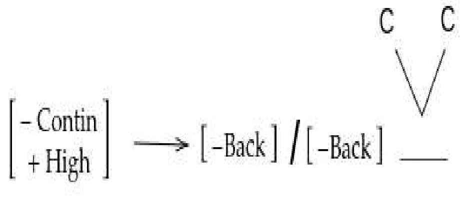


FIGURE 5.1: Malayalam velar paltalization rule revised for single-melody velars

Finally, the morphological environment in which the rule applies/fails is also contentious. Mohanan and Mohanan (1984) point out that while the stem-final /i/ triggers palatalization with the dative suffix, the same vowel fails to trigger palatalization in the context of the plural suffix.

Ex 5.1.3. kutti ’child’ kuttikal ’children’
pu:cca ’cat’ pu:ccakal ’cats’

To summarize the velar palatalization process applies in monomorphemic words (but with numerous lexical exceptions (see table 6.2), across suffix boundaries with the dative, causative and verbalizing suffixes (but fails with the plural suffix), and applies to geminates but never to non-geminate clusters. For the authors this distribution, when considered within the confines of Lexical Phonology, implies that the rule must apply in the lexicon itself. However, this still requires marking numerous lexical entries as "exceptions", an ad hoc solution at best. Given this conundrum the following sections first evaluate alternative theoretical frameworks within which the paradigm can be analysed, and then considers empirical psycholinguistic evidence weighing on whether such a process can at all be considered phonological in a meaningful sense.

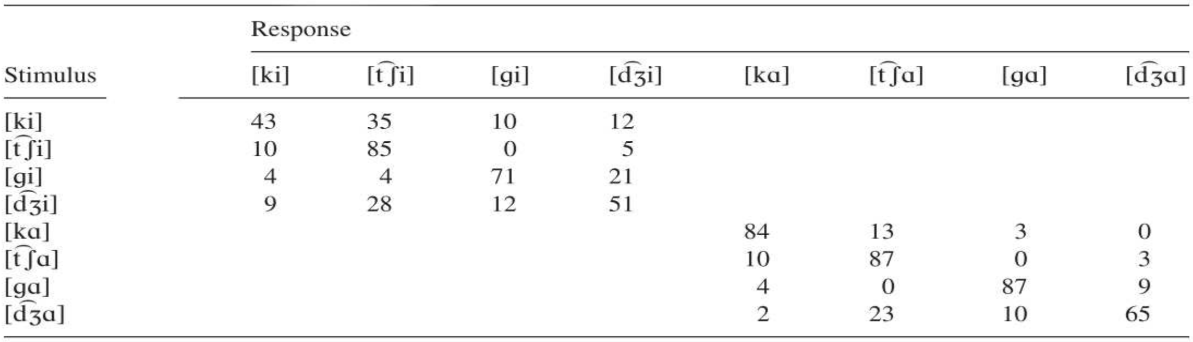
5.2 Psycholinguistic approaches to palatalization

Palatalization, as a cross-linguistic phenomena, is rather well-studied from both phonological (Halle, 2005; Wilson, 2006) and speech processing (Jacobs and Berns, 2012; Kochetov and Alderete, 2011; Rubach, 2003) perspectives. In particular, the acoustic, articulatory and perceptual properties of velars and palato-alveolars have been extensively documented. As a general rule, velar consonants are articulated further forward when they appear before front vowels. Often discussed under the rubric of anticipatory co-articulation, this phenomena is stable enough that Keating and Lahiri, 1993 conclude that the *“more front the vowel, the more front the velar”*. One would, of course, expect this pattern of co-articulatory effects to reflect in the acoustic properties of the externalized forms, and indeed Keating and Lahiri, 1993 confirm this. The more fronted a vowel’s articulation is, the smaller the resonant cavity gets, resulting in higher frequency peaks. The peak spectral frequencies of velar releases are directly proportional to the relative frontedness of the contextual vowel (Guion, 1996; Guignard Guion, 1998).

In comparison, the palato-alveolar affricates display spectral frequencies that are (a) on average higher than those of velars, and (b) remain roughly constant across the frontedness of the contextual vowels (Guion, 1996). Guion's (1996) book-length treatment extends the articulatory and acoustic findings to predictable behavioral evidence from speech perception studies. Under experimental conditions native listeners tend to confuse velar stops with palato-alveolar affricates more as the frontedness of the following vowel increases. In other words subjects are more likely to confuse [ki] with [ʧi] (or [kʲi]), than [ka] with [ʧa] (or [kʲa]), and least likely to confuse [ku] with [ʧu] (or [kʲu]). It is worth noting here for explicatory purposes that velar palatalization itself may refer to one of two different types of alternations – the first involves an underlying velar consonant (e.g. /k/ → [ʧ]) while the second involves a velar consonant taking on a secondary palatal articulation (e.g. /k/ → [kʲ]). For the purposes of this article we will restrict our discussion to the latter type of palatalization since this is the one attested in Malayalam. In the next few sections we first outline the behavioral and neurobiological psycholinguistic investigations that have reported on velar palatalization, and then outline two pairs of behavioral studies conducted with adult early/simultaneous Australian English - Malayalam bilinguals.

5.2.1 Perception

Perceptual studies with native speaker-listeners have provided a wealth of evidence highlighting increased confusability between non-palatalized and palatalized consonants before front vowels (Guion, 1996; Guignard Guion, 1998; Wilson, 2006). Guion (1996, 1998), in particular, report multiple experiments confirming this claim. For instance, in one of the experiments involving forced-choice methodology native speakers of English made noticeably more identification error in the presence of masking noise. The stimuli comprised of CV sequences comprising of four consonants ([k, ʧ, d, ʤ]) and three vowels (([i,a,u]), excised from fast running speech and then truncated to ensure that the vowel length was always 100ms. The stimuli were presented to participants under two conditions, normal (without any masking noise) and in the presence of a +2dB signal-to-noise white masking noise. Consider the table below, reproduced from a discussion of Guion by Wilson, 2006, that shows the pattern of confusion matrix in the concerned experiment.



	Response							
Stimulus	[ki]	[t͡ʃi]	[gi]	[d͡ʒi]	[kɑ]	[t͡ʃɑ]	[gɑ]	[d͡ʒɑ]
[ki]	43	35	10	12				
[t͡ʃi]	10	85	0	5				
[gi]	4	4	71	21				
[d͡ʒi]	9	28	12	51				
[kɑ]					84	13	3	0
[t͡ʃɑ]					10	87	0	3
[gɑ]					4	0	87	9
[d͡ʒɑ]					2	23	10	65

FIGURE 5.2: From "Learning Phonology with Substantive Bias", by Colin Wilson (Wilson, 2006), pp. 949

It can be readily observed from the table above that the rate at which [ki] is misidentified as [ʧi] is much higher than the rate of confusing [ka] with [ʧa]. Similar results are obtained with the voiced consonants – [gi] is more liable to be confused with [ʤi] than [ga] with [ʤa] [6]. Overall, however, the presence/absence of masking noise played an important role in Guion's study, with the rate of confusion being 31% in the presence of masking noise compared to just 5% in the absence of noise. As pointed out by Wilson, 2006(pp.949) such confusion patterns are readily explained in terms of perceptual and articulatory (reviewed below) effects well-known from empirical studies.

5.2.2 Production

The articulatory/production based evidence of palatalization also largely vindicates the role of front vowels. Butcher et al., 2004 provide an overview of the articulatory and acoustic underpinnings of dorsovelar consonants from several Australian aboriginal languages, and compare

---

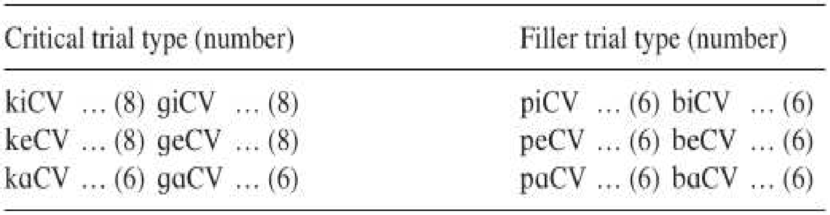
[6] Although the relative rate of confusion is higher for the voiceless pairs. However, pure voicing-confusion (confusing [ki] with [gi]) was rare.

this data with English. Direct palatography and F_2 transition based evidence show that Australian languages have distinct velar targets for each vowel. In particular, the authors conclude (pp. 49) that "allophones of Australian velars before non-front vowels are articulated further back on the soft palate than the corresponding allophones in English". In general, the authors point out, that Australian languages display a much tighter and co-ordinated co-articulatory effect with the following vowel than English. For immediate interest to us, of course, is the finding that velars are articulated in a fronted fashion in the context of fronted vowels and further back when the contextual vowel is non-front. Likewise, Keating and Lahiri, 1993 review X-ray and other articulatory evidence of such co-articulation in English and other languages, and report that the degree of frontedness in articulation of velars is directly proportional to the degree of frontedness of the contextual vowel, and also that this generalization "holds in all of the languages for which data were available" (pp.89). This is relevant, especially in light of the evidence discussed in the immediately preceding section, because the more fronted the articulation of a velar consonant the greater the chance that it will be confused with its palatalized counter-part. This is, of course, to be expected given the tight link between articulatory configuration and the resultant acoustic cues associated with a phoneme (Lenneberg, 1967).

5.2.3 Acquisition

The articulatory changes induced in velars in the context of front vowels give rise to pronounced substantive cues. Keating and Lahiri, 1993 point out that spectrum peak immediately following a consonant's release is due largely to resonance in the oral front cavity, and that the frequency of this resonance is largely a function of the following vowel. When followed by a front vowel the velar, likewise, is articulated in a fronted manner. This has the direct result of reduction in size of the resonance cavity leading to higher frequency values. One issue of theoretical importance here is the extent to which such substantive information is used during the process of acquisition to induce generalizations about the grammar being acquired? Wilson, 2006 provides one of the more recent empirical reports on the role of substance in acquisition of abstract patterns by native listeners. In Wilson's (2006) study two experiments are conducted to test, respectively, asymmetry in direction of generalizations (Experiment 1, pp. 961) – are learners exposed to palatalization in the context of [i] and [e] separately more likely to extend the generalization in one direction than the other – and learners' tendency to generalize palatalization in the context of one consonant (say, [ki] → [ʧi]) to another consonant and vice versa (e.g. [gi] → [ʤi]; pp. 969). In the first experiment, using a novel language game, the native American English speaking participants were divided into two groups for two experimental conditions (mid and high) and presented with recorded examples of pairs of words. The participants were informed that the words were from an imaginary language and did not represent any real languages. Participants in the "high condition" were presented with pairs such as [kinə] ... [ʧinə], with "...' representing a brief pause. Participants in the "mid condition", likewise, were exposed to pairs such as [kenə] ... [ʧenə]. In each condition, therefore, participants were being exposed to examples of velar palatalization in the context of either high or mid vowels. Crucially, participants in each condition were never presented with words representing velar palatalization in the context of the "other" vowel – "mid condition" participants never heard [kinə] ... [ʧinə] and vice versa. Wilson (2006; 962) argues that his "method deliberately withholds crucial information—in this case, whether palatalization applies in the novel context—and thereby forces participants to rely on their ability to generalize from limited exposure to a new phonological pattern" and should be referred to "as the poverty of the stimulus method (PSM)". Both groups of participants were then presented with words containing the velar followed by the vowels [i,e,a], in separate trials, which they had to pronounce out loud. Wilson (2006; 962) notes that for each participant only one vowel would have conditioned velar palatalization, and any extension of this generalization in either backward ([i] → [e]) or frontward ([e] → [i]) direction would imply the generalization being extended to novel conditions. Given that velar stops and palato-alveolar affricates are more perceptually similar before [i] than before [e], and that palatalization before [e] asymmetrically implies palatalization before [i] in most attested languages, if participants have a system of substantively biased phonological generalization then one would expect more generalization of palatalization in the "mid condition" than the "high condition". On the other hand, if participants do not have such a system then one would expect generalizations being extended in both directions without any asymmetry. Wilson's (2006, pp. 967-968) results largely appear to conform to the substantive and typological [7] predictions, with participants in the "mid condition" extending palatalization in the context of the high front vowel [i] based on experience in the context of the mid vowel [e], but not the other way around.

In a second experiment (Wilson, 2006, p. 969) the author tests whether experience with palatalization of one velar stop, say [k], is extended to the other stop [g]. The experiment was run in two conditions – voiced and voiceless. In a practice run subjects were exposed to three trials, with one trial containing word pairs that were phonologically identical and two in which they were related by velar palatalization (e.g. Voiceless: [kiwə] → [ʧiwə], Voiced: [gipə] → [ʤipə]). The exposure phase had thirty-four trials, with the trials being grouped into four blocks. Each block contained two examples of velar palatalization ([k] → [ʧ] or [g] → [ʤ] before [i] and [e]), one or two examples of velars that did not palatalize ([k] or [g] before [ɑ]), and four or five fillers. The testing phase consisted of a total of eighty trials shown in the figure below. The stimuli items used and response collection from the subjects followed protocols similar to the first experiment.



Critical trial type (number)	Filler trial type (number)
kiCV … (8) giCV … (8)	piCV … (6) biCV … (6)
keCV … (8) geCV … (8)	peCV … (6) beCV … (6)
kɑCV … (6) gɑCV … (6)	pɑCV … (6) bɑCV … (6)

FIGURE 5.3: Test Block. From "Learning Phonology with Substantive Bias", by Colin Wilson (Wilson 2006), pp. 965

---

[7] Cross-linguistically languages that exhibit palatalization in the context of mid-vowels also exhibit palatalization in the context of front vowels.

Wilson, 2006, pp. 970–971 reports two main within-participant effects – consonant (exposure vs. novel) and vowel context (high-front [i] vs. mid-front [e]). However, there was no statistically significant effect of condition suggesting that the subjects did not respond with different rates based on exposure conditions. Further, a lack of any significant effect of consonant-condition interaction suggest that participants do not extend generalizations about velar palatalization from [g] to [k] at a higher rate than in the other direction. While overall both generalization rates were lower than those observed in the first experiment, Wilson (2006) concludes that the results of the second experiment suggest that there is no significant effect of voicing of the consonant on generalizations about velar palatalization.

5.2.4 Form and substance in palatalization

A review of the psycholinguistic literature provided in the preceding section would immediately highlight that the results are directly attributed to phonology itself. For instance, Wilson, 2006 attributes the asymmetries in his artificial language learning experiment directly to phonological grammar, concluding that phonology is substantively biased. But it remains unclear what motivations, if any, could exist for importing such phonetic motivations within the formal grammar. Chabot, 2021 provides an in-depth book-length review of phonological alternations, focusing especially on the so-called *"crazy rules"* – rules that lack any phonetic motivation defined in terms of "naturalness" (Mielke, 2008). For instance, a certain environment, say intervocalic consonants, (VCV) might very well be a well-known context of alternation (e.g. [-Syl,-Voice] → [+Voice] / V_V) but in a given language produce an output that is phonetically unlikely (e.g. /l/ → [ʁ] / V_V) (Chabot, 2021, p. 86). Chabot provides a detailed overview of crazy rules across a wide variety of languages, but of particular interest for our discussions is the fact that there exist little to no phonetic motivation for the structural change effected by the rule:

> [...] either in that the input and the output are unrelated, or the output and the triggering context are unrelated. In the first case, an input segment X is realized as Y, where X and Y diverge unexpectedly in manner, voicing, place of articulation, or some combination thereof. In the second case, a change X → Y is conditioned by an environment Z such that Y and Z diverge unexpectedly along the same axes.

Chabot, 2021 points out that the apparent craziness of so-called crazy rules stems not from some characteristics inherent to the rules themselves, but rather the pre-theoretical assumption that phonology itself is phonetically motivated, or natural in some sense. If we let go of this naturalness, then crazy rules are not so much crazy, as they are rare. Chabot, thus, argues that crazy rules are not born crazy, but become so through diachronic processes. They are derived from rules that started out as phonetically plausible, but through a sequence of individual diachronic changes they steer further and further away from their initial functional motivations. Chabot, 2021, p. 161

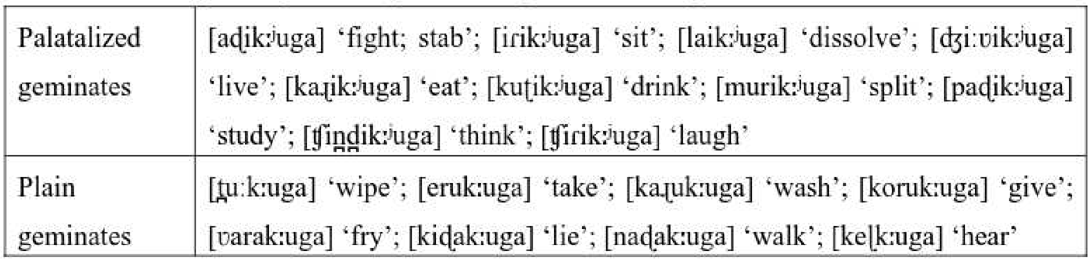
> Since phonology is isolated from phonetic influence, once phonologized, the structural description of a phonological rule is no longer sensitive to the phonetic context at its origin—rules are no longer yoked to phonetic exigencies. If a further change to the target, trigger, or conditioning environment occurs, the rule may loose its natural description while continuing to operate.

The reader might wonder here what is the relevance of a discussion of crazy rules in a behavioral psycholinguistic study of velar palatalization? The relevance, of course, lies in Chabot (2021)'s discussion of phonology and naturalness itself. Recall, from our discussion of Wilson, 2006's *implicational laws of palatalization* that typological and behavioral trends in attested patterns of palatalization are usually argued to be evidence for substantively-biased grammars. But by Wilson, 2006's own account phonological grammars are nonetheless able to overcome this alleged bias, and palatalize in the context of unnatural vowels. Just as they are able to overcome naturalness and compute crazy rules. So, it behooves the theorist to ask, if crazy rules are not really crazy even if they might seem so due to various non-phonological factors then might substantively-biased "grammars" also not seem so for extra-grammatical reasons? In other words, if it is possible to show that Wilson, 2006's "substantive biases" in acquisition can be accounted for outside of grammar, then perhaps the grammar itself is not biased at all. Rather, explanations for such bias in attested trends in acquisition studies should be sought elsewhere, perhaps in third-factors (Samuels, 2009).

The next section describes two psycholinguistic studies of Malayalam velar-palatalization, and illustrates that asymmetries of the sort reported by Wilson, 2006 are indeed proliferate in *performance*. The following sections, then, attempts to account for such asymmetries, without importing substantive biases into the grammar itself.

5.2.5 Two behavioral investigations

The previous sections have provided an overview of a cross-linguistic pattern of alternation known as velar palatalization, or simply palatalization, and discussed a particularly intriguing pattern of palatalization from the Dravidian language Malayalam (Mohanan and Mohanan, 1984). Of particular interest, as concerns the implications of Malayalam palatalization for phonological theory, are the facts that (a) in Malayalam palatalization is conditioned by preceding vowels ([i,e,a]), as opposed to the typologically prevalent environment of following front vowels, and (b) the numerous "lexical exceptions" to the process reported by Mohanan and Mohanan, 1984, pp.585 that seemingly defy dialectal or idiolectal generalizations (see Table 5.3). Yet, as reported by both Mohanan and Mohanan, 1984 and Jiang, 2010, the process exhibits productive application, especially across certain morpheme boundaries. Recall from section 1.2 that the initial velars in the dative, causative and verbalizing suffixes routinely palatalize (table 5.2), while the plural suffix never does (Ex. 5.1.3). Further, Jiang, 2010, pp.23 reports that the velars in the suffix -kkuka][8] also routinely undergoes palatalization when preceded by a front vowel.



Palatalized geminates	[aɖikːʲuga] ‘fight; stab’; [iɾikːʲuga] ‘sit’; [laikːʲuga] ‘dissolve’; [ʤiːʋikːʲuga] ‘live’; [kaɻikːʲuga] ‘eat’; [kuʈikːʲuga] ‘drink’; [murikːʲuga] ‘split’; [paɖikːʲuga] ‘study’; [ʧin̪d̪ikːʲuga] ‘think’; [ʧiɾikːʲuga] ‘laugh’
Plain geminates	[t̪uːkːuga] ‘wipe’; [erukːuga] ‘take’; [kaɻukːuga] ‘wash’; [korukːuga] ‘give’; [ʋarakːuga] ‘fry’; [kiɖakːuga] ‘lie’; [naɖakːuga] ‘walk’; [keɭkːuga] ‘hear’

FIGURE 5.4: Velar From "Malayalam: a Grammatical Sketch" by Haoway Jiang (Jiang, 2010)

Further, to our knowledge there exist no minimal pairs that contrast only in terms of palatalized and non-palatalized velars. The lack of such minimal pairs, coupled with the systematic triggering of palatalization in the context of a subset of suffixes while simultaneously not applying in the context of the plural suffix, makes the existence of a selectively productive rule a possible scenario. While the existence of the "lexical exceptions" in monomorphemic words, mentioned by Mohanan and Mohanan, 1984 (see table 6.2) obviously requires a deeper investigation in light of the Malayalam consonant system, the systematic palatalization of certain suffixes offers an opportunity to empirically probe productivity across morpheme boundaries. In order to evaluate this we conducted two pairs behavioral studies with early/simultaneous bilinguals, with the first experiment targeting productivity in perception while the second tests productivity in production.

The primary motivation behind our study is to compare the predictions of two contrasting hypotheses about Malayalam velar palatalization to the results obtained from psycholinguistic experimentation. As discussed in the previous sections, there exist good reason to believe that substantive factors play an important role in shaping palatalization processes. For instance,

---

[8] Note that the second velar in this suffix regularly undergoes intervocalic voicing and surfaces as -kkuga]

given the evidence provided by Keating and Lahiri, 1993 and Guion, 1996 it is entirely possible that Malayalam velars undergo palatalization due to coarticulatory effects. If, in fact, this is the case then there exist little reason to invoke phonological grammar to account for the same. This line of reasoning that is adopted by substance-free phonology (Fudge, 1967; Hale and Reiss, 2000b; Scheer, 2022; Chabot, 2022a; Samuels, 2011) that holds that phenomena that can be explained by phonetic motivations should not be imported into phonological grammar. Proponents of substance-freedom argue that doing so is problematic on, at least, two counts. For one thing it constitutes a basic premise of modularity, namely informational encapsulation and proprietary vocabulary of modules (Hammarberg, 1981; Scheer, 2010a). According to this view, related closely to Marr's tri-level framework (Marr, 1977; Marr, 2010), cognition is best understood in terms of encapsulated but interfacing modules, whereby each module is (a) specialised for a very narrow set of computations that are opaque to every other module, and (b) accomplishes the same by using proprietary information stated in representational terms that is illegible to all but the concerned module. Given that the foundational definition of phonology, as contrasted with phonetics, is that phonology deals with abstract and categorical information the same module cannot also be burdened with computing over gradient phonetic information. Second, as argued by Hale and Reiss, 2008 in their book-length treatise, if something is explicable purely by appeals to phonetic motivations then further invoking phonological grammar for the same purpose leads to unnecessary duplication of information. According to this view, further, a defining hallmark of pure phonological computation is a process' ability to apply regardless of whether phonetic motivations are present or not. As argued by Chabot, 2021; Chabot, 2022a and Mielke, 2008, for instance, although a large number of phonological operations do also exhibit phonetic motivations the very existence of processes that defy the same (e.g. crazy rules) suggest such overlap is incidental and non-explanatory. Consider, for instance, the fact that despite meeting identical phonetic criterion the Malayalam plural suffix -kal] never palatalizes. In light of these facts we hypothesized that one of the two following scenarios must be true of the Malayalam process, at least as it applies across morpheme boundaries.

1. Scenario 1: The process is purely co-articulatory. If this is truly the case then one would expect the process to apply to all nonce stem+suffix combinations that meet the required co-articulatory criteria
2. Scenario 2: The process involves phonological computations that apply only in the presence of suffixes that are licensed for palatalization. If this is the case then one would expect the plural suffix to not undergo palatalization with nonces. Not unlike with actual lexical items.

Finally, note that the use of nonce-stems further helps us avoid the problem of a subset of Malayalam lexical items being marked as "exceptions". Since nonce stems do not contain any entries in the lexicon the issue of "lexical exceptions" do not apply to them.

5.3 Experiment I: Two forced-choice tasks

The history of the perceptual forced-choice (FC) task can be traced back to the early development of experimental psychology and psycholinguistics. Wilhelm Wundt (Wundt, 1893), a pioneer in experimental psychology, laid the groundwork for the method in the late 19th century. However, its application to language research gained prominence only in the mid-20th century with the advent of more sophisticated experimental designs and statistical analyses. The perceptual forced-choice task typically involves presenting participants with stimuli and asking them to make judgments or selections based on predefined criteria, and limited response-choices – hence, *forced-choice*. For example, in a linguistic context, participants may be presented with ambiguous sentences and asked to choose the most plausible interpretation, with the experimental design limiting the number of choices available. The forced-choice nature of the task minimizes the possibility of random responses and allows researchers to gain insights into underlying cognitive processes (Marslen-Wilson, 1975). A typical early application of the FC methodology involved experimentation on ambiguity resolution. By presenting sentences with multiple possible interpretations, but a priori limiting the options for response to a subset of theoretically motivated options, researchers were able to examine the factors influencing participants' choices, the speed (reaction-time) at which disambiguation occurs, asymmetries in ambiguity resolution given specific syntactic constructions etc. Likewise, the fundamental motivation behind employing forced-choice tests in speech perception lies in the need to systematically investigate how individuals process and discriminate among the myriad acoustic features present in running speech. By constraining participants to make explicit choices among predefined alternatives, researchers can discern the fine-grained perceptual distinctions that underlie speech comprehension. As with ambiguity resolution in sentence processing, this method offers a controlled environment for probing the mechanisms that govern auditory processing, shedding light on the cognitive intricacies involved in decoding speech. Indeed, the utility of FC as an experimental toolkit has been extended to computational linguistics, for instance, by McClelland and Elman, 1986 in their TRACE model. In their study, they used forced-choice tests to investigate how the model performed in discriminating between similar-sounding words. Participants were presented with auditory stimuli, and they had to make forced choices among competing lexical candidates. The study was instrumental in providing insights into how context and feedback mechanisms influence word recognition and how the model simulated human-like performance in resolving lexical ambiguity. Of more immediate interest to us, however, is the study by Norris, McQueen, and Cutler, 2003 on listeners' ability to adapt to artificial accents. The primary goal of the study was to examine how perceptual learning occurs in the domain of speech perception, particularly when listeners are exposed to an artificial accent. The authors made use of speech stimuli that included words pronounced with an artificial accent, thereby introducing phonetic variations that were not typical in the participants' native language. Participants then underwent a training phase where they were exposed to the speech stimuli with the artificial accent. This training aimed to simulate a real-world scenario where individuals encounter novel accents and need to adapt for effective communication. Forced-choice tests were then employed to assess participants' ability to discriminate between words with and without the artificial accent, with the participants having to make choices among alternative responses, indicating their perception of the speech sounds. Dutch listeners first made lexical decisions on Dutch words and nonwords. The final fricative of 20 critical words had been replaced by an ambiguous sound, between [f] and [s]. One group of listeners heard ambiguous [f]-final words (e.g., [wItlo?], from *witlof*, 'chicory') and unambiguous [s]-final words (e.g., *naaldbos*, 'pine forest'). Another group heard the reverse (e.g., ambiguous [na:ldbo?], unambiguous *witlof*). Listeners who had heard [?] in [f]-final words were subsequently more likely to categorize ambiguous sounds on an [f]–[s] continuum as [f] than those who heard [?] in [s]-final words. Control conditions ruled out alternative explanations based on selective adaptation and contrast. The authors, thus, argue that lexical information can be used to train categorization of speech. This use of lexical information differs, they point out, from the on-line lexical feedback embodied in interactive models of speech perception. In contrast to online feedback, lexical feedback for learning is of benefit to spoken word recognition (e.g., in adapting to a newly encountered dialect).

A specific version of FC, called *two-alternative forced choice* (2AFC) is often employed to test participants' preference between two closely related versions of stimuli. Originally developed as a psychophysics toolkit by Gustav Theodor Fechner (Fechner, 1860) to measure sensory thresholds. In this method, a participant is presented with a stimuli-pair and must choose between two alternative versions, typically indicating whether they consistently prefer one variety over the other. A closely related use for 2AFC involves testing of participants' discrimination threshold wherein the stimulus intensity is varied across trials, allowing researchers to determine the threshold at which the participant can reliably detect or discriminate between the paired versions. Fechner's work, particularly in the context of his psychophysical laws, was instrumental in laying the foundation for the quantitative study of the relationship between physical stimuli and perceptual experience. In our study we employed a closely-related version of 2AFC, called *two-interval forced choice* (or, 2IFC). In 2IFC, participants are presented with two successive intervals, each containing a stimulus. One of the intervals contains the target stimulus, while the other interval serves as a reference or control, with successive trials randomizing the relative positions of the target and control stimuli in each pair. The participant's task is to either identify which interval (the first or the second) contains the target stimulus, or to indicate a preference for either the first or the second member of a pair.

5.3.1 Participants

The study was conducted at the MARCS Institute of Brain, Behavior and Development, at the Western Sydney University with native Malayalam and early/simultaneous bilingual Australian English speakers from the greater Sydney region of Australia. We recruited participants in three categories, namely SP(eaker), C(onsultant), and SUB(jects). Only one male participant each were recruited under the SP and C categories. The participant in the SP category was used to record all audio stimuli used in the experiments. The participant in the C category provided consultation in the capacity of naive native listener by listening to and verifying the native-ness of, both, the test stimuli recorded from the participant in the SP category as well as the during-task recordings made from the actual test subjects in the SUB category while they performed articulatory/production tasks.

A total of thirty right-handed early/simultaneous Malayalam - Australian English bilinguals were recruited in the SUB group. The participants were recruited through targeted advertisements placed around the Milperra campus of Western Sydney University, and through the Sydney Malayali Association. All participants were adults (over 18 years of age) and literate in English. A majority of the participants were also literate in Malayalam, but some having been born and raised in Sydney were not. However, this was not considered an issue since (a) all of our participants were screened for fluency in speaking and hearing through targeted speaking and listening tests, and (b) our tasks did not require the participants to be literate in Malayalam.

The participants were tested at the auditory laboratory of the MARCS Institute, located (at the time) in Milperra in Sydney. Subjects were reimbursed for their travel costs to the institute, and received compensation in the amount of AUD 50.00 per hour for participation in the test. The consultant participant in the C category and the speaker in the SP category were further reimbursed at the same rate for the additional hours of work towards recording stimuli and helping conduct the language proficiency tests with the participants in the SUB category.

5.3.2 Stimuli

As mentioned before the process of velar palatalization in Malayalam tends to exhibit many dialectal and idiolectal variations (Mohanan and Mohanan, 1984). In order to accomodate the possibility suggested by Mohanan and Mohanan that velar palatalization is indeed a phonological process that simply works around numerous "lexical exceptions" we opted to test native speakers' behavioral responses to nonce words. The stimuli utilized consisted of nonce-stems of the CV_1CV_2 pattern, where the initial V_1 was randomly chosen from the set [a,o,u] while the stem-final V_2 was either [i], [e] or [a]. There were fifteen samples of each possible stem-final vowel, yielding a total of forty-five nonce-stems.

We conducted this study in two phases, separated by around two months. In the first phase of the study, experiment 1, we utilized the forty-five nonce stems alongside two Malayalam suffixes that are reported to routinely trigger palatalization in suffixation – the suffix -kkə] is used to mark the dative form (Mohanan and Mohanan, 1984) while the suffix -kkuka] is used for the citation form of Malayalam verbs (Jiang, 2010). Fifteen repetitions of each stem final vowel was combined with each of the suffixes two yield nonce dative and citation forms respectively. The desired form/suffix was induced during experimentation by the use of appropriate carrier sentences. Table 6.3 below provides a sample of the stimuli utilized in this first phase of the experiment.

Nonce stem with final /i/	Nonce stem with final /e/	Nonce stem with final /a/	Suffix -kkuka]	Suffix -kkə]
pati	pate	pata	patikkuka patekkuka patakkuka	patikkə patekkə patakkə
suti	sute	suta	sutikkuka sutekkuka sutakkuka	sutikkə sutekkə sutakkə
...	...	...	...	...
lopi	lope	lopa	lopikkuka lopekkuka lopakkuka	lopikkə lopekkə lopakkə

TABLE 5.4: Examples of experimental stimuli used in exp.1A

Given that both of the suffixes mentioned above are reported to routinely trigger palatalization, a fact also observed in our experiment, we decided to include a third plural suffix (-kal]) in a second experiment (2). The sole purpose of this suffix was to provide a control case in order to tease apart grammatical constraints on palatalization from purely co-articulatory or phonetic ones. We hypothesized that a purely phonetic co-articulatory effect, especially in the context of nonce-stems, should induce palatalization with the plural suffix as well given the identical phonetic environments across the morpheme boundaries. The table below illustrates the stimuli paradigm utilized in the second part of the study.

Nonce stem with final /i/	Nonce stem with final /e/	Nonce stem with final /a/	Suffix -kkuka]	Suffix -kkə]	Suffix -kal]
pati	pate	pata	patikkuka patekkuka patakkuka	patikkə patekkə patakkə	patikal patekal patakal
suti	sute	suta	sutikkuka sutekkuka sutakkuka	sutikkə sutekkə sutakkə	sutikal sutekal sutakal
...	...	...	...	...	...
lopi	lope	lopa	lopikkuka lopekkuka lopakkuka	lopikkə lopekkə lopakkə	lopikal lopekal lopakal

TABLE 5.5: Examples of experimental stimuli used in exp.1B

We recorded spoken samples of each of the nonce stems, as well as the actual suffixes, used in our experiment with the help of the male native speaker recruited under the SP category. Further, we also recruited the suffixed forms derived by combining each of the forty-five nonce stems with the three suffixes chosen for the study. Each of these suffixed forms were recorded by the phonetically trained SP subject in both standard (palatalized for the dative and citation forms, unpalatalized for the plural forms) and deviant forms (unpalatalized for the dative and citation forms, and palatalized for the plural forms). This yielded palatalized and unpalatalized forms for all stem+suffix forms. These recordings were then subjected to both auditory inspection by both the SP and C subject, as well as laboratory phonetic verification by the researchers. Any token found to be ambiguous or unclear with regard to its acoustic phonetic qualities were eliminated, and subsequently re-recorded until all tokens passed our standards tests. All recordings were made using Praat software on a Lenovo Thinkpad P14 laptop running Ubuntu LTS, a Sennheiser MKH-50 microphone, inside MARCS' anechoic chamber.

5.3.3 Procedure

The subjects underwent testing at the Milperra campus of the MARCS Institute of Brain, Behavior and Development, Western Sydney University, in Sydney, Australia. All subjects were required to sign a consent form, and given basic instructions on the procedures involved in the study. Each participant attended a single one and a half hour session. During the first half hour they went through a training phase to familiarize them with the forced-choice task. They listened to pairs of English words in which one member of each pair had a deviant pronunciation – e.g. for the word dog the two members of the pair were [dɔg] and [d^{h}ɔg]. The subjects responded with a keypress to indicate whether they preferred the first (1) or second (0) member of the pair. The positioning of the deviant form was randomized across trials.

Experiment 1 In the testing phase of experiment 1A the subjects were seated in a testing booth, in front of a Thinkpad P14 laptop running on Ubuntu LTS. The auditory stimuli were presented in pairs, like in the training phase, over a pair of open-back Sennheiser HD-650 headphones. They listened to 45 (nonce-stems) x 2 (suffixes: -kkuka] and -kkə]) pairs of stimuli, presented randomly using the psychopy stimuli presentation software. Each stimuli pair consisted of a pair of suffixed nonce stems, with the positioning (first or second member of the pair) of the standard pronunciation randomized across trials. Each member was approximately 1.5 seconds in duration, and were separated by an interval of 1 second. 500 ms after the second member of the pair was presented by software the participants heard a beep indicating that they were required to provide a response. The participants indicated their preference for either the first (1) or second (0) member of the pair by pressing the pre-determined numeric keys on a keyboard.

Experiment 2 This second phase was conducted under identical conditions at the MARCS Institution, about five weeks after the first phase of the trials. The same participants returned to take part in this phase of the experiment. In this experiment we used the original nonce-stems, the two suffixes that figured in our first experiment, plus an additional plural sufffix (-kal]) that Mohanan and Mohanan, 1984 reported to be non-palatalizing. We recorded similar standard and deviant pronunciation suffixed forms with the third suffix using the same participant from the SP category as the original experiment. In order to ensure comparability the suffixed forms with the original two suffixes were not recorded again, and we opted to reuse the material from the first experiment. The method of stimnuli presentation and response collection was identical to the first phase of the experiment.

Note: After presentation of each pair of standard-deviant stimuli pair the software waited until the participants' responses before proceeding with the experiment. While we did not collect reaction-time measures for this set of experiments, the response epoch was not time-limited in order to allow the participants enough time to mull over their choices appropriately. However, the design did not allow them to replay the stimuli pairs once it was presented the first time. A total of fifteen tokens per stem-final vowel type per suffix was presented in each phase of the experiment.

5.3.4 Results

As a first pass we had our phonetically trained male native listener in the C category listen to all recorded samples, both standard and deviant, and grade them as S(tandard) or D(eviant). All 45x3 =135 recorded samples recorded were also subjected to phonetic analyses by the researchers. 127 of these 135 samples yielded phonetic measures typical of velar stops in palatalized and non-palatalized articulations. As is typical vowels in palatalized utterances yielded lower F1 values and higher F2 values compared to non-palatalized samples. Typical F1 values of the most fronted vowel in our tests, [i], tended to be approximately 220hz to 270Hz, while the F2 values tended to be around 2000Hz. Conversely, F1 values for non-palatalized vowels were found to be higher than palatalized velars, with the highest values (approximately 350 Hz to 400 Hz) coming from [i] and slightly lower (but still higher than palatalized items) from [e] (approximately an average of 350Hz) and [a] (around 300Hz to 320hz). F2 and F3 values were generally lower than palatalized velars, with a preceding [e] yielding an average of around 1600Hz and [a] around 1250 to 1300Hz.

Auditory judgements of the native listener largely reflected judgements that adhered to these phonetic cues. 129 out the total 135 recorded utterances were categorized correctly, with only 6 samples yielding mismatches between the intended and reported judgments. Out of these 6 miscategorized samples, 5 were palatalized samples with a preceding [a] that were miscategorized by the subject C as non-palatalized, and the last one was a non-palatalized form with a preceding [i] being reported as palatalized. Unsurprisingly, the intended palatalized samples that were reported to be non-palatalized were all in the context of [a], the least fronted of all three triggering vowels. These were also observed to have F1 values that were relatively higher than those in similar contexts that were categorized properly as palatalized. Likewise, the non-palatalized form with a preceding [i] that was miscategorized as palatalized yielded lower F1 value than other similar samples. These forms were re-recorded, and consequently subject C returned, on a separate day, and performed two perceotual categorization runs. The first run with the newer recordings yielded only one error, with non-palatalized form after [i] being miscategorized as palatalized. However, this error could not be replicated on a second run of the perceptual categorization with the same speaker, performed approximately two hours after the first run. All phonetic measurements were found to be as expected, and accordingly we considered these samples to qualify for the actual testing phase.

Experiment 1

In the first phase of the experiment 15 adult native early/simultaneous bilingual listeners listened all 45 nonce-stems attached to the two palatalizing suffixes (-kkuka] and -kk]). Responses to each pairs of stimuli were recorded using the psychopy software. We conducted a 3x2 mixed design analyses of variance, with the three stem-final vowels as independent between-subjects variables and the two suffixes as within subject dependent variables. Here we found a significant effect of vowel-identity, $F(1, 14) = 20.01$, $p < .01$, but no effect of suffix-identity, nor any suffix x vowel interactions. Further, planned contrasts revealed that the vowels [i, e] induced significantly more palatalization than [a] ($p < .05$).

These results were expected on both counts as we predicted that any phonological rules (competence) notwithstanding, the gradient frontedness of the triggering vowels, especially the phonetically low-back vowel [a], will manifest observable effects in performance given the psycholinguistic and phonetic evidence discussed in the preceding sections. Further, the lack of any suffix effects is also unsurprising as the literature on Malayalam phonology is quite unambiguous about both of these suffixes being subject to palatalization productively. The figure below provides a graphical overview of the patterns of perception reported by the subjects.

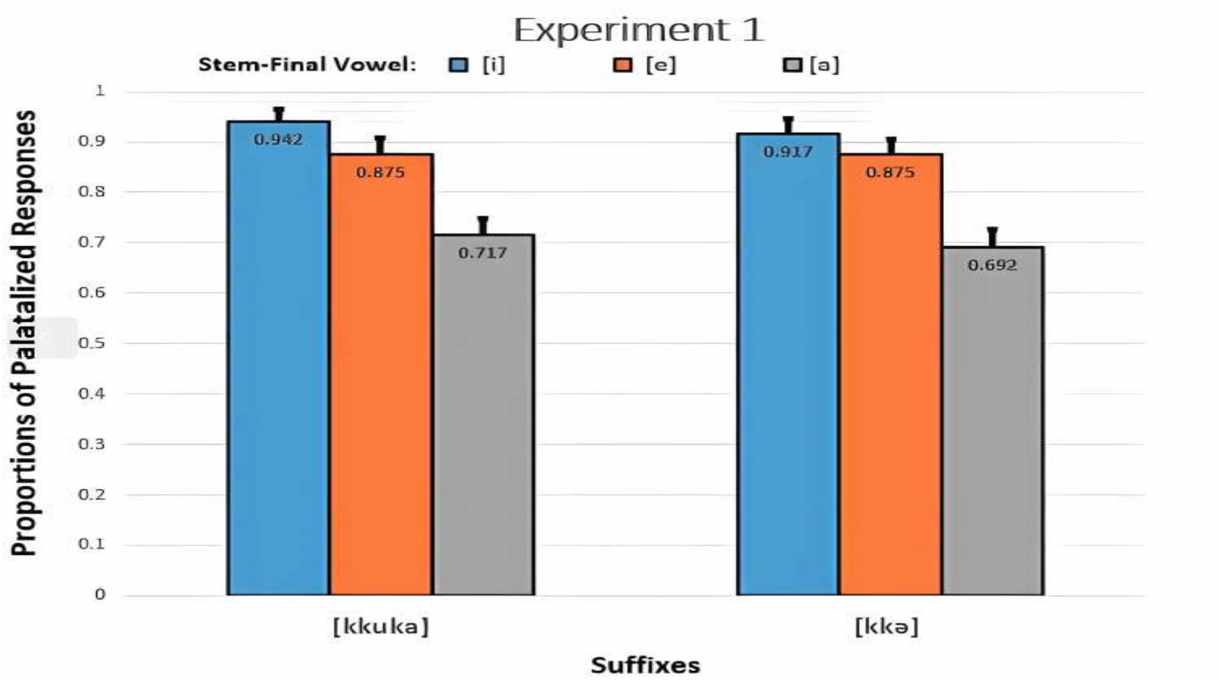


FIGURE 5.5: Experiment 1: Proportion of Palatalized Responses

These results, however, do not speak to the core concern of our study which involves testing possible phonologization of the Malayalam velar-palatalization process. On the one hand the gradient vowel effects are exactly as one would predict given the nature of phonetic co-articulation, while on the other hand the inclusion of two suffixes, both of which are reported to be palatalizing in the literature, left us without a proper control case. We rectified this in a second run of the study, the results of which are reported below.

Experiment 2

The second phase of this experiment was run about five weeks after experiment 1A. The procedure, location, and participants were identical to the first phase of the study (1A). For stimuli we used all of the nonce-stem+suffix combinations used in the first-phase of the study, but included an additional pluralizing suffix (-kal]) as a control case (see table 4 above). This means that this run of the experiment had a total of 45(nonce-stems)x3(suffixes), compared to the 45x2 for the first run. Stimuli presentation and response collection was done with the same version of Psychopy running on the same Thinkpad P14 laptop with Ubuntu LTS.

Here, once again, we found a significant main effect of vowel-identity ($F(1, 14) = 24.13$, $p < .01$), but unlike the first run of the study we also noticed significant effects of both suffixes ($F(1, 14) = 917.72$, $p < .01$) and vowel*suffix interaction ($F(1, 14) = 7.09$, $p = .01$). Planned contrasts again found [i, e] more effective than [a], p < .01, but no significant difference between [i, e]. Crucially, notice here that nearly identical results are obtained in two separate runs with the first two suffixes. Given this the lack of a suffix effect in the first phase, which is nevertheless observed in phase two, can be reliably attributed to the third suffix which (a) is claimed to be non-palatalizing in the theoretical literature (Mohanan, 1982; Mohanan and Mohanan, 1984; Jiang, 2010) and subsequently (b) fails to trigger palatalization to any noticeable level in our experimental task. The figure below provides a graphical overview of the rates of palatalized responses.

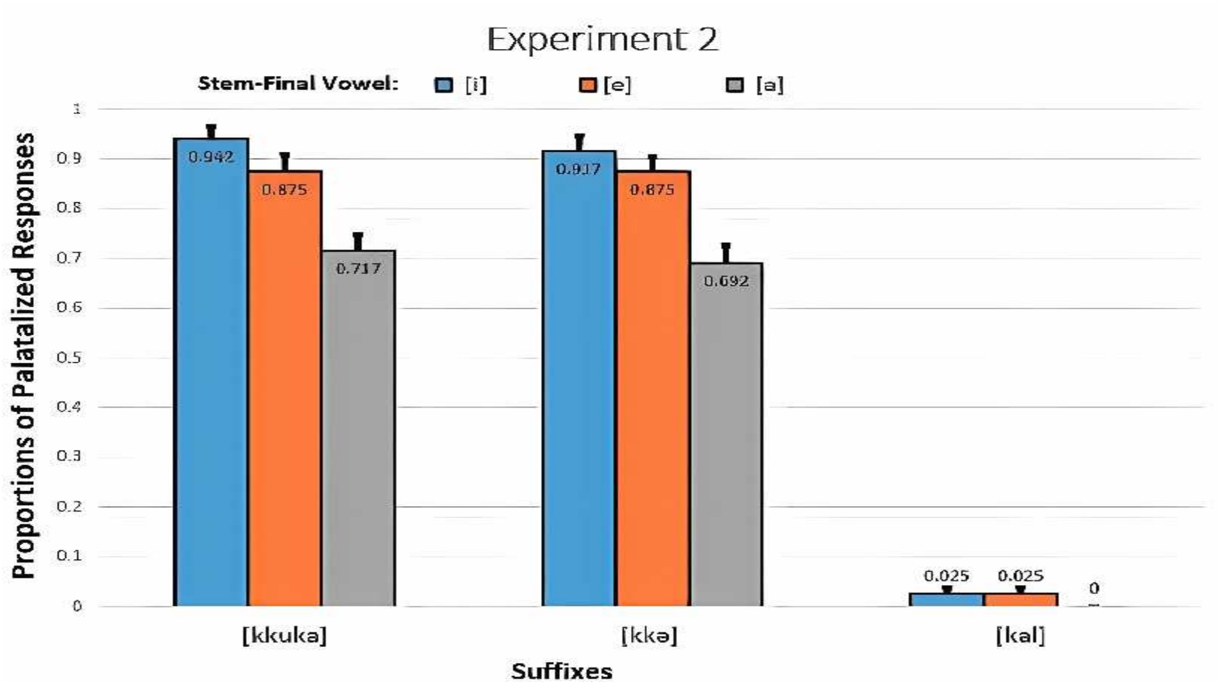


FIGURE 5.6: Experiment 2: Proportion of Palatalized Responses

A second interesting facet to notice here, particularly in light of the implicational laws of palatalization (Wilson, 2006), is asymmetry observed in the effect of vowel-identity (frontedness) on the observed rates of palatalized responses. Recall here that instrumental phonetic and laboratory phonology studies have repeatedly furnished evidence that a more-fronted vowel is more likely to result in a contextually palatalized velar stop (Keating and Lahiri, 1993), and that languages which palatalize velars in the context of a relatively less fronted vowel (e.g. [e]) are predicted to obligatorily palatalize in the context of a more-fronted vowel (e.g.[i]). The figures 6 and 7 both show this gradient effect with the two palatalizing suffixes, with the highest rate of palatalizing responses being observed with [i], and then gradually decreasing through [e] and [a]. However, from a statistical point of view in both of our tasks the two phonetically fronted vowels [i,e] fail to show any statistical difference in rates of triggering, while together they manifest significant differences in both tasks when compared to the typically low-back vowel [a].

Finally, notice that while neither of the three triggering vowels seem to have much of a palatalizing effect on the plural suffix -kal], the rare occassional/accidental palatalized responses observed with this suffix are again predictably with the two phonetically fronted vowels. While of no statistical significance, such accidents nonetheless appears to adhere to Keating and Lahiri's observation that the more fronted a vowel is the more likely it is to induce fronting in the articulation of contextual velars. Whether such contextual phonetic fronting has any significance for the phonological phenomena of velar palatalization is, of course, a function of the theoretical framework within which a discussion is couched. We will return to this issue, and others of equal theoretical import, in section 6.

5.3.5 Experiment II: Two modified WUG-tests

In phonological theory, at least since the decomposition of the phoneme into sub-phonemic primes of representation(Chomsky and Halle, 1968), distinctive features have been appealed to as the abstract encoding of those specific aspects of speech sounds that are linguistically relevant (Courtenay, 1870[1972]; Sapir, 1925; Jakobson, 1973). For Saussure, 1916[1967], p. 166, for instance, the phoneme itself was a manifestation of this linguistically relevant information in a system of contrasts – *"dans la langue il n'y a que des différences."*[9] Chabot (Chabot, 2021; Chabot, 2022b) points out, rightly, that while symbols as contrastive markers enjoy a quite different (sub-phonemic) status in modern theories when compared to this original conception, most of them inherit the assumption that phonological primes are abstract – a property of human minds and the uniquely individual experiences encountered during the critical period. This, inevitably, leads us to the issue of what features/primes abstractions of? In other words, if we admit that the primary purpose served by features/primes in phonological theory, as algebraic symbolic markers, is "to abstract from the welter of descriptive complexity certain general principles governing computation that would allow the rules of a particular language to be given in very simple forms" (Chomsky, 2000, p. 122) then the logical question that follows, at least in the

---

[9] In language, there are only differences.

sound domain (Reiss, 2007), concerns whether this abstraction pertains to acoustic or articulatory substance? In the experimental psycho- and neurolinguistic literature the basis of primes, accordingly, has been debated from multipled perspectives – some claim primes are primarily acoustic (Stevens, 1989; Liberman and Mattingly, 1985) in nature, yet others argue for an articulatory origin (Browman and Goldstein, 1989; Browman and Goldstein, 1992; Best, 1995b), and more recently the emerging view argued for Poeppel, Idsardi and colleagues have conceptualized them as anchor-points for co-ordinate transformation between articulatory and acoustic spaces (Poeppel, Idsardi, and Van Wassenhove, 2008; Poeppel and Idsardi, 2011). While this latter view explicitly argues for primes as links between articulation and perception, that primes figure in a non-trivial way in both directions has been the staple of the literature since the advent of phonological theory (Chomsky and Halle, 1968; Goldsmith, 1976b; Cutler, 1992; Best et al., 1994; Best, 1995b; Halle, Vaux, and Wolfe, 2000; Halle, 2002; Halle, 2005). In this view features play an important role in both perceptual (Mandal et al., 2020) and articulatory processes, and the way we reason about such representational primitives is in the light of phonological alternations (Scheer and Mathy, 2021) which figure in production as much as they do in perception of speech signals. In order to address the articulatory component we complement our perceptual forced-choice study with a psycholinguistic investigtion of velar palatalization in production of Malayalam suffixed forms.

For probing productive palatalization in externalization we opted to resort to a slightly modified WUG test. The original WUG test was developed in a seminal study investigating children's acquisition of morphological processes in English (Berko, 1958). A psycholinguistic method designed to assess children's ability to generatively apply grammatical rules to words they have never heard before, the methodology is primarily intended to probe the possibility that children are innately capable of formulating algebraic rules that they then instinctively apply to any form that meets the structural description of said rule – in this case a singular noun that needs to be turned into a plural, a process that in English typically (though not always [10] requires concatenating the plural marker -z] at the end of the singular form. The name "WUG" itself is derived from one of the test items, where children are shown a picture of a single bird, which they were told is called a "WUG". This is not an English word at all, and thus could not possibly be part of the lexicon or have any associated learned plural form. Berko, thus, argued that if children instinctively produce the plural form "WUGS", such pluralization would imply the application of a productive grammatical rule that children are equipped with. In her original study this position is reinforced as the subjects, when shown two of these figures, and responded with the predictable "WUGS". The test provides insights into the development of morphological rules and demonstrates that even young children possess a remarkable ability to grasp grammatical structures. The WUG test has been widely used in the field of psycholinguistics to explore language development in children and gain a deeper understanding of the cognitive processes involved in acquiring language. The significance of the WUG test lies in its ability to unveil the implicit knowledge that children have about linguistic rules. By examining children's responses to novel words and grammatical structures, researchers can draw conclusions about the underlying mechanisms of language acquisition. The WUG test has influenced subsequent research on language development and has become a foundational tool in the study of psycholinguistics.

As originally designed the WUG test relies on visual and auditory cues – subjects are usually shown a pictorial representation of a noun (a car, a dog, or a WUG!) and they are told what that thing is called (This is a WUG!). The participants then respond with the plural form. Given the number of repetitions involved with each stem-final vowel in our task, and also the different grammatical functions of each of the suffixes used, we felt that this approach was not feasible for our purposes. Instead, we opted for a combination of auditory cues that required the subjects to listen to recordings of stems and suffixes separately, and then respond with suffixed forms.

---

[10] For instance, the plural of 'child' is an irregular form, 'children'.

Our WUG test, thus, involves perception, recognition, mental suffixation followed by overt articulation.

5.3.6 Participants

One again, the study was conducted at the MARCS Institute of Brain, Behavior and Development, at the Western Sydney University with native Malayalam and Australian English bilingual speakers from the greater Sydney region of Australia. The same group of participants who took part in the forced-choice study were also recruited for the production task. Once again, we recruited participants in three categories – C, SP and SUB. The members in each group were the same individuals who participated in the perceptual FC task. We had a priori confirmed with each participant their willingness to take part in the second study, and the contact information provided by the participants as part of their signed information-sheets and consent forms was used to call them back for this study. The speaker in the SP category was utilized in the recording of the auditory stimuli, while the participant in category C provided native listener judgments on the acceptability of the recordings. Further, this particular listener (C) also provided native listening judgements and helped categorize the utterances recorded from the SUB participants' as part of their articulated responses to the experimental task. A total of seventeen participants were recruited, one each under the C and SP categories, and fifteen under the SUB category.

5.3.7 Stimuli

The stimuli used in this second experiment consisted of the same 45 nonce-stems and the 3 suffixes that figured in our pair of FC tests. We separately recorded individual instances of all 45 nonce stems, and all 3 suffixes with the help from participant SP. The standalone suffixes were recorded with non-palatalized initial velar stops. Further, we also recorded three instances of unrelated real stems (e.g. [ad̪ʰjapika] "teacher-Fem") and (-mar] "Num-Pl"), alongside the suffixed forms (e.g. [ad̪ʰjapikamar] "teachers-FEM.Pl"). These forms, unrelated to the phenomena of palatalization, were used only in the training phase to familiarize the SUB participants with the suffixation task. No suffixed forms were recorded for the nonce-stem+suffix combinations for this task. All recordings were first judged by the native listener C, and were subsequently subjected to acoustic-phonetic examination by the researchers using the PRAAT software.

5.3.8 Procedure

Experiment 1 In the first testing phase of experiment II the subjects were seated in a testing booth, in front of a Thinkpad P14 laptop running on Ubuntu LTS. The auditory stimuli were presented in pairs, like in the training phase, over a pair of open-back Sennheiser HD-650 headphones. They listened to 45 (nonce-stems), selected randomly, followed by one of two randomly picked suffixes (-kkuka] and -kkə]). The nonce-stem was presented first, followed by a 10ms silent gap, and then the suffix – e.g. [but̪i] – 10ms – [kkuka]. Each of the nonce-stems were paired with each of the suffixes, thus resulting in 15 repetitions per stem-final vowel per suffix, for a total of 45x2 trials.

Experiment 2 The second phase of this experiment mirrors experiment, and we used all of the original nonce-stems, the two suffixes that figured in our first experiment, plus the non-palatalizing plural sufffix (-kal]) as a control caae. The method of stimnuli presentation and response collection was identical to the first phase of the experiment

Note: After presentation of each pair of stem+suffix stimuli pair, where the suffix always occured after the 10ms silent gap period, the software presented the participants with a short beep indicating that they could now provide their oral response. No time-limit was imposed on the response epoch, and the software waited until the participants' responded by pressing the spacebar before proceeding with the experiment. Once again, we did not collect reaction-time measures for this set of experiments, and the response epoch was not time-limited in order to allow the participants enough time to provide a confident response. The design did not allow them to replay the stimuli pairs once it was presented the first time. It is also worth noting that for both experiments I and II the first phases (1) were performed on consecutive days, while the second phases (2), also performed on consecutive days, were separated from the first runs by around five weeks.

5.3.9 Results

Just like the first set of experiments, all auditory stimuli items were first subject to perceptual verification by subject C. The native listener confirmed that none of the velars sounded palatalized, and subsequently phonetic verifications were performed paralleling those from the first set of experiments. Once instrumental data and native judgements were aligned the items were deemed fit for the actual testing phase.

Experiment 2A

The first phase of this study, once again, was run with only the palatalizing suffixes (-kkuka] and -kkə]). The participants' production of the suffixed forms were recorded, and subsequently put through two stages of verification. As a first pass, the native listener, participant C, listened to all of the produced forms and categorized them as either P(alatalized) or NP(non-palatalized). This categorization task was performed by subject C twice, on two separate days. Judgments were mostly found to be matching on both occasions, with just three instances of mismatch, all involving the stem-final vowel [i], two of which also involved the suffix -kkə] and one involved the suffix -kkuka]. In the first run of categorization task participant C had categorized all responses as palatalized, except the three instances noted above. However, on the second run this observation could not be replicated. Given the low number of mismatches, we next moved on to phonetic inspection of the recorded forms. Confirming the native listener's intuitions, the acoustic measures revealed a systematic difference between palatalized and unpalatalized forms. Mean F2 values of vowels were highest for [i] and lowest for [a] ([i] > [e]

$>$ [a]), while palatalized forms of each vowel demonstrated higher F2, indicating frontedness, relative to unpalatalized forms. Likewise, palatalized velars displayed higher mean centroid, intensity and duration values than non-palatalized velars.

We next conducted a 3x2 mixed design ANOVA, with the three stem-final vowels as independent variables and the two suffixes as the dependent variables, with the scores for each combination coming from the categorization task performed by subject C. In this phase the palatalization pattern revealed a main effect of vowels, $F(1,14) = 7.56$, $p < 0.01$, and a lower but still significant suffix effect, $F(1,14) = 6.52$, $p = 0.023$, but no interactions. This observation, once again, mirrors the patterns observed in experiment 1A. Planned contrasts revealed that vowels [i] and [e] were better triggers of palatalized articulation than [a] ($p < 0.01$), but no difference was observed between [i] and [e] ($p > 0.05$). The two suffixes differed slightly in their rate of palatalizing the forms ($p < 0.05$). This suffix-difference was not observed in the first-phase of the FC task. The figure below provides an overview of the response patterns.

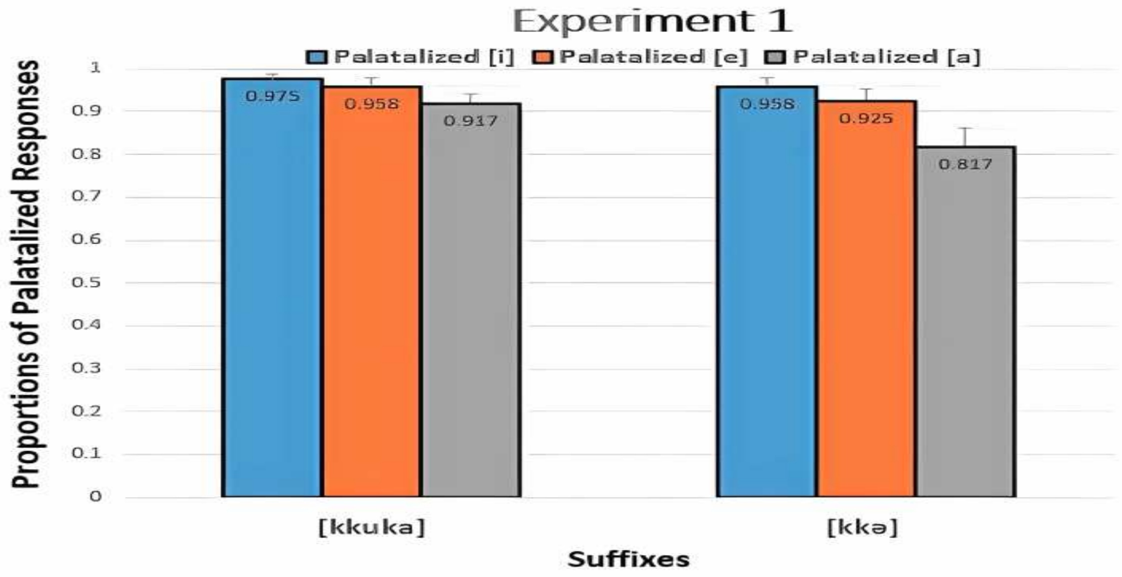


Figure 1: Experiment 1

FIGURE 5.7: Experiment 2A: Proportion of Palatalized Responses

Experiment 2B

The second phase of our modified-WUG test was conducted with the original 45 nonce stems, 15 with each of the three stem-final vowel types, the two original suffixes (-kkuka], -kkə]) and a third, non-palatalizing control case in the form of the pluralizing suffix -kal]. We conducted the same native-listener categorization tasks with the recorded participant responses, twice over. Surprisingly for this phase of the test the native listener's judgments did not produce any mismatches at all with subsequent phonetic tests. All phonetic measues obtained reflected those from the first phase, with a systematic phonetic difference between palatalized and unpalatalized forms. Mean F2 values of vowels were, once again, highest for [i] and lowest for [a] ([i] > [e] > [a]), while palatalized forms of each vowel demonstrated higher F2. Likewise, palatalized velars displayed higher mean centroid, intensity and duration values than non-palatalized velars.

Next, we ran a 3x3 mixed design analyses of variance in order to probe the statistical effect of the triggering vowels and the suffixes on the participants' production of palatalized vs. non-palatalized forms. Here, we observed main effects of vowels, $F(1,14) = 10.38$, $p < 0.01$, and suffixes, $F(1,14) = 1747.79$, $p < 0.01$, but no interactions at all. Planned contrasts on the vowels conformed to the findings reported in the previous production experiment (2A), with the exception that all vowels failed to exhibit triggering with the additional suffix [kal]. Further, for suffixes, we found the first two suffixes (-kkuka] and -kkə]) to have a significantly higher palatalizing effect than the unlicensed [kal], $p<0.01$. Unlike the first phase of this production task, however, the two licensed suffixes did not exhibit any significant differences between them.

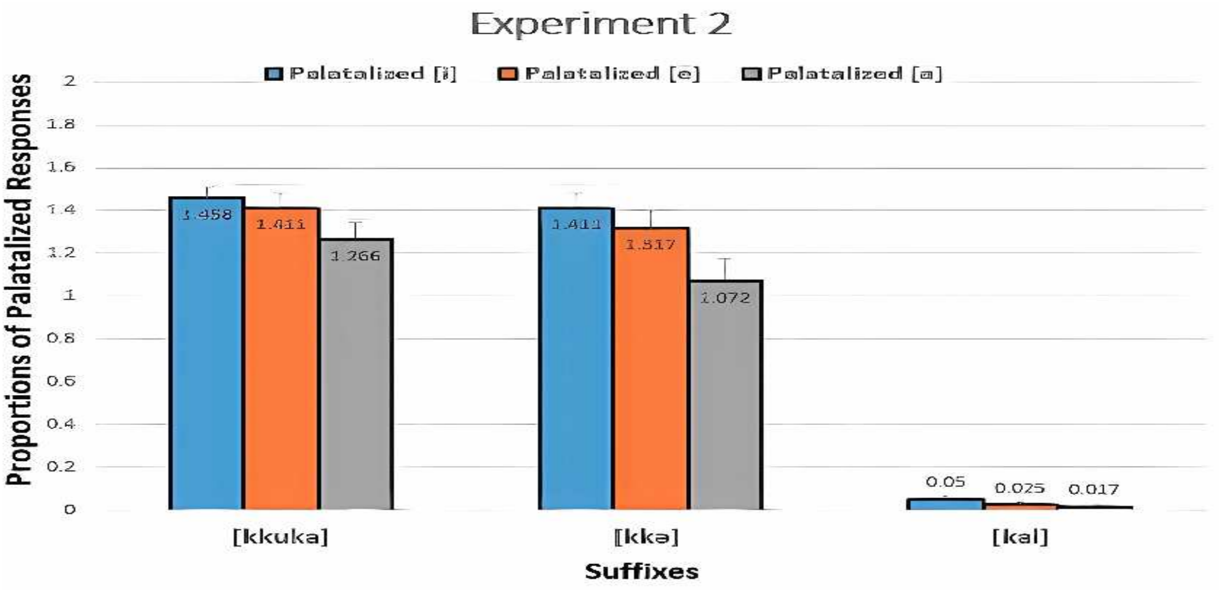


Figure 2: Experiment 2

FIGURE 5.8: Experiment 2A: Proportion of Palatalized Responses

The following section delves into a detailed discussion of the implications of these results, particularly for the various competing theoretical accounts of velar palatalization in phonological theory. But note here that both pairs of tests reported here exhibit certain idiosyncrasies. First, note that there is clearly an interaction going on between a productive grammatical rule that demands palatalization in certain contexts, phonetic naturalness notwithstanding. This is the case, for instance, for velars that surface with palatalization in the context of the typologically atypical low-back vowel [a]. Yet, this rule seems subject to performative restrains arising from phonetic pressures. Consider, for instance, the 28% lapse in proportions of palatalized responses in the FC task. Conversely, though the pluralizing suffix -kal] in our tests appear in phonetic environments that are identical to the other two (single-melody velar stop preceded by a stem-final [i], [e] or [a]) it rarely triggers any palatalization in either direction. Likewise, note also that while gradient phonetic frontedness of the vowels clearly have an effect on the proportions of palatalized responses in both studies, only the least-fronted vowel [a] exhibits any statistically significant difference. Finally, recall that the two palatalizing suffixes did not exhibit any significant differences in the first phase of the FC task, with a suffix-effect only emerging in the second phase of the study due to the inclusion of the third suffix that does not palatalize at all. However, while in the production based task (a) a suffix effect was observed in the first phase of the study, and consequently the two palatalizing suffixes showed statistically observable differences, (b) the inclusion of the third non-palatalizing suffix in phase two retained the suffix-effect in variance but wiped out any differences between the two palatalizing suffixes, with only the non-palatalizing suffix -kal] differing significantly from the former two.

5.4 Discussion

The results of the empirical studies reported above appear to be along the lines of what Wilson, 2006 calls the implicational laws of palatalization. Palatalization in the context of [a] asymmetrically implies palatalization in the context of [i], and this is clearly true in the results reported above. Participants routinely palatalize, in both perception and production, in the context of all three vowels, but do so with higher rates of success in the context of [i], and to a slightly less extent with [e]. However, unlike Wilson, 2006 we are not ready to attribute this gradient effect to the grammar itself. For us the crucial issue concerns how to interpret the results of the empirical studies in a manner that affords a proper distribution of labor between phonology and phonetics. In what follows we first attempt to establish a proper delineation between the modules, and then attempt to provide an account of the role played by each in the phenomena of velar palatalization in Malayalam as evidenced in the results reported above.

A typical trend in the phonological literature, both theoretical (Goldsmith, 1976b; Yip, 2003; Sanyal, 2012; Bale and Reiss, 2018) and experimental (Duanmu, 2000; Pierrehumbert, 2003; Wilson, 2006; Mandal et al., 2020), that has increasingly become a cause of concern for those interested in the architecture of natural language concerns the fact there appears to be very little agreement among the practitioners themselves as to what constitutes phonology (Scheer, 2010a; Scheer and Mathy, 2021; Chabot, 2021; Chabot, 2023). In other words there is no agreement on the set of phenomena that are phonological in kind, and hence should constitute the input for theory-building. Scheer and Mathy, 2021, thus, rightly point out that even simple phenomenon such as the alternation between [k] and [s] in the word pair electri[k] - electri[s]-ity (often termed *velar softening*) are treated differently by different schools of research: it might be considered phonological in kind by some (i.e. there is one single underlying representation, /k/, and phonological computation turns it into [s] in the derivational pipeline, while the same is placed outside of the purview of phonology by yet others who claim there are two distinct lexical recordings, either of whole words (electric and electricity), or of allomorphs (electri[k]- and electri[s]-) which are then selected in the appropriate morphological context. This is problematic at a very basic level because the shape of any scientific theory is inherently determined by the range of phenomena that lie exclusively within its purview, to the exclusion of everything else. Thus, for instance, a theory of gravity is meant to explain, for instance, why a feather and pebble dropped simultaneously from a height will, under ideal circumstances, hit the ground at the same time. That in real world situations a feather is prone to be carried away, while the pebble travels straight down, is due to the interacting effects of a various number of factors (wind velocity and direction, surface area etc.) all of which lie outside the purview of a theory of gravity itself. A theory of gravity would, indeed, look very different than it currently does if such phenomena were to be considered by a subset of physicists to be within its purview. Likewise, a theory of phonology that considers velar softening (among any number of other alternations) to be outside its remit will have a very different shape than one that considers the same to be phonological. At present, phonologists by and large appear to pick and choose phenomena to explain depending on personal and subjective preferences for what they want their theory to look like. This problem in phonological theory transcends broader ontological commitments (e.g. nativist vs. emergentist, or substantive vs. substance-free theories) – for instance, within the broader nativist camp Berent, 2013 (see also Gervain, Berent, and Werker,

2012) argue that children are innately sensitive to positional binding constraints restricting adjacent co-occurrence of consonants, while Reiss, 2017 argues that such phonotactic patterns are of no consequence to phonology. Of course, positional binding either *is* phonology, or it is not, but in the absence of a unified yardstick to settle such issues Popperian competition among theories becomes impossible (Scheer, 2010b; Scheer and Mathy, 2021). Much the same holds for velar palatalization.

5.4.1 Null-Hypothesis: modularity

The standard scientific methodology in both theory-construction and subsequent evaluation of empirical data in light of said theory is to assume a null-hypothesis – a default set of axiomatic assumptions in the context of which "raw" empirical data is "cooked" (Hammarberg, 1981). As discussed in the initial sections of this article, even with the structuralist linguists the primary function of the basic units of natural language phonology (in this case, the *phoneme*) was the encoding of only those aspects of speech sounds that were linguistically relevant. By linguistically relevant, of course, the structuralists meant a system of contrastive differences that set a phoneme in opposition to yet others. While modern phonological theory continues to maintain this "system of differences" approach (Hale and Reiss, 2008; Scheer, 2022; Chabot, 2021), the most significant departure from the structuralist zeitgeist comes in the form of Roman Jakobson's[11] decomposition of the basic primes of phonology from the phonemic segment to the sub-phonemic features. Writing to the First International Congress of Linguistics held in April 1928 at The Hague Jakobson, 1971, 3ff. argued:

---

[11] While Jakobson is usually credited with this sub-phonemic view, Halle, 2005 points out that such decompositional views actually well predate Jakobson. For instance, in his now forgotten work on *Visible Speech* Bell, 1867, p. 38 argues that segments are like chemical compounds that can, and should, be decomposed into their constituent elements:

> The true element of articulation, I think, is a constriction or position of the vocal organs rather than a sound. Combinations of positions yield new sounds, just as combinations of chemical elements yield new substances. Water is a substance of very different character from either of the gases of which it is formed; and the vowel oo is a sound of very different character from that of any of its elementary positions.

> Every scientific description of the phonology of a language must above all include a characterization of its phonologic system; i.e., a characterization of the repertory—specific to this language – of the distinctive differences [features] among its acoustico-motor images [segments] [...] Comparative phonology must formulate the general laws which govern the relations among the correlations [features] within the framework of a given phonological system.

This position is taken up and further developed in the foundational years of the cognitive revolution by Chomsky and Halle, 1968 who retained Jakobson's original assumption that all features were binary, e.g. ±F, but proceeded to replace several of the acoustic features (e.g. grave and acute, compact and diffuse, flat and plain) with ones that were primarily of an articulatory nature (±Back, ±Highm ±Low). Halle, 2005, p. 30, thus, notes:

> The overall effect of these changes has been to emphasize the primary role of articulation in phonology. I now believe that for phonology, articulatory considerations are paramount and that acoustic aspects of speech play at best a subsidiary functional role.

Implicit in such assumptions is the fact that features/primes, while they are algebraic symbolic markers, still contain intrinsic substantive (phonetic) content. Indeed, in the famous chapter nine of SPE Chomsky and Halle grapple with the issue of form and substance. In this much debated chapter (Hale and Reiss, 2008; Reiss, 2017; Reiss and Volenec, 2021; Chabot, 2021), the authors first bemoan the fact that the theory they have thus far developed has become overtly formal. That is, the authors consider seriously the problem that the computational system has become powerful enough that it is now capable of generating patterns that have never been attested in any language. This issue, often discussed as *overgeneration* is problematic given the core concern of generative grammar is a computational system that generates all ***and only*** the grammatical patterns to the exclusion of everything else. Chomsky and Halle, 1968, p. 400, thus, write:

> The entire discussion of phonology in this book suffers from a fundamental theoretical inadequacy. Although we do not know how to remedy it fully, we feel that the outlines of a solution can be sketched, at least in part. The problem is that our approach to features, to rules, and to evaluation has been overly formal. Suppose, for example, that we were systematically to interchange features or to replace $[\alpha\ \mathrm{F}]$ by $[\ -\alpha\ \mathrm{F}]$ (where α is $+$, and F is a feature) throughout our description of English structure. There is nothing in our account of linguistic theory to indicate that the result would be the description of a system that violates certain principles governing human languages. To the extent that this is true, we have failed to formulate the principles of linguistic theory, of universal grammar, in a satisfactory manner. In particular, we have not made any use of the fact that the features have intrinsic content. By taking this intrinsic content into account, we can, so it appears, achieve a deeper and more satisfying solution to some of the problems of lexical redundancy as well as to many other problems that we have skirted in the exposition.

The authors' solution to reigning in the generative power of SPE came in the form of *markedness*, the argument that not all valued features are of equal status. Particularly, they argue that certain values of certain features are unmarked, and thus enjoy a priviledged status. These values are more likely to emerge than unmarked ones, are often the outcome of assimilation processes etc. While markedness did in fact have significant constraining effects on the generative power of the computational system, Chomsky and Halle were fully cognizant of the fact that this merely leads to a tighter theory, but not necessarily a more explanatory one – is X less likely because it is marked, or is it marked because it is less likely? Similar issues lead them (Chomsky and Halle, 1968, p. 428) to conclude in the same chapter (emphasis added) that,

> It does not seem likely that an elaboration of the theory along the lines just reviewed will allow us to dispense with phonological rules that change features fairly freely. The second stage of the Velar Softening Rule of English and of the Second Velar

> Palatalization of Slavic strongly suggests that **the phonological component requires wide latitude in the freedom to change features, along the lines of the rules discussed in the body of this book.**

In spite of this clear admission by Chomsky an Halle, however, the notion of markedness became ubiquitous in later generative theories, appearing in various conceptualizations (Goldsmith, 1976b; Donegan and Stampe, 1979; Sagey, 1986; Yip, 2003; Blevins, 2004; Flemming, 2013) but serving essentially the same purpose everywhere – constraining phonological patterns by appeals to notions such as phonetic naturalness, typological trends, acoustic salience, articulatory ease etc. Often overlooked in these approaches, however, is the fact that phonological theory is *not* meant to be descriptive of the attested sound patterns of languages. Rather, it is a theory of what is humanly computable as a possible phonological grammar at all. Hale and Reiss, 2008 provde an illustration of this logic as set-hierarchy, illustrated below with some modifications.

Ex 5.4.1. ATTESTED $\subset$ ATTESTABLE $\subset$ COMPUTABLE $\subset$ STATEABLE

1. Attested: English, Old English, Swahili, French etc.
2. Attestable: English in 23rd Century, Nissart in the 14th Century etc.
3. Computable: /p/ → [s] /_ /r/.
4. Stateable: /V/ → [V:] in prime-numbered syllables.

They argue (see also Chabot, 2021; Chabot, 2022a; Chabot, 2023; Scheer, 2022; Berent, 2013; Samuels, 2011; Dresher, 2014; Dresher and Hall, 2022; Odden, 2022 that phonologists' primary concern is the characterization of the architecture of phonological grammar – what are the primes of phonological computations, and what are the limits on the combinatoric operations that operate over said primes. Such a theory of phonology is a theory of what could possibly count as a phonological computation to a sapien brain. Consider the figure below, adapted with modification from Reiss (2007, p. 59).

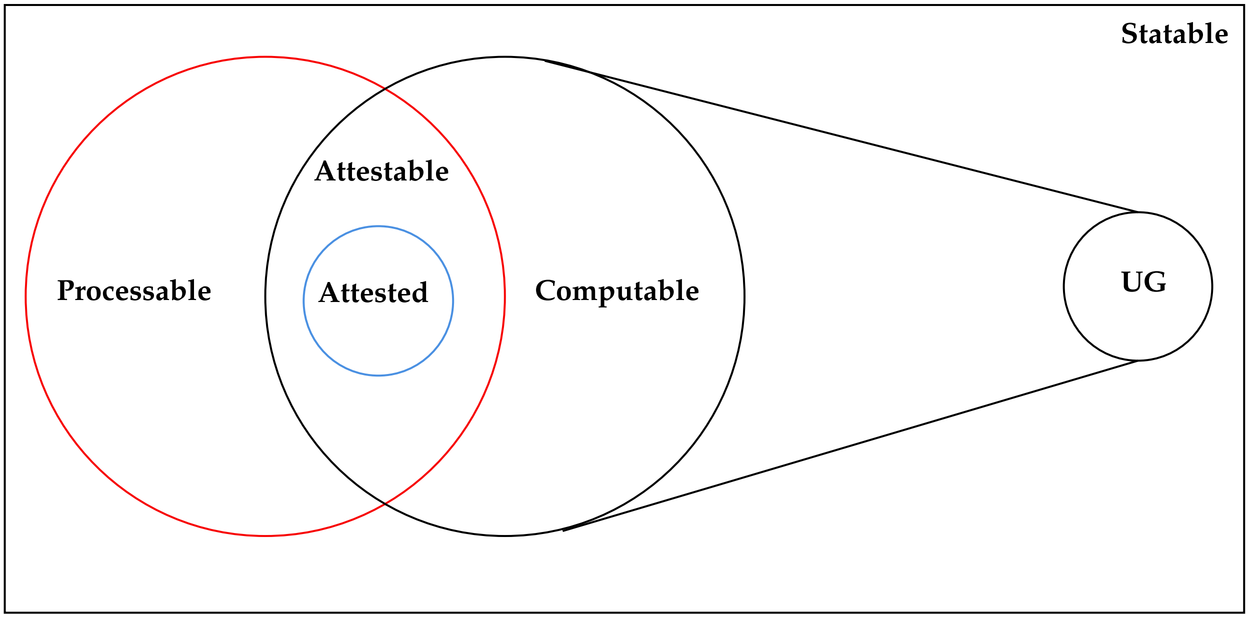


FIGURE 5.9: Biolinguistic typology

The large rectangle represents the space of grammars that could be stated within the expressive space of formal languages, for instance the *Chomsky Hierarchy*. This is the largest set containing any grammar that could, in theory, be computed by a universal Turing Machine – for instance, a grammar that lengthens every vowel in every prime-numbered syllable. Reiss (2017) points out that while such grammars are statable in formal language theory, there is no independent reasons available to assume that the faculty of language (FL) has access to notions like prime-numbers. Crucially, while such grammars are in principle computable by the human mind, they are not computable by FL. In Chomskyan terms, then, UG represents that sub-set of computable grammars that are computable *by* the FL alone. UG is represented by the small black circle on the right. Reiss argues that phonological UG contains only the set of phonological primes, and basic rules for combining them into sets, or equivalence classes. Crucially, this UG casts a wide net because it is devoid of any kind of sensory-motor or performance related substantive constraints. The scope of UG, the set of computable (by phonological UG) is represented by the big black circle. Reiss further points out that there exist other, grammar-external, constrains on the types of languages that are actually observable in reality. For instance, while UG contains no restrictions on the length of a possible grammatical sentence, memory and processing limitations automatically limit the length of observable sentences. These grammar-external biological constrains are represented by the big red circle, which contains the set of processable languages [12]. The intersection of the processable and computable grammars yields the set of attestable grammars. However, the set of attested grammars, represented by the small blue circle, is even smaller. As can be seen from the smallest blue circle being contained within this black circle, the scope of phonological UG is much larger. Reiss (2007) points out that the set of actually attested grammars are limited by a range of factors – historical accidents, genocide, the propensity of linguists to study certain languages, existence of uncontacted populations, the limitations of documented history etc. – that are neither biological nor psycholinguistic in nature. It is, thus, not sensible to burden either biolinguistic or psycholinguistic theories with explaining phenomena that do not fall within their remit.

The issue of determining what does, and does not, fall within the remit of a particular theory is a complex one. It is also one that intimately ties in with the issue of *modularity* in cognitive science (Fodor, 1983; Coltheart, 1999; Coltheart, 2011; Pylyshyn, 1980; Pinker, 2005; Chomsky, 2017). Fodor's (1983) opus is often cited as the most influential work in this domain, arguing that the mind consists of specialized cognitive modules, each dedicated to specific type of information processing. This approach strictly delineates the remit of each module, making it highly efficient in processing a specific type of information while being blind to everything else. Coltheart, 1999 and Chomsky, 2017, likewise, propose distinct modules that are each specialized to solve specific and unique problems. In functional terms what modularity implies is

---

[12] One question that often arises in this context is why would the processable languages not be a sub-set of the ones computable by UG? Reiss' (2017) argument centers around the fact that in principle one could think of grammars that are humanly processable, e.g. vowel-lengthening in prime-numbered syllables for instance, but that are not computable by phonologbical UG alone.

that a certain module deals in information encoded in a specific format (proprietary vocabulary, cf. Scheer, 2010a, 23ff.), any other type of encoding is illegible to it (informational encapsulation), and that any communication with other modules is facilitated by a transductive interface tasked with format conversion. In linguistics, however, the notion of modularity is often discussed in rather narrow ways. In a near-thousand pages long treatise Scheer, 2010a, pp. 23–24 provides an in-depth characterizatin of the deep, yet problematic and complicated, relationship that linguistics has had with the notion of modularity.

> Modularity is one of the deepest layers of generative thinking (see §623), and its presentation in the Interlude reflects its central status: in the 50s, Noam Chomsky participated in the development of the general computational paradigm (Turing - von Neumann, see §603) that underlies much modern science and grew into the standard paradigm of how the mind works (Cognitive Science). [...] Despite the fact that modularity is so deeply rooted in generative thinking, it is not a loss of time to recall all this and to apply the modular referee to phonological and interface theories. This is one thing that I had to learn while reading through the literature: modularity usually appears in introductory classes to linguistics and in first chapters of linguistic or phonological textbooks but then disappears from the radar. It is not uncommon for theories to explicitly subscribe to the modular architecture of the mind in general and of language in particular, but then to live in overt violation of modularity.

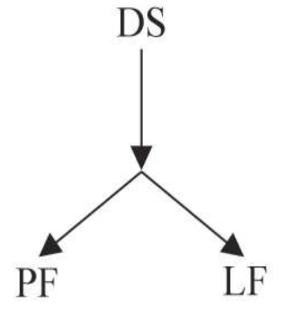
This is surprising given the fact that modularity has been a cornerstone of generative grammar since its earliest conceptions (Newmeyer, 1986/2014). In the discussions below we attempt to broaden the narrow scope in which modularity is usually discussed in the linguistic literature, and attempt to situate it in the context of broader (modularist approaches to) cognitive science (though, see the next section for some contrarian perspectives on these discussions).

> The bridge with Cognitive Science does not accommodate much traffic: linguists may not know what it takes to be a cognitive module, what its properties are, how it works, how it is defined, how it is detected and so forth.
>
> **Scheer, 2010a, p. 24**

The idea that phonology is an independent module, just like syntax, is elaborated on by Chomsky et al., 1963, p. 308 who argue that the phonological component of FL contains only those processes that are responsible for ascribing a phonetic shape to the terminal elements that appear in the output of syntactic trees "given the morphemic content and general syntactic structure of this utterance"(306ff.).

> The phonological component can be thought of as an input-output device that accepts a terminal string with a labelled bracketing and codes it as a phonetic representation.

Scheer, 2010a, p. 541 points out that this early modular conception of grammar itself has a nested, modules-within-modules structures, visualized clearly in the inverted-Y structure – DS (deep structure in even earlier terminology) ships off its output to two distinct modules (PF and LF) each of which is blind to the other one.



Inverted Y-model

This kind of inter-modular opacity is reminiscent of informational encapsulation that characterizes Fodorian (1983) modularity, but Scheer, 2010a, 514ff points out that (a) the idea that language and its sub-modules are cases of modularity qua modularity in the cognitive sciences is absent from Chomsky's works in the 1980s, and (b) consequently Chomsky, 1980 does not provide any detailed discussion of the specific properties that characterizes cogniive modularity. Scheer rightly points out this oversight is not in line with Chomsky's personal contributions to Cognitive Science, and often causes much confusion to external observers looking in on generative grammar. For instance, Scheer refers to Hirschfeld and Gelman, 1994 discussion of modularity in linguistics which betrays a sense of confusion:

> Chomsky and other maintain that these findings provide compelling evidence for the claim that the mind is modular, comprising a number of distinct (though interacting) systems (the language faculty, the visual system, a module for face recognition), each of which is characterized by its own structural principles. [...] Chomsky, however, has also suggested that the mind is modular in a somewhat different way. [...] This, in other more technical writings, Chomsky has described 'modules of grammar' (e.g., the lexicon, syntax, bounding theory, government theory, case theory, etc.) (1988:135). Here the notion of modularity appears to be tied to specific subcomponents or subsystems of the language faculty rather than to the modular uniqueness of the language faculty itself. The grammar, in the traditional sense, is located at the intersection of these distinct modules. It is not clear whether these two notions of modularity are to be distinguished, and if so how to interpret the relationship between them.

There are, obviously, as many ways to approach modularity as there are proponents. In his later work *The Mind Does Not Work That Way,* Fodor, 2000, Ch.1 expresses deep skepticism about the evolutionary psychological position (Plotkin, 1997; Pinker, 2003; Pinker, 2005) that the mind is entirely modular all the way down and that computationalism is capable of an exhaustive account of all aspects of mental phenomena. Particularly, while Plotkin and Pinker argue that every function of the mind is a biological ability (a point that, incidentally, Fodor admits as well (Fodor and Piattelli-Palmarini, 2011).) that has evolved through natural selection under specific adaptive pressures (domain specificity), Fodor maintains that central systems, such as those responsible for belief-fixation, require information from various different sources (hence of various different vocabularies) to be integrated, and hence could not possibly be modular. For Fodor, and in the position that will be adopted in the rest of this section, only peripheral input systems are modular.

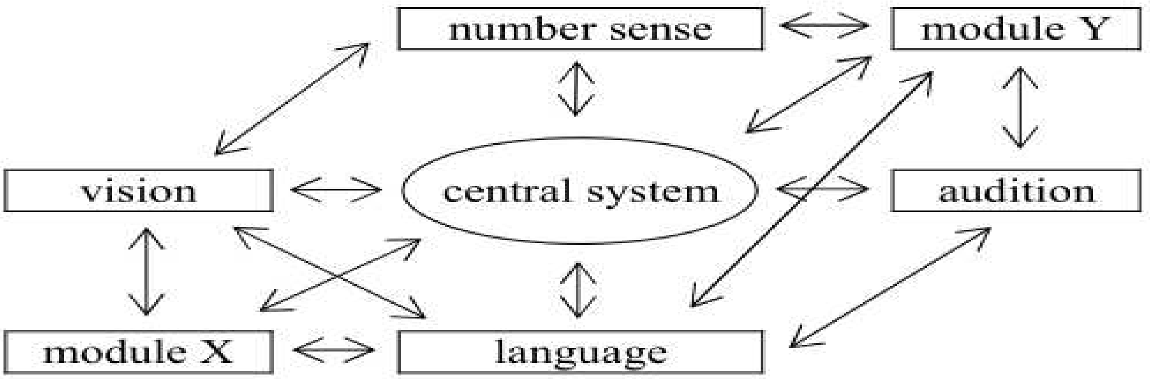


FIGURE 5.10: Modularity of Input Systems. From Scheer, 2010a.

Fodor and colleagues provide a number of criteria that are used to decide on whether something counts as a module (e.g. proprietary vocabulary, obligatory filtering, fast speed, shallow outputs etc.), though for language the most relevant qualities are domain specificity and informational encapsulation – i.e. a module solves a very specific purpose (for empirical coverage see Berent, 2013; Berent, 2021 and Scheer, 2010a) and to this end it works with a proprietary vocabulary. A similar proposal also emerges in the information theoretic works of Hammarberg, 1981, p. 261 who writes:

> Given that an *Information Processing System* (IPS) cannot process data except in terms

> of whatever representational language is inherent to it, data could not even be apprehended by an IPS without becoming representational in nature, and thus losing their status of being raw, brute, facts. The representational language of the IPS provides the categories in terms of which the IPS 'views' reality[...]

Scheer, 2010a, 526ff, accordingly, rightly points out that SPE (Chomsky and Halle, 1968, p. 60) had set similar standards, albeit indirectly, almost two decades before Hammerberg and Fodor, arguing that the phonological module can only solve problems of a very limited kind and only those expressed in a proprietary vocabulary.

> The rules of the grammar operate in a mechanical fashion; one may think of them as instructions that might be given to a mindless robot, incapable of exercising any judgment or imagination in their application. Any ambiguity or inexplicitness in the statement of rules must in principle be eliminated, since the receiver of the instructions is assumed to be incapable of using intelligence to fill in gaps or to correct errors.

In other words phonology, conceived of as a module, is a dummy computational system that works mechanically over information represented in terms of its primes, also called distinctive features. Two immediate conclusions emerge out of this assumption – (a) any information relevant to the phonology of palatalization must be stated in algebraic terms, and (b) beyond manipulating such algebraic symbols in accordance to the language-specific rules of alternation phonology is incapable of either filling in gaps or of exhibiting preferences in affecting alternations.

5.4.2 Grammar cannot be biased

The previous section discussed the relevance of modularity for phonological theory as a part of broader cognitive sciences. In particular, it was argued that modular arguments in Science are commonplace and adopting the same as null hypothesis in phonological research (and phonetic) should be commonse practice. This has been the position, for instance, in the re-emerging substance-free program in phonology (Fudge, 1967; Blaho, 2008; Hale and Reiss, 2008; Reiss, 2007; Samuels, 2011; Dresher, 2014; Chabot, 2021; Idsardi, 2022; Scheer, 2022). Recall here that it was pointed out in section 6.1.3 that the structuralists put a bulkhead between phonetics and phonology, most famously encoded in Saussure's *Langue-Parole* distinction, one that was systematically dismantled with an accompanying reduction in the conceptual distance between the two domains in post-SPE generative phonology. In canonical SPE features served three primary purposes – (a) they encode the linguistically relevant system of binary contrasts in an inventory, (b) they make explicit that shared property of segments that group them into equivalence (natural) classes, and (c) the "intrinsic phonetic content" of features serve as instructions from phonology to the sensory-motor system (Chabot, 2022a, 433ff.). This latter aspect meant that substantive properties relevant to the production and perception of sounds were directly accessible to grammar, and thus could be appealed to in constraining overgeneration, particularly ruling out those patterns that are typologically rare or (as yet) unattested. One way of achieving this end of fitting phonological theory more tightly with the set of attested languages, commonplace in both generative (Archangeli and Pulleyblank, 1994) and non-generative (Donegan and Stampe, 1979) circles, took the form of a reinterpretation of the notion of *markedness* inherited from the Prague school Chabot, 2022a, p. 434. Crucially, while for the Prague school markedness was a property of the individual grammars under consideration the neo-markedeness proponents adopted an universalist stance where one value of a binary feature is considered marked, or in some sense less optimal. In essence such approaches allow phonological computations direct access to functional substantive concerns, and essentially views phonology as trying to optimize the production and perception.

Within the generative tradition this trend of gradually reducing the conceptual distance between form (phonology) and substance (phonetics) is, perhaps, most clearly visible in Wilson, 2006's formulation of the *implicational laws of palatalization*, discussed in detail in section 2.3. Recall that per Wilson's laws (a) learners exposed to palatalization in the context of a less fronted vowel (e.g. [e]) is more prone to extend the generalization to more fronted vowels (e.g. [i]), and (b) the former observation implies that phonology is substantively-biased. The claim that the grammar itself is biased towards substantively motivated patterns is surprising, especially given Wilson, 2006, p. 974's own account that this bias is "not so strong that it excludes unfavored patterns". We argue that this is not accurate. To be sure, Wilson's (2006) empirical data appears valid given his experimental design, however the conclusion he draws from them appears not to be conceptually necessary for a theory of phonology. Wilson (2006: 968) argues that "the substantively biased model yields detailed qualitative and quantitative fits to the pattern of behavioral data [...]", even though it has been amply clear since the onset of generative phonology that the object of study is not patterns of behavioral data nor the observed trends in cross-linguistic typology. Rather, the object of study for phonologists is the system of representations and computations that generate possible phonological patterns across languages, because phonological theory is a theory of knowledge (Chomsky and Halle, 1968; Bromberger and Halle, 2000; Samuels, 2015a; Chabot, 2021; Chabot, 2023). The trends observed in typology, and the probabilistic asymmetries in speaker-listener behavior, are likewise not a direct function of just phonological competence, but are rather confounded by a myriad of third-factors (Chomsky, 2005; Samuels, 2009; Samuels, 2015a; Fitch and Cutler, 2005; Dresher and Hall, 2022; Graf, 2022; Chabot, 2021; Hale and Reiss, 2008; Reiss, 2007). For instance, by Wilson's (2006) own account the intricate perceptual and articulatory variations introduced into velars by contextual vowels determine their confusability with corresponding palatals. Likewise, perceptual confusability implies misperception and miscategorization by listeners, which diachronically precipitates as language change, and often leads to phonologization of patterns. However, neither perception nor articulation is a phonological concern, which is merely concerned with a system of language-specific network of differences or contrasts. Thus, for instance, while vowels can be phonetically more or less fronted, in phonology there is no such notion of gradient

[2.5 Back]. A vowels is either [+Back] (or, [Back] if one prefers a privative feature system), or it is not. A computational system that is capable of generating substantively unfavored patterns at all is simply capable of generating them. That substantively favored patterns are more often attested is a fact that is borne out of extra-phonological factors, and as such their explanations must be sought outside of phonology. The notion of bias itself is a gradient one, and as Hale and Reiss (2008) rightly argue, to ask gradient questions in phonology is a conceptual error akin to asking "How fast do WH elements move?".

With these ground assumptions in place, the next section elaborates on two distinct lines of discussion. First, we provide a phonological account of velar palatalization in Malayalam, and next we discuss the gradient effects observed in our tests in the light of modular approach to the phonology-phonetic interface, and on inter-modular communications. Section 6.2, then, proceeds to provide some alternatives to this highly modular framework, leaving the reader to draw their own conclusions as to which of the two competing frameworks they prefer.

5.4.3 A phonological account

While the mainstream trend in post-SPE generative phonology has been to increasingly imbue phonology with phonetic motivations (Archangeli and Pulleyblank, 1994; Prince and Smolensky, 2004), arguments from the opposite direction has also been forwarded intermittently. Similar observations are found in pre-SPE literature as well. For instance, reviewing a wide variety of alternations patterns and their phonetic motivations Fudge, 1967, p. 26 writes (emphasis added):

> The logical conclusion of this is that phonologists (above all, generative phonologists) ought to burn their phonetic boats and turn to a genuinely abstract framework. By so doing they will escape the fate of not only falling between two stools (**the result of attempting to handle systematic phonemic and systematic phonetic**

> **levels in the same terms**), but also ending up sitting in the very place which they have expended such strenuous and well-justified efforts to avoid.

Against this tide Hale and Reiss, 2000a mark a critical break and argue for a reinstating the structuralist bulkhead [13] between form and substance, between phonological computations (e.g., feature insertion or deletion) and phonetic motivations (e.g. velars are perceptually similar to palatals in the context of front vowels). Their original arguments has led to the resurgence of what has come to be known as *substance-free phonology* (SFP; Hale and Reiss, 2000a; Hale and Reiss, 2003; Hale and Reiss, 2008; Reiss, 2003; Reiss, 2007; Reiss, 2017; Samuels, 2009; Samuels, 2015a; Dresher, 2014; Odden, 2022; Dresher and Hall, 2022; Boersma, Chládková, and Benders, 2022; Blaho, 2008; Idsardi, 2022; Chabot, 2022b; Scheer, 2022, which holds that the use of functionalist principles, derived from facts about articulation and perception, in explanations for phonological patterns obscures, both, the nature of the formal system, and the precise role played by mind-external physical considerations in shaping speech signals. On the one hand the use of functional explanations in phonology, which they term substance-abuse, duplicates phonetic information in phonology and saps the formal apparatus of explanatory prowess. On the other, invoking (substantive) phonological grammars in order to account for phenomena that can be easily accounted for through channel considerations (Moreton, 2007; Moreton, 2008), for instance gradient effect of vowel frontedness in palatalization, dilutes the crucial role played by phonetics in both acquisition and diachrony.

The SFP approach is closely tied to Marr, 1977; Marr, 2010's tri-level computational framework. Briefly, Marr argues that a causal explanation of cognitive phenomena requires decomposing it into three distinct levels.

1. **Computational Level:** A formal/mathematical statement of the problem to be solved. E.g. What kind of expressive power is minimally required to capture vowel harmony?

---

[13] Indeed, Hale and Reiss describe their framework as "neo-Sausseurean".

2. **Algorithmic Level:** What are the possible algorithms that can be deployed to solve this problem. E.g. Bale and Reiss, 2018 argue that vowel harmony can be adequately explained through set-theoretic tools – set unification, set-subtraction etc. – and a linear Search-n-Change algorithm (Mailhot and Reiss, 2007; Dabbous et al., 2021).

3. **Implementational Level:** How are these algorithms implemented on neural wetware. E.g. Poeppel and Idsardi, 2011; Hickok, 2022 argue that areap Spt in the human planum temporale plays a crucial role in representation of features by performing sensory-motor integration of acoustic and articulatory information.

At the computational level, SFP adopts the minimalist position of positing the bare minimum numbers of algebraic primes necessary to express the linguistically relevant system of binary contrasts between segments. It is here that the fissure within the program begins to take shape, with Hale and Reiss (henceforth, *The Concordia School* (Hale and Reiss, 2000b; Hale and Reiss, 2003; Reiss, 2003; Reiss, 2007; Hale and Reiss, 2008) arguing, following SPE, that the set of primes are innate and universal endowments of UG, while the radical school of SFP (Dresher, 2014; Dresher and Hall, 2022; Samuels, 2009; Chabot, 2021; Odden, 2022) arguing that only the ability to create symbolic markers from experience is innate, while the features themselves emerge as a function of experience during the critical period. In other words, for the Concordia school the set of features are fixed and invariant across languages, even though not all languages may make use of all features to express contrast Reiss, To Appear. On the other hand, the radical school argues that features are meaningless symbolic markers, that the familiar feature-labels (Labial, Dorsal, Anterior etc.) are historical baggage that can and should be reduced to simple variables like α or β Chabot, 2021. The algorithmic level, likewise, continues to divide the program with Reiss, 2003 arguing for abandoning the familiar feature-geometry with feature-algebra where intra-segmental features are simply an unordered, unstructured set, as opposed to being structurally organized into a geometric tree [14], while the radical school by and large maintains the feature-geometric hierarchical notation (Dresher, 2014; Dresher and Hall, 2022; Scheer, 2022). The uniting aspect of SFP, however, is in the common assumption that phonological rules are blind to phonetic substance, and rules manipulate features as arbitrary symbols (recall, Chabot, 2021's discussion of crazy rules, as well as Chomsky and Halle, 1968's elaboration of the second stage of the velar softening process in chapter 9 of SPE). In other words, phonology being a dummy module of computation over symbolic markers simply manipulates symbols per the syntax of the concerned rule. Thus, for instance, in Malayalam the velar palatalization rule posited by Mohanan and Mohanan (1984) (repeated below) contains two structural descriptions of note – (1) it applies only to single-melody velars, and (2) it applies to all such velars that appear immediately after any vowel that bears the feature [-Back].

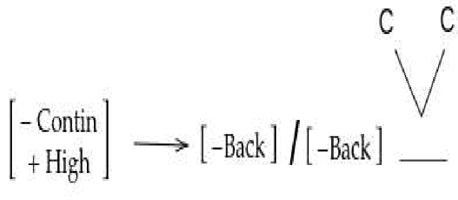


FIGURE 5.11: Malayalam velar palatalization rule. From Mohanan and Mohanan, 1984

Crucially, one consequence of being a dummy module is that such rules are incapable of introspective applications, of displaying biases towards, or against, phonetic motivations. Phonological computations, as a theory of competence, simply encodes the syntactic details of the rules that capture a native speaker's knowledge of palatalizing velars. Any gradient effects of

---

[14] Reiss's arguments center around the fact that there is no feature-geometric way to represent the non-identity condition – the requirement that two segments, C_1 and C_2, be distinct in terms of some arbitrary feature. In other words, it does not matter what feature creates the contrast, as long as some feature does. In feature-algebra, Reiss(2003:316) points out, this is expressed through the use of existential quantifiers: $\exists\ [F_i\ ] \in \mathbf{F}$ such that $[(\alpha F_i\ )_1] \neq [(\beta F_i\ )_2]$, where 1 and 2 refer to features on different segments

phonetic vowel quality are to be explained extra-phonologically, through inter-modular interface theories. We tackle the issues concerning gradient effects in our discussion of the architecture of the lower interface in the next section, while below we sketch out what we consider to be a representative formalism for Malayalam-type I-languages active in our participants.

We will assume a hybrid version of substance-free phonology that retains the core assumptions of substance-free phonology common to both the Concordia and radical schools, but diverge somewhat from both in certain assumptions about both the nature of features, the role of UG in phonology and the interface with phonetics. For instance, like the Concordia school we assume that UG endows us with sense for the sum total of possible featural contrasts, we do not make the strong claim that phonological UG comes with all possible features linked to specific substantive values from birth. Likewise, while the framework proposed here concurs with the radical school that the critical period is critical precisely because the child undergoes the process of creating language-specific networks of contrasts, it maintains that features exhibit a partially veridical relationship with their phonetic correlates instead of being completely arbitrary (cf. Chabot, 2021; Dresher, 2014). Conceptually, the framework proposed is most similar to Dresher, 2014's approach to innateness and borrows heavily from Halle, 2002; Halle, 2005's DAX (Designated Articulator) feature theory.

Like Halle (2002,2005) we make a distinction between articulator-free major class features and articulator-bound features. The latter are inextricably linked to specific articulators and from a neuro-muscular point-of-view can be thought of as controllers for specific sets of articulators – e.g. the feature [±Voice] applies to voicing and this effect is executed by relaxing or tensing of the vocal cords. The former, however, are not permanently linked to any one articulator, and depending on the specific segment into which they are bound are executed by different articulators – e.g. [±Continuant] is executed by the lips in segments such as [p,b,f] while the same is executed by the tongue body in [k,g,x]. Further, since these features are not intrinsically linked to any specific articulators it follows that every segment that bears such a feature must distinctly specify which articulator is to execute them. Following Halle, we will call this the *Designated Articulator - X*. The x is a variable that functions as a place-holder for one of the six main articulators[15] (Halle, 2005, p. 31):

1. Labial (Lips)
2. Coronal (Tongue Blade)
3. Dorsal (Tongue Body)
4. Radical (Tongue Root)
5. Rhinal (Tongue Root)
6. Glottal (Vocal Folds)

Halle, 2005, p. 32 points out that in cases where a segment involves more than one articulator-free features they are executed by the same DAX. For instance, in the segment [f] both the radical narrowing of the air passage through the oral cavity, encoded by [+Consonantal], is executed by the lips. This same articulator is also resposible for executing [±Continuant] in this segment. Conversely, Ladefoged and Maddieson, 1996, 333ff. point out that in cases of labiovelar stops, such as [kp] and [gb], the feature [±Continuant] is executed simultaneously by two diferent articulators. This aspect, in particular, will be relevant to out exposition of Malayalam velar palatalization.

The idea of a designated articulator is not new, being present in, for instance, Sagey, 1986's discussion of feature geometry. Halle (2005), however, points out that Sagey assumed that DAX features encode a fundamentally different kind of information than other segmental features. Particularly, she assumed that this information is not subject to feature-spreading and

---

[15] Unlike other features, DAX features are assumed to be privative. While in binary segmental features both the +ve and -ve specification imply distinct functions of the articulators, for DAX features the absence of a specific feature simply implies that some other articulator executes that function.

was thus represented by the use of pointers. Following, Halle we will argue that DAX features do in fact spread, and must be represented as any other segmental feature in autosegmental notations. Halle, 2005, p. 32 provides the well-known example of nasal place of articulation in English, where the nasal sonorant surfaces with the place of articulation of the adjacent consonant – co[mp]etent, co[nt]inent, co[ŋg]regate etc. This process is formalized in the figure below, adopted from Halle, 2005, p. 32.

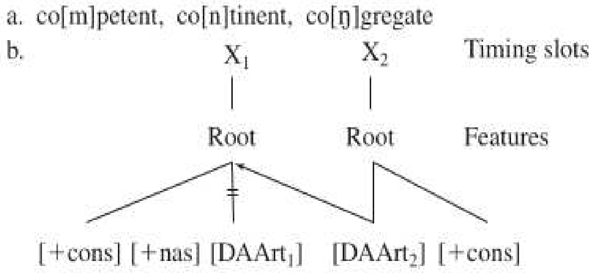


FIGURE 5.12: Nasal Place Assimilation as DAX Spreading. From Halle (2005)

Briefly, in the sub-figure (b) above the nasal (X_1 and the following consonant (X_2) have different designated articulators that execute the feature [+Cons] in each of them. During the derivation process the link to the [DAArt$_1$] node for the nasal is delinked, and the following consonant spreads its [DAArt$_2$] node to the preceding nasal. This results in the nasal surfacing with the designated articulator of the following nasal. Adopting a framework that allows DAX features to spread like any other feature allows for a straightforward account of velar palatalization, cross-linguistic in general and Malayalam in particular.

We will work with the assumption that [DADorsal] is involved in the articulation of all vowels (Halle, 1992). Further, Clements, 1993 and Halle, 2005 argue that [-Back] vowels should also be considered [Coronal], implying that these vowels are specified for two designated articulator features – [DADorsal] and [DACoronal](see figure below).

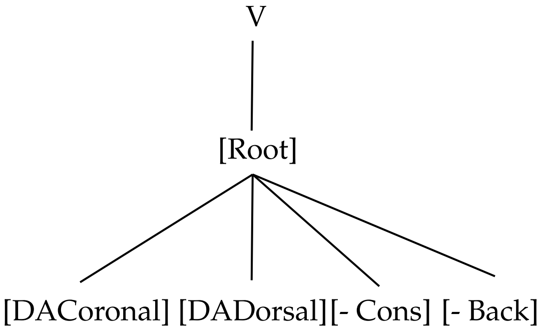


FIGURE 5.13: Feature tree for [−Back] vowels

Now consider the figure below. Here the single-melody velars (singletons and geminates) are specified for [-Back], [-Contin] and [DADorsal], indicating their status as velars. The preceding high front vowel, [i], is accordingly specified for [+High], [DACoronal], and the obligatory [DADorsal]. The rule for velar palatalization in figure 5.12 above, accordingly, spreads the [DACoronal] from the preceding [-Back] vowel to the following consonant. Crucially, in Malayalam the process of velar palatalization results not in the softening of the velar to a [s], but rather in the velar taking up a secondary palatal articulation – /k(k)/ → [k^jk^j].

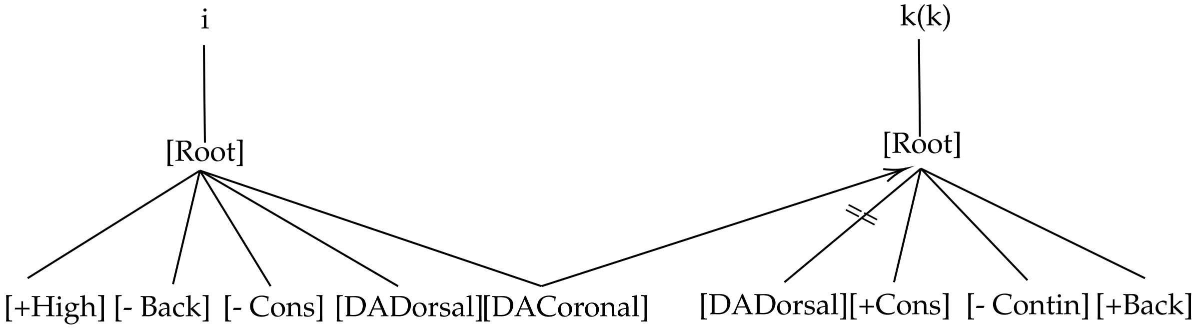


FIGURE 5.14: Malayalam velar palatalization after[i]

We indicate the feature-spreading with the directed arrow leading from the DAX feature of the following vowel to the root node of the preceding velar. The crossed out line to the original [DADorsal] of the velar indicates that this link is detached, and the designated articulator for the velar changes to [DACoronal]. However, note that this segment is still specified for [+Back], which crucially ensures that it surfaces as [$k^k(k^k)$] – i.e. a velar with a secondary palatal articulation. The process works similarly for the vowels [e] and [a], figures 5.15 and 5.16 respectively.

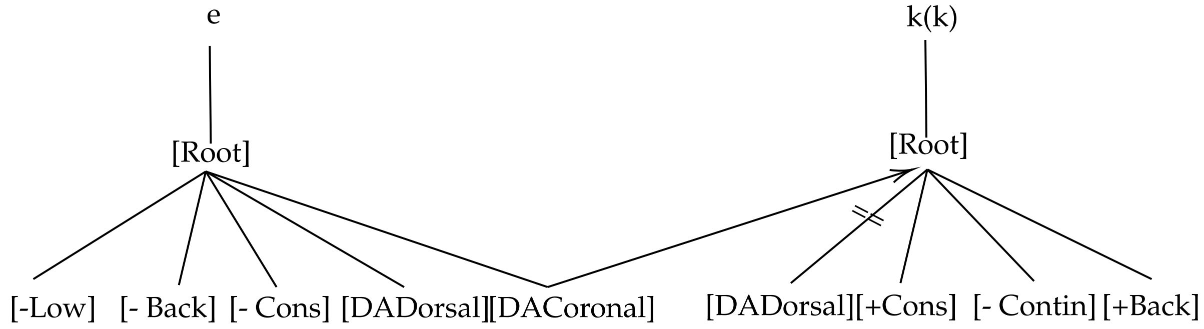


FIGURE 5.15: Malayalam velar palatalization after [e]

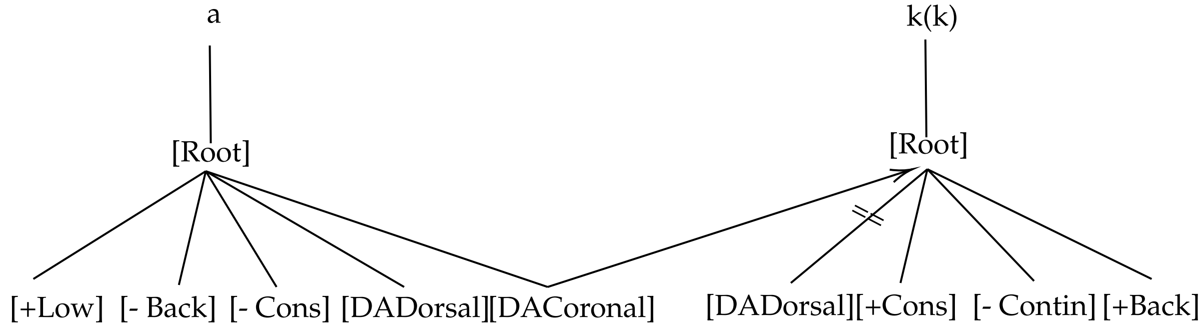


FIGURE 5.16: Malayalam velar palatalization after [a]

In figure 5.15 above the vowel is specified for [-Low], while in figure 5.16 the vowel is specified for [+Low]. The designated articulator configurations, and accordingly the process of DAX spreading, remain unchanged. This specification of the vowels the three Malayalam vowels [i,e,a] with an additional [DACoronal], out of the five vowel inventory [i,e,a,o,u], drives the palatalization process as only the former three meet the structural description of rule in figure 5.11.

One final aspect of the process begs explanation at this point. Recall from our discussions so far that the Malayalam plural suffix, -kal], does not undergo palatalization. Consider the data from 5.1.3 repeated below.

Ex 5.4.2. kutti ’child’ kuttikal ’children’
pu:cca ’cat’ pu:ccakal ’cats’

In the two examples above, the velar /k/ appears after /i/ and /a/ respectively. Recall, also, from 5.2 above, that both of these vowels trigger palatalization when followed by velar stops. The phonological account developed so far will apply to the forms in the example above and yield ungrammatical surface forms with palatalized velars. Mohanan and Mohanan, 1984 argue that in order to exempt the suffix-initial velar in -kal] from undergoing the rule it is necessary to posit a diacritic feature — ±P — such that only segments marked +P form input to the palatalization rule. Thus, they argue, the velar in the suffix -kal] is marked -P while those in the palatalizing suffixes are marked +P. We take issue with the postulation of diacritic features on the ground that such ad hoc stipulations are incompatible with generative grammar’s commitment to neurobiological realism. The distinctive features posited within phonological theory serve the purpose of abstractly encoding systematic computational contrasts, and features usually correlate to some substantive dimension via transduction. Postulating a feature of the form ±P raises the difficult question of how many such diacritics could reasonably be allowed within the grammar, and how such limits could be determined? Rather, we argue that the observed surface forms can be derived without appeals to such diacritics by reconsidering the contents of the Malayalam segmental inventory.

Consider the possibility that the Malayalam inventory consists of two distinct underlying voiceless velar stops — /k/ and /K/. The two segments bear the same phonetic realizations, but differ in their underlying representations. Specifically, while /k/ is underlying specified for DA:Dorsal, /K/ is underspecified for this feature. The proposal we have in mind is identical to the *inalterability as prespecification* hypothesis forwarded by Inkelas and Cho, 1993, and makes use of an underlyingly prespecified feature to block assimilation. Consider the figures below.

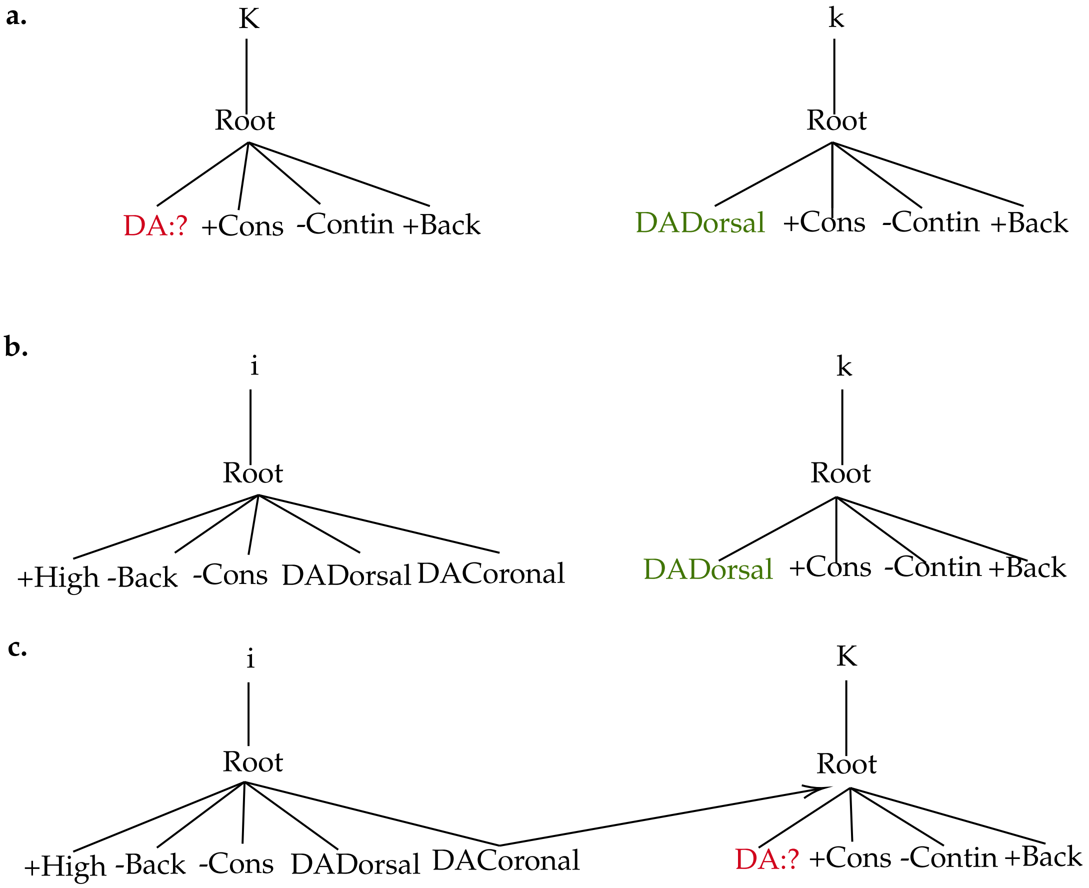


FIGURE 5.17: Prespecification of /k/

In sub-figure in 5.17a above the segment /K/ has an underspecified DA node (represnted by the red DA:?). In comparison, the regular /k/ is prespecified for DA:Dorsal. Assuming that /k/ appears in the Malayalam plural suffix (-kal]) while /k/ appears in -KKuka] and -KKə], it can be straightforwardly argued that the prespecification of the DA node in -kal] blocks the spreading of the DA:Coronal in -kal] (illustrated in sub-figure 5.17b), while the underspecified DA node in -KKuka] and -KKə] allows spreading to take place (sub-figure 5.17c). In this formulation we only need one further criterion that the palatalization rule can only spread DA:Coronal if the target is underspecified, but delinking a prespecified DA:Dorsal is not allowed. This approach is preferable in our opinion because it allows us to account for -kal] straightforwardly using the representational and computational machinery already available to the grammar, and making adjustments to (a) the underlying representations of Malayalam inventory and (b) the structural description of the velar palatalization rule. Crucially, note that in the account forwarded there are no appeals to any ad hoc diacritic features.

5.4.4 The Lower interface: explaining gradience and variation

As discussed in the preceding sections, since its inception the generative tradition has been modular in nature (Mandal, 2021; Scheer, n.d). Recall that in the inverted-Y model a firm distinction is made between one concatenative (morpho-syntax) and two interpretive systems (semantics and phonology) in EXT(ernalization). These systems are mutually unintelligible due to the distinct vocabularies they process. The substance-free program in general extends this modular approach to phonology and phonetics based on the observation that just like syntax and phonology, phonetics and phonology also make use of different conceptual granularities. Namely, phonetics is concerned with real-world and physiological phenomena that are continuous and analog in nature, while the vocabulary of phonology is completely discrete – a segment is either [+Continuant] or it is not, but never [2.573 Continuant], for instance.

Module	Vocabulary	Type
Morphosyntax	Person, Gender, Aspect, Tense, Number etc.	Algebraic/Discrete/Digital
Phonology	Labial, Coronal, Continuant, Dorsal, High, Low etc.	Algebraic/Discrete/Digital
Phonetics	VOT, Harmonics-to-Noise Ratio, Amplitude, Duration etc.	Continuous Analog

TABLE 5.6: Module-Specific Vocabulary

One consequence of modules being mutually incommunicado is the need for interface driven computations, where a transductive intermodular interface is tasked with format-conversion, converting information encoded in the vocabulary of one module – the output of this module and the input to the transduction process – into the vocabulary of another module – the output of the transduction process and the input to the second module. Thus, for instance, morphosyntax generates a tree stated explicitly in its own vocabulary consisting of non-terminal nodes (DP, TP, VP, V, v etc.) and terminal nodes where vocabulary items are inserted. These vocabulary items, as opposed to syntactic constituents (VP, DP etc.), that are endowed with a phonological form (the upper-interface of phonology with morpho-syntax). If phonological forms are substance-free then these forms, by themselves, are illegible to the sensory and motor systems that are tasked with production and perception of speech signals. The interface with phonetics (the lower-interface of phonology) is responsible for mapping these phonological forms onto their corresponding phonetic shapes (in the direction of production), and the other way round (for perception). Within the substance-free program there has been two broad proposals for handling the lower-interface. For the Concordia school the interface consists of a set of transducers, articulatory and auditory. Reiss, 2003, p. 306 argues that:

> Unlike the computations of the phonology, whose inputs and outputs are both described in the same representational alphabet, a transducer converts between different representational formats, or even between physical/neurological states and symbolic representations (see Pylyshyn, 1984)

The relevant mappings are shown in the figure below.

Relevant mappings (*I* and *O* are featural, symbolic representations):
– Grammar (Phonology): $I \longleftrightarrow O$
– Auditory Transducer: $O \longleftrightarrow$ *Auditory Percept*
– Articulatory Transducer: $O \longleftrightarrow$ *Gestural Score*

FIGURE 5.18: Transductive interfaces

This proposal is further developed by Volenec and Reiss (Volenec and Reiss, 2017; Reiss and Volenec, 2022a) who argue that their proposal adheres to strict modularity as neither the transducers nor the grammar can look into each other, as only the output of one module can be fed into the other one. This, however, leaves aside two issues. First, there is the issue of how articulatory and acoustic parameters are co-ordinated. In other words, how is co-ordinate transformation from acoustic to articulatory space, and vice versa, achieved? In the diagram above there is no indication of any interaction between the two transducers. Second, this conception of transduction appears to violate at least one condition of Fodorian modularity – proprietary vocabulary. Volenec and Reiss' transducers, by their own account, accepts a digital input and produces an analog output. This is either an oversight that consists of gross violation of modularity, or there is an implicit assumption that transducers are not modular. This latter case, if intended, will constitute a rather unique claim, though no such claim is made explicit by Volenec and Reiss, 2017. To our knowledge there are only two broad approaches to modularity – the standard Fodorian modularity (also accepted by Chomsky and Halle, 1965; Chomsky, 1985; Chomsky, 1980) holds that only input systems (which transducers should count as, and thus be subjected to the proprietary vocabulary requirment) and the evolutionary psychology (see Pinker, 2005; Pinker, 2003; Plotkin, 1997) position of massive modularity that views all aspects of the mind being modular. It is, of course, possible that these two phonology-phonetic transducers are conceived of in the second sense, and as modules whose vocabulary consists of both phonological and phonetic primes, but no such position is made explicit by Reiss.

In contrast, Tobias Scheer (Scheer, 2010a; Scheer, n.d) argues that if the mind is modular then all intermodular interfaces should adhere to a similar architecture. No one argues, for instance, that the interface between morpho-syntax and phonology is arbitrary and consists of a dictionary-based spell-out system. For instance, for weak verbs in English the specification of past tense in a node results in the insertion of *-ed* morpheme. Scheer points out that at least synchronically there is no systematic link between a morpho-syntactic feature and the vocabulary item that is inserted corresponding to it at the spell-out – -ed could as easily have been -lp given an alternative diachronic history of English. In this view the architecture of the intermodular interface remains the same at both the upper and lower interfaces of phonology. Just as a lexical spell-out inserts vocabulary items into the terminal nodes of the trees that morphosyntactic computations generate, a *post-phonological spell-out* system maps phonological features onto language-specific substantive correlates. This process is illustrated in the figure below, to which we suggest minor embelishments and provide an analysis-by-synthesis (Bever and Poeppel, 2010) driven account of, both, Malayalam velar palatalization (considering, in turn, first perception and then production) and the lower-interface in general.

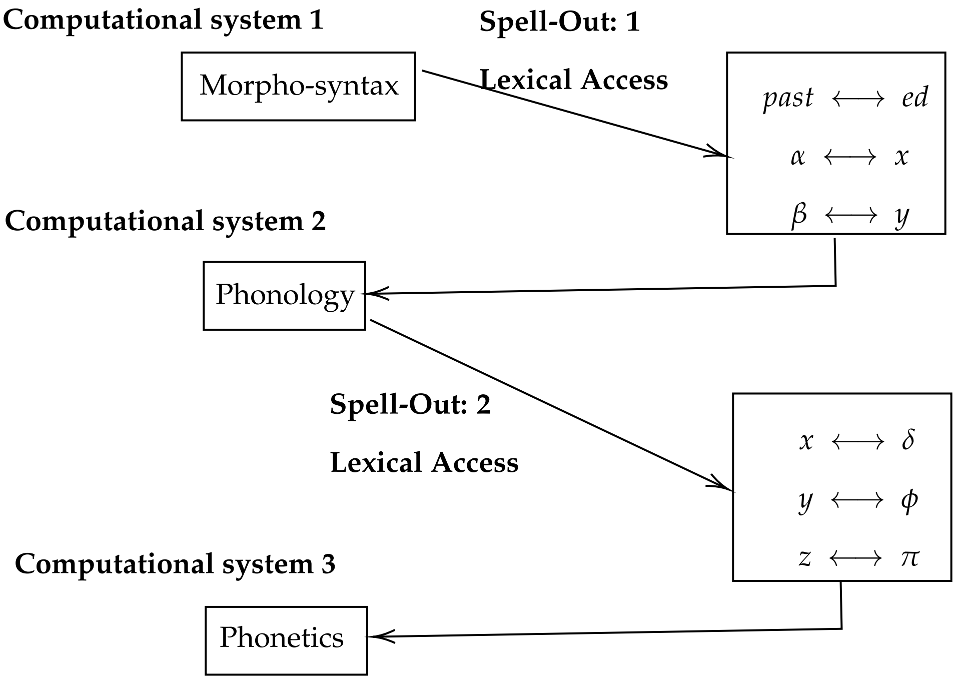


FIGURE 5.19: Upper and lower interfaces of phonology. From Scheer, 2010a

As a starting point we will adopt the hypothesis, emerging out of the last half a century's worth of phonological (Chomsky and Halle, 1968; Goldsmith, 1976b; Browman and Goldstein, 1989; Reiss, 1998; Halle, Vaux, and Wolfe, 2000; Halle, 2002) and psycho- (Norris, 1994; Cutler and Fay, 1982; Cutler, 2012; Levelt, 2013; Best, 1995b) and neurolinguistic (Poeppel and Idsardi, 2011; Berent, 2013; Poeppel and Embick, 2017; Hickok, 2022) that phonological primes, or distinctive features, form the basis of lexical representation of the words that are conveyed through an analog and varying signal. Emerging neurobiological evidence suggests that specialized sub-populations of neurons in the auditory-cortex display selective preferences for spectro-temporal properties (STRFs) in the signal, in other words distributions of frequency over time (Poeppel, Idsardi, and Van Wassenhove, 2008; Hickok and Poeppel, 2007; Hickok, Houde, and Rong, 2011; Kemmerer, 2014). These properties are argued to be hardwired (Poeppel and Idsardi, 2011), though they are usually assumed to be ecologically useful, shared across the mammalian brains and not specific to language (Fitch and Cutler, 2005). These neuronal preferences for STRFs afford a relatively high-resolution encoding of the signal, over which an initial estimate of major-class feature distributions can be made. The proposal we have in mind is based on Stevens, 2000′ *landmark hypothesis*(LH), and has been argued for by Poeppel and colleagues in conjunction with an analysis by synthesis (AxS) framework (Bever and Poeppel, 2010; Poeppel and Idsardi, 2011). The underpinning idea behind LH is that the speech signal is interspersed with *landmarks*– locations in the stream that are relatively rich in information regarding the configurations of the larynx and the vocal tract articulators. For instance, Stevens, 2000 points out that landmarks are often expressed in terms of expressed in terms of acoustic cues for consonant closures, consonant releases, glide minima and vowel maxima, and their distribution offers cues to the relative positions and distribution of the major class features in the signal. In other words, the neuronal representation of the acoustic cues in the signal triggers a hypothesize-and-test process that yields a probabilistic estimate of the distribution of major class features. Poeppel and Idsardi, 2011, p. 1079 argue:

> Frequency information is encoded throughout the system, so that the search can proceed on a 'best-first' basis, with more probable parses assigned greater weight in the system. More specifically, we see the landmarks of Stevens, 2002, which correspond to the articulator-free features of Halle, 2002, and which define the 'major' classes of phonemes (stop, fricative, nasal and approximant), as defining a PPS [phonological primal sketch] of the segmental time scale. This primal sketch gives a neighbourhood of words matching the detected landmark sequence.

Once this distribution of major class features is attained, what the authors above term the primal sketch, conditional probabilities are then evaluated around these landmarks to obtain the articulator-bound features.

> That is, we see this layer of analysis as evaluating conditional probabilities of the sort p([labial] | [+nasal]). The hypothesized temporally synchronized feature sequences are then matched against the main lexicon and a list of candidates is generated, and then the rules of the phonology of the language are used to resynthesize the predictable features, again using Bayes's rule [...]
>
> **Poeppel and Idsardi, 2011, p. 1080**

The f illustrates the general idea of feature detection using a conjunction of AxS and multi-time resolution processing. Here, the boxes at the lowest level represent the distinct levels of mapping from "vibrations in the ear" to "abstractions in the brain". The topmost tier represents the AxS module, while the ones in between represent the hypothesized computational processes that interact with the top and the bottom.

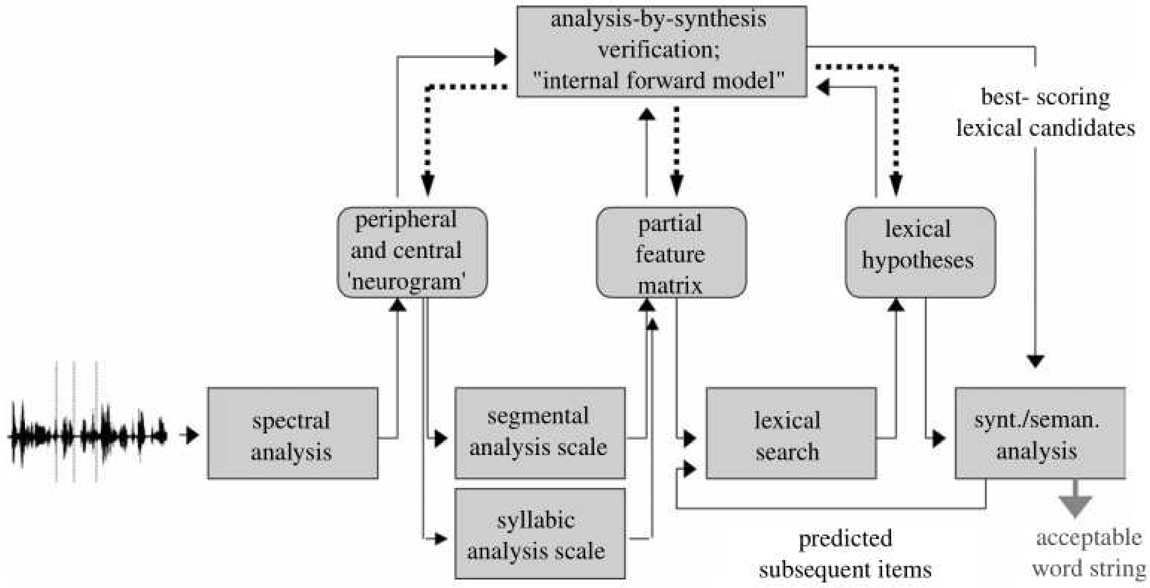


FIGURE 5.20: Processin steps in AxS model.

Crucially, note that while the neurogram is used to generate a PPS by the internal forward model, which in turn yields a cohort of possible lexical matches, the lexical hypotheses talk back to the internal model, which is in turn used to re-evaluate the neurogram. In this way, both bottom-up and top-down processes mechanically interact until a best match is found. We argue that this process plays a crucial role during the critical period in the construction of the Scheer, n.d's post-phonological spell-out lexicon (Lexicon 2 in figure 19 above).

In the view adopted here, we propose that the contribution of phonological UG to the adult synchronic grammar take the form of (a) an instinct for the set of possible featural contrasts in any natural language (distinguished into major class features that correspond to broad phonetic qualities resulting from totality of articulator configurations, and articulator-bound features that link possible configurations of specific articulators with expected acoustic targets), and (b) a a priori limited set of computational operations available to manipulate these symbols algebraically. The computations themselves are substance-free, as are the features. Exposure during the critical period allows the acquisition process to fine-tune the active and necessary contrasts in the language being acquired through the AxS process outlined above. Crucially, note that in order for this model to work it is essential to have available, at birth, both an idea of the possible contrasts and the ability to segment and analyse substantive cues in order to assign readings to features. In the figure below, the oval shape on the right represents the set of possible feature-based contrasts allowed for any natural language. We take this set, an expectation of what is at all possible, to be given by UG [16]. But what we have in mind is the rather trivial claim that there exist clear correlations between features, such as [+Round] (or [Round] if one prefers a privative system) and [-Back] and the resulting substantive cues – in the case of the former rounding of the lips (articulation) and a decrease in frequencies across the spectral range, and pushing forward of the tongue body and an increase first formant values.

---

[16] In our view, however, this endowment constitute faculty of language in the broad sense (FLB; Hauser, Chomsky, and Fitch, 2002), and is not language-specific

In this sense, then, features and their readings are partially-veridical, rather than completely arbitrary or deterministic.

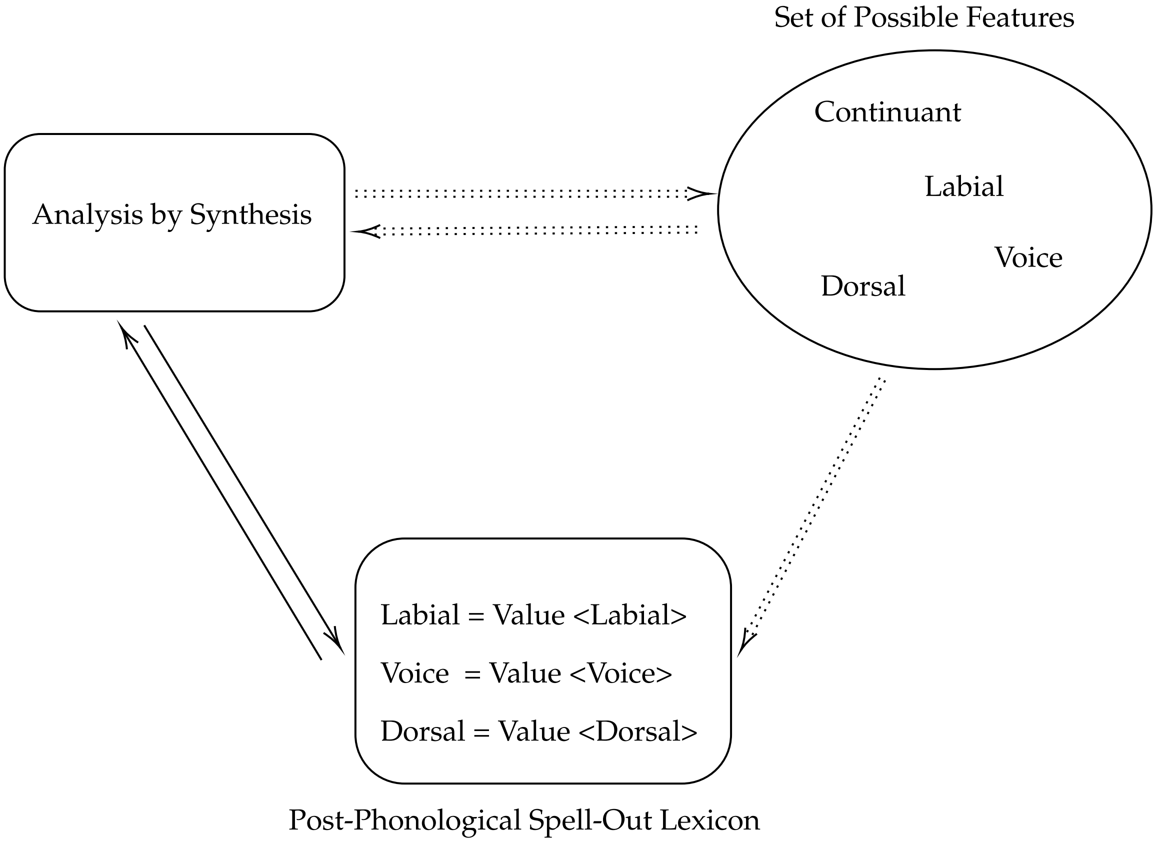


FIGURE 5.21: Post-phonological lexicon and spell-out

Within this framework, the results observed in our experimental study are rather to be expected. Consider the fact that the more fronted the contextual vowel the lesser the phonetc dissimilarity between an adjacent velar and its corresponding palatalized form. Thus, phonetically more fronted vowels are more likely to reliably yield a match with the learned cues for a palatalized velar. Conversely, the further back the vowel is articulated the more towards the edges of the permissible values the consonantal cues. In other words, while the phonological grammar presented in the preceding section will always transform all velars following a [-Back] vowel into palato-alveolars, the gradient vowel frontedness effects implies that less fronted values will yield phonetic cues that fall towards the edges of the learned cue-values. We see this most clearly in the fact that while all three vowels show a gradient effect in triggering rate, the least fronted vowel [a] is the only one that shows a statistically significant difference. Likewise, in production, while the output of phonological computations demand that all single-melody velars be palatalized following [-Back] vowels, hitting the required acoustic targets becomes less viable in the context of vowels articulated further back. This fact follows straightforwardly from articulatory inertia, especially in Malayalam where the vowel precedes the target consonant.

In our view the ability to configure the articulators to hit an acoustic target, and the expected acoustic signature of a particular configuration, are both innate. Though it is highly unlikely that such innate predispositions are specific to FL (Fitch, 2010; Fitch and Cutler, 2005; Samuels, 2015a). This line of reasoning is not new and follows straightforwardly from the FLN-FLB distinction made by Hauser, Chomsky, and Fitch, 2002. The process of acquisition, indicated by the dotted arrows, makes use of AxS in order to determine (a) which features are active in the language being acquired, and (b) what is the range of possible of values that correspond to this feature (i.e. its acoustic signatures, and necessary articulator configurations). The former helps create a language-specific set of active features, while the latter creates a language-specific post-phonological spell-out lexicon. It is important to note that we do not suggest that the values acquired are in some sense point-values. Rather, keeping in track with the probabilistic nature of the internal feedforward process, we hypothesize that the learner constructs probabilistic ranges for each feature-value – a mass of probabilities defining an acoustic map that is linked with specific articulatory configurations. The one way dotted arrow between the oval set of possible features and the lexicon indicates the critical-ness of the critical period, suggesting that once L1 contrasts have been lost the post-phonological lexicon remains frozen in time. On the other hand, the interaction between the AxS module and the post-phonological lexicon is an active process in all synchronic I-languages, indicated by solid two-way interaction arrows. The combination of top-down and bottom-up processes try to match observed values in the signal to a possible feature-value combination in the lexicon, until the best match is found. Given the probabilistic nature of this process, of course, this process is liable to performative failures, leading to misperception.

The defining property of the proposal presented here lies in the fact that it maintains, both, strict-modularity and an entirely substance-free phonology (both representations and computations), while simultaneously affording phonetic naturalness all the explanatory burden necessary. Phonological features in this framework are neo-Sausseurean symbolic markers, and phonological UG (FLB) endows every child with an expectation for the set of possible contrasts. The phonetic module works in close tandem with known neurophysiological mechanisms, including AxS, sensory transductions and landmark detection, while the task of mapping between the two is attained via a post-phonological spell-out that makes use of a lexicon created during the critical period, and does not require operating over distinct vocabularies. In this way, we try to have our cake and eat it too.

5.5 Conclusion

Our purpose here has been two-fold. First, we reported on two behavioral psycholinguistic studies investigating velar palatalization in Malayalam. The results of the experiments exhibit two interesting phenomena. First, in grammatically unlicensed contexts velar palatalization ignores all notions of phonetic naturalness and consistently fail to apply. Second, where demanded by the grammar of Malayalam velar palatalization applies with statistically significant rates across both natural and unnatural contexts. The results, however, also showed an observable effect of gradient vowel-frontedness. We then discussed our findings within a strictly modular approach to phonology and phonetics, and proposed a phonological rule for palatalization utilizing Halle, 2005 designated articulator theory. We then proposed an account of the interface that operates largely along Bayesian probabilistic lines, and maintains partial-veridicality between features and their phonetic interpretations. It was argued that the gradient effects observed in our study fall out of the interface interactions between the phonological grammar and the interpretation of the features over which the grammar operates.

5.6 Special acknowledgment

The data for this research was collected at the MARCS Institute for Brain, Behavior and Development in Sydney. I will like to thank Prof. Catherine T. Best for the support she provided in designing the experimental procedures, and throughout the process of data collection and analyses. I also need to thank Prof. Jason Shaw for the help he provided with acoustic analyses and statistical design, and also for the many hours of his time he has spared discussing phonetic theory with me. This work is currently being finalized for publication in Phonetica.

Chapter 6

What phonology is NOT

6.1 Introduction

Within the cognitive neurosciences the neurobiological implementation of linguistic operations and structures are typically less well-understood than, say, vision or audition. This is partly because unlike the latter there exists no animal analog for language, and partly because human-compatible tools like functional magnetic resonance imaging or studies on brain surgery patients are coarser in space or time than the underlying linguistic events in neural circuitry. A seminal study by Sahin et al. (2009) recorded local field potentials from depth electrodes implanted in specific cortical regions – a methodology with high spatial and temporal resolution – to report that lexical retrieval of words from memory (e.g. reading a word, "dog"), concatenation of units (dog+[z] = dog[z]) and phonological processing (cat+[z]=cat[z]→cat[s]) occur in a specific order, with respective temporal signatures at 200ms, 320ms and 450ms respectively. Based on this finding Scheer and colleagues (2020) have argued that a 450ms entrainment can be reliably used as a neurophysiological detector for phonological computations in the brain. An initial study replicated the results from Sahin et al. (2009) by recording event-related potentials (ERP) using electroencephalography (EEG), while also obtaining neurophysiological evidence to support the argument that English velar softening – i.e. the [k]~[s] sound alternation in words like electri[k] ~electri[s]ity – involve phonological computations, with a similar

450ms ERP signature. This is welcome result, especially considering that ERPs are easier to record than depth electrodes' local field potentials, but nevertheless the results retain certain ambiguities. Linguistic theory is characterised by the fine-grained distinctions it makes between various levels of computations – syntactic, morphological, phonological and phonetic. The last two are particularly important because one too many phonological phenomena hew too close to the phonetic bone, making it difficult to objectively determine when a certain alternation falls under the purview of the phonological as opposed to the phonetic module. This is particularly true of the velar softening study by Scheer and colleagues, and as such I describe in this chapter a project that specifically seeks to use ERPs to distinguish phonology from analogy, and subsequently from phonetics, and does so across unrelated languages.

6.1.1 The issue of types of alternations/computations

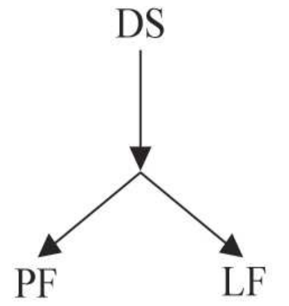
A cornerstone of the cognitive and generative revolutions in linguistics entails the recognition of abstract and hierarchical levels of representations of native speakers' knowledge of their language. This knowledge is abstract because it encodes only those aspects of language that play a linguistically relevant role in a system of grammatical computations and not high resolution recordings of acoustico-articulatory and statistical information – in Saussure, 1916[1967], p. 166 terms *dans la langue il n'y a que des différences* – and also because the nature of this knowledge is implicit and not directly accessible to introspection. It is also hierarchical in that the primitive units of representation, the syntactic (tense, aspect etc.) and phonological (labial, coronal etc.) primes, can be combined and recombined to yield increasingly larger granularities (DP, NP, segment, rhyme, mora etc.) that form levels of representations – distinctive features group into segments, which can be grouped into syllables, syllables into words into phrases and so on. Thus, for instance, to produce a verb embedded in a sentence, say 'talk', one must first determine the appropriate morphosyntactic context which determines whether a present or past tense form is required, and then fetch the relevant form from the lexicon and apply the necessary alternations to pronounce either 'talk' (present) or 'talked' (past). Sahin et al., 2009 point out that this superficially simple phenomena involve multiple distinct levels of neurophysiological processes – lexical retrieval of a form, concatenative computation (talk + past tense suffix), phonological alternations (e.g. a vowel is inserted between the base form of the verb and the suffix in pairs like 'pat and patted' but not in 'talk-talked'), and subsequent processing that determines the shape of the externalized form (e.g walk+[d]=walk[t], but jod+[d]=jog[d]) and assigns sensory-motor commands for articulation. The conceptual necessity of such logical decomposition is rather well-accepted in the linguistic theory, but whether each representational level corresponds to distinct stages and/or circuits of processing in the brain, and if they do what their spatial and temporal signatures are, is not quite as well-studied in the neuro- and psycholinguistic literature, often leading to what Poeppel and colleagues (Embick and Poeppel, 2015; Poeppel and Embick, 2017) term the *granularity mismatch* problem in cognitive neurosciences.

The granularity mismatch problem has mostly been approached from the perspective of neuroscience (Poeppel and Idsardi, 2011; Krakauer et al., 2017; Poeppel and Embick, 2017), although it has non-trivial consequences for linguistic theory as well. Among other reasons, since its inception the generative program has been realist in its assertions — i.e. the primitives and principles postulated by generative grammar are taken to be real neurobiological events and processes instantiated in the brain (Chomsky, 1959; Chomsky, 1980; Chomsky, [1966] 2009; Chomsky, 2005; Piattelli-Palmarini and Uriagereka, 2008; Scheer, 2022). If this is the case then linguistic processes must exhibit unique neurophysiology and brain functioning. Conversely, if two (neurobiologically real) lingustic processes are claimed to be different in type – e.g. syntactic versus semantic mismatches in pronoun binding — then the brain must have some ways of telling them apart. In principle the separation in *type* of linguistic phenomena could be detected at the implementational level (Marr, 2010) either through spatial (specific types are localized to specific neural anatomy) or temporal (specific types exhibit specific temporal latencies and windows) means. Thus, for instance, Moro et al., 2001 utilized functional localization of syntactic and morphological processing by positron emission topography (PET) scans to report that a selective deep component of Broca's area and of a right inferior frontal region is involved in both syntactic and morphosyntactic processing, while within this system, the left caudate nucleus and insula respond only to syntactic processing demands. Based on spatial localization of linguistic processes Moro and colleagues provided in vivo evidence for the existence of specialized language networks in the brain, with specific sub-components exhibiting preferences for specific levels of linguistic structure. Such studies help validate the realist position of generative grammar, and contribute significantly to theory building by validating sub-portions of broader hypotheses.

Recall, here, that generative grammar has a modular architecture (Scheer, 2010a), with specific modules of grammar carrying out computation of specific levels of the linguistic hierarchy – lexical access, phonological alternations, syntactic concatenation etc. The idea that sub-parts of the grammar, like syntax and phonology, are separated modules is elaborated on by Chomsky et al., 1963, p. 308 who argue that the phonological component of language faculty, for instance, contains only those processes that are responsible for ascribing a phonetic shape to the terminal elements that appear in the output of syntactic trees:

> The phonological component can be thought of as an input-output device that accepts a terminal string with a labelled bracketing and codes it as a phonetic representation.

Scheer, 2010a, p. 541 points out that in this early modular conception of grammar the three internal components of the inverted-Y structure – DS (deep structure in even earlier terminology) ships off its output to two distinct modules (PF and LF) each of which is blind to the other one – are granted modular status, i.e. they constitute grammar qua grammar. Their interaction with the conceptual device and pragmatics, both of which are grammar-external, produces externalized language.



Inverted Y-model

Further work in the wake of the inverted-Y model has developed the modular conception of the faculty of language, arguing that modules are informationally encapsulated computational systems specialized for particular kinds of computations,working with a proprietary vocabulary (Fodor, 1983). Particularly relevant for the current discussion is the substance-free (Hale and Reiss, 2008; Chabot, 2021; Scheer, 2022) argument that, not unlike morphosyntax and phonology, phonetics is also an independent module concerned with substantive considerations that is phonology-external. This is crucial because, as discussed in the preceding chapters, a large many phonological phenomena hew relatively close to the phonetic bone. If phonology and phonetics are separate modules however, operating over independent vocabularies, then there must be some objective yardstick to determine when some phenomena actually falls under the purview of phonology.

A modular architecture of grammar demands a modular (set of) theories. To an extent this is observable in the nature of the primes in term of which various sub-theories in linguistics are constructed. In morphosyntax theories are built around number, aspect, tense etc., while in phonology we typically talk about coronals, labials, velars etc. The problem encountered with phonology stems from the fact that, unlike the primes of morphosyntax which have no observable real world correlates, phonological primes are routinely intepreted by the sensory-motor systems, endowing them with phonetic shapes, and subsequently acoustic and articulatory forms. Thus, for instance, a segment specified [+Labial] inevitably leads to the rounding of the lips through the enervation of the orbicularis oris muscle, accompanied by a down sweep in spectral frequencies. One of the core concers of substance-free phonology is that phonetic interpretation of phonological primes requires transduction through an inter-modular interface, and as such phonetic effects observed in speech are not directly attributable to phonology (Reiss, 2017; Chabot, 2022a). This raises the obvious question of how do we decide when something *is* attributable to phonology, especially given the fact that for various reasons, including but not limited to diachrony and phonologization, phonological and phonetic processes often mirror each other. Considering the fact that the purview of a theory, any theory, is a critical component that decides its shape, which in turn helps narrow down what novel phenomena can be accommodated within its limits, this is not a trivial issue.

To summarise, the issue of granularity mismatch between linguistics (as cognitive science) and neuroscience are of non-trivial significance to any attempts to properly align theories of cognition with neurobiological theory. The former studies computational capacities and properties of the human mind, which itself is an abstract characterization of the brain. The latter studies anatomy and physiology of the functioning brain which instantiates the mind, along with all its purported properties. For linguistics, a field whose primary importance lies in its focus on (perhaps) the only uniquely human aspect of cognition, Language, this directly translates to elaborating on how the intricate levels of computations uncovered by linguistic theory can be vindicated by the functioning brain, and where linguistic operations can better help us understand the functional properties of neurophysiology and anatomy (Poeppel and Embick, 2017).

6.2 Neurobiology of language

The earliest studies on language and brain date back to the early 19th century with the foundational post-mortem lesion studies by Paul Broca and Carl Wernicke (Broca et al., 1861; Schiller, 1992; Blank et al., 2002; Poeppel et al., 2004). These studies were almost exclusively focused on documenting the lesion patterns in patients suffering from aphasia – a condition where patients are unable to process speech sounds, even though their non-linguistic audition, and indeed general cognitive functions remained unimpaired. Paul Broca's seminal contribution came in the mid-19th century when he discovered that damage to a specific area in the left frontal lobe of the brain resulted in language impairment. This region, now known as Broca's area, was linked to expressive language function. Broca's landmark case study of "Tan," a patient who could only utter the single syllable "tan" due to a lesion in this area, demonstrated the localized nature of language processing within the brain (Broca et al., 1861; Blank et al., 2002). Building upon Broca's findings, Carl Wernicke identified another critical area in the posterior portion of the left hemisphere associated with receptive language comprehension. In his work, Wernicke observed patients with damage to what is now termed Wernicke's area, who exhibited fluent but nonsensical speech, a condition known as Wernicke's aphasia (Keyser, 1994). This delineation between expressive and receptive language deficits provided crucial insights into the neural organization of language. Ludwig Lichtheim further expanded upon these ideas with his famous model of language processing, known as the Wernicke–Lichtheim model. This model proposed a connectionist framework wherein information flowed through pathways connecting auditory comprehension (Wernicke's area), speech production (Broca's area), and conceptual processing (posterior association areas). Lichtheim's model elegantly explained various language phenomena, such as conduction aphasia, where patients have intact comprehension and production but struggle with repetition due to damage to the arcuate fasciculus connecting Wernicke's and Broca's areas (Lichtheim, 1885). Together, the works of Broca, Wernicke, and Lichtheim provided a comprehensive understanding of the neural basis of language, laying the groundwork for modern neurolinguistic research and clinical practice. Their contributions remain foundational in the field, guiding investigations into language disorders and brain-behavior relationships. Likewise, Lenneberg, 1967; Lenneberg, 1969 points out that these early studies were the first set of evidence that cortical regions displayed selective preference for linguistic processes.

Early functional neurolinguistic studies began appearing in the later-half of the 20th century, bolstered by the availability of neuroimaging techniques, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). One landmark study in this domain was conducted by Peterson et al., 1988, who used PET scanning to investigate regional cerebral blood flow changes during language tasks. Their findings revealed distinct patterns of activation in various brain regions depending on the linguistic task performed, providing the first direct evidence of functional specialization within the language network. Subsequent studies utilizing PET and fMRI further elucidated the neural substrates of language comprehension, production, and semantic processing. For instance, studies by Price, Wise, and Frackowiak, 1996 and Binder et al., 1997 highlighted the involvement of the left inferior frontal gyrus (Broca's area) and posterior superior temporal gyrus (Wernicke's area) in syntactic and semantic processing, respectively. These early functional neuroimaging studies not only confirmed the localization of language functions within specific brain regions but also demonstrated the dynamic interplay between different cortical areas during language tasks, providing a nuanced understanding of the distributed neural networks underlying language processing. Moreover, functional neuroimaging techniques allowed researchers to explore language processing in various populations, including individuals with neurodevelopmental disorders and neurological conditions affecting language function. For instance, studies employing fMRI have investigated language impairments in conditions such as autism spectrum disorder (ASD) and specific language impairment (SLI), shedding light on the atypical neural organization of language in these populations (Just et al., 2004; Redcay and Courchesne, 2008). Additionally, functional neuroimaging studies have also contributed to our understanding of language recovery and plasticity following brain injury, such as stroke-induced aphasia. By examining changes in brain activation patterns over time, researchers have identified neural mechanisms underlying language rehabilitation and the potential for functional reorganization in spared brain regions (Thulborn, Carpenter, and Just, 1999; Saur et al., 2006). These early functional neuroimaging studies not only advanced our knowledge of the neural bases of language but also paved the way for translational applications in clinical diagnosis, prognosis, and therapeutic interventions for language-related disorders.

6.2.1 Correlational behavioral neuroscience

A turning point in the study of the functioning brain in recent times have come in the form of what John Krakauer and colleagues have termed correlational behavioral neuroscience (Poeppel, Idsardi, and Van Wassenhove, 2008; Krakauer et al., 2017). Correlational behavioral neuroscience represents a paradigm shift in the field, and advocates for a more holistic understanding of brain function and behavior than is possible by just imaging. This approach, articulated in their influential 2017 paper, emphasizes the reciprocal relationship between neuroscience and behavior, challenging the reductionist view that neural activity can be fully understood in isolation from the complexities of real-world behavior. Krakauer et al., 2017 assert that neuroscience must integrate behavioral observations to comprehensively elucidate the neural basis of cognition and action. By emphasizing the bidirectional interplay between brain and behavior, correlational behavioral neuroscience aims to uncover the intricate relationships between neural activity, environmental factors, and individual differences in behavior.

In their paper, Krakauer et al., 2017 critique the prevalent reductionist bias in neuroscience, which often prioritizes controlled laboratory experiments over naturalistic observations of behavior. They argue that while experimental manipulations are valuable for probing specific hypotheses, they may oversimplify the rich tapestry of real-world behaviors and fail to capture the full complexity of neural processing. Instead, correlational neuroscience embraces the diversity and variability inherent in human behavior, recognizing that neural activity arises from the dynamic interaction between genes, environment, and experience. By leveraging large-scale datasets and advanced analytical techniques, such as machine learning and network science, researchers can identify robust correlations between neural patterns and behavioral outcomes across diverse contexts. This approach enables the discovery of neural signatures associated with individual differences, developmental trajectories, and pathological conditions, offering insights into the underlying mechanisms of cognition and action. Furthermore, the authors also highlight the importance of interdisciplinary collaboration in advancing cognitive neuroscience. They advocate for the integration of expertise from neuroscience, psychology, computer science, and other fields to tackle complex cognitive issues and develop linking hypotheses between brain function and behavior. By combining neuroimaging techniques with behavioral assays, genetic analyses, and computational modeling, researchers can elucidate the multifaceted nature of brain-behavior relationships. For example, studies employing multimodal imaging approaches have revealed how neural networks dynamically reconfigure during cognitive tasks, providing a nuanced understanding of brain function. Moreover, correlational neuroscience holds promise for translational applications, such as personalized medicine and targeted interventions for neurological and psychiatric disorders. By identifying biomarkers and predictive markers of treatment response, researchers can optimize therapeutic strategies and improve outcomes for patients.

6.2.2 Towards a new neurobiology of language

The emphasis on correlational neuroscience has been influential in neurolinguistic studies by virtue of its emphasis on what (Embick and Poeppel, 2015) call the *granularity mismatch* and *ontological incommensurability* problems. The former, as pointed out above, stresses that linking cognitive science (including, but not limited to, linguistics) requires nuanced linking hypotheses, because the two fields appeal to conceptual primitives of different granularities. The latter further points out that cognitive and linguistic ontological primes are not directly reducible to neural wetware. This approach, which Poeppel and Embick (Poeppel et al., 2012; Embick and Poeppel, 2015) have termed *a new computationalist neurobiology of language*, emphasizes the importance of integrating behavioral observations and neuroimaging data to unravel the complexities of language comprehension and production. Poeppel's work, for instance, focuses on the dynamic neural mechanisms underlying speech perception, exploring how auditory signals are processed and transformed into meaningful linguistic representations. Through sophisticated neuroimaging techniques such as magnetoencephalography (MEG) and electroencephalography (EEG), Poeppel and colleagues have identified neural oscillations in the gamma and theta frequency bands that are associated with phonological and syntactic processing, respectively (Giraud and Poeppel, 2012). These findings highlight the distributed nature of language processing in the brain, with different frequency bands serving distinct linguistic functions. Hickok, on the other hand, has contributed to our understanding of the neuroanatomical substrates of language, particularly through the dual-stream model of language processing. According to this model, proposed in collaboration with David Poeppel, language comprehension and production are supported by two parallel pathways: the dorsal stream, responsible for mapping sound to articulatory motor commands, and the ventral stream, involved in mapping sound to meaning (Hickok and Poeppel, 2004) By examining lesion data and functional neuroimaging studies, Hickok and colleagues have provided empirical support for the dual-stream model, elucidating how disruptions to these pathways can result in distinct language deficits such as conduction aphasia or semantic dementia (Hickok, 2022).

Likewise, Monahan and colleagues' works have focused on bridging the gap between neural activity and linguistic behavior through computational modeling and large-scale data analyses. Monahan and collaborators have developed novel frameworks for analyzing natural language processing using machine learning algorithms and neural network models. For example, their recent study employed deep learning techniques to investigate semantic representations in the brain, revealing that hierarchical neural networks trained on linguistic corpora could predict brain activity patterns during semantic processing tasks with high accuracy (Monahan, Lau, and Idsardi, 2013). This approach highlights the potential of computational methods to uncover the latent structure of language in the brain and to elucidate the neural mechanisms underlying language comprehension and production. Together, these works exemplify the power of correlational neuroscience in advancing our understanding of language processing, from the basic perceptual mechanisms of speech to the higher-order cognitive functions involved in semantic and syntactic analysis. By integrating behavioral data, neuroimaging findings, and computational modeling approaches, researchers can gain deeper insights into the complex interplay between neural activity and linguistic behavior, paving the way for more comprehensive theories of language processing in the human brain.

Perhaps the most phonologically relevant work in correlational neuroscience has been carried out by Aditi Lahiri and colleagues, investigating the neural basis of phonological underspecfication. Phonological undespecification is the theoretical claim that not all segments are specified for all features in the mental lexicon. Indeed, for reasons of economy, it is often assumed that certain features, for instance the feature [Coronal], remains unspecified because it is, in a sense, a default feature that can be predicted by the computational system. Aditi Lahiri's innovative work in using EEG to explore underspecification in phonology has been pivotal in unraveling the neural mechanisms underlying phonological processing in the human brain. Lahiri and her colleagues have developed a sophisticated experimental paradigm that utilizes EEG to examine how the brain responds to phonological features that are either present or absent in speech stimuli (Lahiri and Reetz, 2002; Lahiri, 2012). The team implemented a systematic approach comprising three conditions: match, mismatch, and no-mismatch. In the match condition, both speech sounds contain the target phonological feature, eliciting a typical neural response associated with processing phonologically congruent stimuli. Conversely, in the mismatch condition, one speech sound contains the target feature while the other lacks it, leading to a neural response reflecting the detection of a phonological discrepancy. Finally, in the no-mismatch condition, both speech sounds lack the target feature, providing a baseline comparison for neural activity when no phonological contrast is present (Lahiri, 2012).

One notable study by Lahiri and colleagues focused on investigating the neural correlates of underspecification in vowel nasalization, a common phonological process in various languages. Participants were presented with pairs of vowels, with one vowel nasalized and the other oral, resulting in three conditions: match (both vowels nasalized), mismatch (one nasalized, one oral), and no-mismatch (both oral). The study analyzed EEG data to examine the brain's response to the presence or absence of nasalization in speech sounds. Results revealed that during the mismatch condition, characterized by a nasalized vowel followed by an oral vowel or vice versa, there was a notable increase in the amplitude of the N1 component, an early negative deflection in the EEG signal associated with auditory processing. This enhanced N1 response indicated the brain's sensitivity to the phonological mismatch between nasalized and oral vowels, demonstrating the detection of a phonological contrast. Importantly, during the no-mismatch condition, where both vowels were oral, no such enhancement in the N1 component was observed, confirming that the heightened neural response in the mismatch condition was specific to the presence of phonological discrepancy (Kotzor, Wetterlin, and Lahiri, 2017).

6.3 Plastic and stone: neurobiological disambiguation of what phonology is not

The input to linguistic reasoning are patterns of alternations observed at various levels of representations. For instance, in the domain of morphology the definite masculine article in Italian is realized as *lo* before words beginning with /sC/ (*lo studio* 'the studio') or before geminate-initial words (*lo gnomo* [lo ɲɲomo] 'the gnome'), and *il* elsewhere (*il treno* 'the train') (Scheer and Mathy, 2021). Likewise, in the phonological/phonetic domain either [k] or [s] realize the stem-final consonant of the same stem electric, in electri[k] and electri[s]-ity respectively. This specific example is of particular interest as, despite over a half a century's worth of research, there appears to be little consensus on whether this alternation is actually within the purview of phonological theory, or for that matter on what constitutes a phonological alternation at all. (Scheer and Mathy, 2021) point out that there are, at least, three distinct analytical options available within linguistic theory for the aforementioned velar softening phenomenon. First, one could argue that there is no computation involved at all, and electri[k] and electri[s]ity are stored as distinct wholistic entries, requiring simple 1° suppletion. Second, one could argue that this is a case of allomorphy – there are two root allomorphs stored, electri[k] and electri[s], and the latter is marked for concatenation with -ity. This involves a retrieval (suppletion) of the two pieces (the allomorph and the suffix) and their concatenation, or 2° allomorphy/concatenation but no phonological computations. The third option, 3° phonological computation, involves the argument that only one root allomorph is stored, electri[k], and phonological computations turn electri[k]+ity into electri[s]ity. This last case involves proper phonological computations, on top of the usual suppletion/retrieval and concatenative computation.

This issue raises concerns that are relevant for both linguistic and psycholinguistic theories of representation and processing respectively. On the linguistic side, the issue relates straightforwardly to what Bermúdez-Otero and McMahon, 2006 call *the Gordian knot of the discipline* – namely, without a scientifically agreed upon yardstick to determine what counts as phonology, qua phonology, Popperian competition among competing flavors of phonological theories becomes impossible. For obvious reasons, a theory of phonology that assumes the explanatory burden for velar softening will have an architecture that is visibly different from one that considers such phenomena to be outside of the remit of phonology proper (Scheer and Mathy, 2021). In order to ascertain the complexity of phonological representations and computations in the mind, and thus by extension in the brain, one must first delineate the space of interactions between phonology, morphology and morphosyntax, and the lexicon, and establish with scientific clarity what constitutes the input to phonological reasoning.

For those wearing a psycholinguistic hat there are at least two related issues pertaining to processing routines in language production – namely, language-specificity and alternation - specificity of processing routines (Scheer and Mathy, 2021). A processing routine can be non-specific, or universal, in the sense that the same processing routine holds for all alternations in a language, or even for all alternations across all languages. This is the position argued for, for instance, by Bürki, Ernestus, and Frauenfelder, 2010 who, studying French liaison in the indefinite article un (un [ɛñ] accident 'an accident' vs. un [ɛ̃] café 'a coffee"), first conclude that there are two lexical entries for the definite article ([ɛñ] and [ɛ̃], and then generalize this to all languages and claim that systematic variations across languages are better explained through positing multiple lexical entries than through context-sensitive alternation rules. Scheer and Mathy, 2021), however, rightly point out that studies based on a single alternation from a single language are incapable of addressing either alternation or language specificity. They point out that if, for instance, a different processing routine underpins the /t/~∅ alternation in French floating-C – *il part* [paʁ] 'he leaves'   *nous part-ons* [paχt-ɔ̃] 'we leave' – then lessons learned from psycholinguistic investigations of liaison will not carry over to floating-C alternations. Caramazza et al., 2001, on the other hand, study cross-linguistic production of determiners in noun phrases and report that language-specific routines known from perception are also active in production as well. However, Scheer and Mathy, 2021 point out that while this is evidence for language-specificity, Caramazza et al., 2001's study do not speak to alternation specificity. A routine is alternation-specific if it is called upon only in response to processing demands stemming from specific alternations – e.g. if, for instance, routine X with a unique spatio-temporal signature is obtained only for /t/~∅ alternation in French, but not for liaison, then routine X is specific to this specific alternation. Likewise, if routine Y is obtained in, say, velar softening in both English and French, then one could reasonably argue that Y is specific to velar softening, possibly across languages. Crucially, studies on a single alternation cannot speak to alternation specificity, while studies conducted on just one language, even when they identify alternation-specific routines, cannot speak to whether the same routine is obtained for similar alternations in other languages.

Scheer and Mathy, 2021, thus, argue that phonologists are currently in the position of geologists who aim to make a theory of the characteristics of stone, but are unable to scientifically distinguish stone from plastic. Thus, they collect samples on which they build their theory, some of which contain 10% of plastic, others 30%, still others 60% and so on. Unsurprisingly enough, competing theories built on these wildly varying sets of empirical material then significantly diverge – not because of the theorizing itself but because of the plastic. The issue has been identifiable, at least, since the early 1970s (Scheer, 2010a), producing a large body of literature including, for instance, the debate surrounding evaluation metrics. Unfortunately, to date there is no pre-theoretical criterion that wold allow the theorist to decide whether a given alternation is stone or plastic, phonology-internal or phonology-external. Since the early 1980s, Scheer, 2010a points out, phonologists appear to have altogether abandoned the issue, and it has become increasingly clear that regular sources of evidence – corpora studies, behavioral analyses etc.– can do little to disambiguate different kinds of alternations. Recall, for instance, the Malayalam velar-palatalization study discussd in the previous chapter. It was argued that the phenomena is, indeed, phonological in nature (see discussion in the next section for some initial neurophysiological evidence by Scheer and colleagues), and consequently a sketch of a hybrid framework was presented arguing for a specific division of labor between phonological computations and phonetic pressures to account for the observed empirical data. But consider now, for the sake of argument, that this discussion was barking up the wrong tree, and velar palatalization turns out to be *not* phonological. In this case the mechanisms and principles attributed to phonology in the previous chapter becomes unnecessary and superfluous. In other words, the shape of phonological theory will be very different if we one assumes that a certain phenomena, in this case velar palatalization, does not fall within its purview. We need an objective, pre-theoretical tool to disambiguate what is, and what is *not*, phonology, and the next sections outline an ongoing research project that seeks to utilize correlational behavioral neuroscience to this end.

6.3.1 PhonolEEGical alternations

It is typical of EEG studies in phonology to target representational hypotheses involving the presence or absence of features, as discussed in section 7.2.2. This is partly due to methodological issues, since negativity mismatch responses are, both, well-studied in multiple cognitive domains and they follow a basic pattern of detecting deviant or novel occurences amidst habituated stimuli. The other avenue that usually attracts empirical attention from neuroscientists interested in linguistics concerns the broader grammar-internal sub-domains (syntax, morphology etc.) and the presence or absence of tissue-level neural specialization for the same (Moro et al., 2001; Mesgarani et al., 2014). Sahin et al., 2009 were among the first to look at these broad sub-domains from a temporal perspective, and explore whether lexical, morphosyntactic and phonological information are computed in parallel – as predicted by connectionist architectures (McClelland and Elman, 1986) – or whether these computations have a more sequential temporal distribution.

Sahin et al., 2009 recorded local field potentials (LFP) from English speakers through multi-contact depth electrodes while they performed a production task with three conditions or levels. The first level, called “Read”, required verbatim repetition of a cue word, either a singular noun or present tense verb, embedded in the carrier sentence “Repeat". The second level, called “Null-Inflect”, used the carrier sentences “Every day they _ ” (e.g. verb = read) or “That is the _” (e.g. noun = rock) which induced either the singular or present tense of the words. This condition requires inflection, but in English these inflections do not induce any phonological alternations. The third level, called “Over-Inflect”, used the carrier sentences “Yesterday they _.” and “Those are the _.”, and induced the past tense or plural forms which causes a change in the phonological shape of the target words. Sahin and colleagues argue that the three conditions require three distinct processing routines, each corresponding to a distinct level of linguistic computation. The first Read condition requires lexical access, or simple memory retrieval of a form. The second Null inflect condition requires memory retrieval of the root form and the null suffix, but given that the suffix is null the computations only involve concatenation without any overt phonological processing. Finally, the third Overt inflect condition requires the two prior mechanisms, but here the suffix is not null and thus triggers an additional phonological computation. Their results revealed that memeory retrieval produced a significant neural response (ERP, Event-Related Potential) at~200ms. The second condition, null-inflect, which required both memory retrieval and concatenation provoked ERPs at ~200ms and ~320ms. Finally, the overt-inflect condition, which requires retrieval, concatenation and finally phonological alternation, provoked ERP entrainment at ~200ms, ~320ms and finally at ~450ms. The table below provides a summary of the experimental conditions, and results.

Condition	**Memory Retrieval**	**Concatenation**	**Phonology**	**Cue**	**Stimulus**	**Target**	**Entrainment**
Read	Yes	No	No	Repeat	Rock	Rock	200ms
Null	Yes	Yes	No	That is the...	Rock	Rock	200ms, 320ms
Overt	Yes	Yes	Yes	Those are the...	Rock	Rock-[s]	200ms, 320ms, 450ms

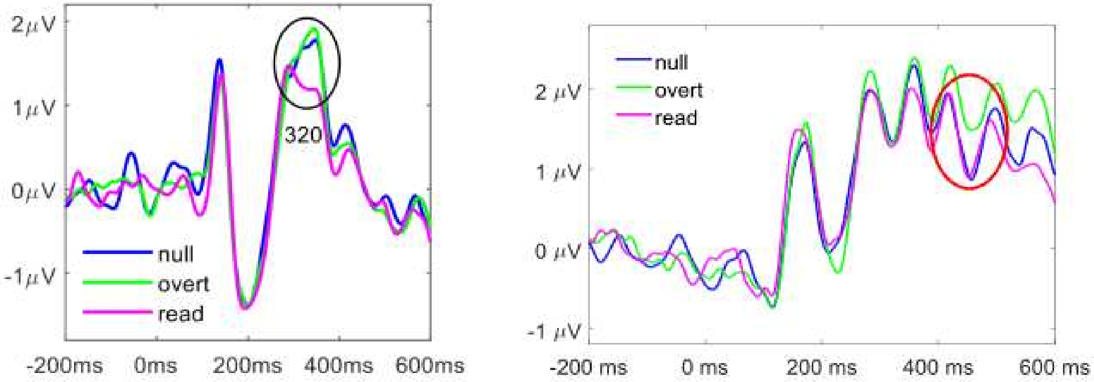
TABLE 6.1: Conditions in Sahin et al. (2009)

The results above speak to the sequential processing view held by most classical symbolic models of computation (Pylyshyn, 1980; Fodor, 1983; Fodor and Pylyshyn, 1988; Chomsky, 1995), evidenced by incremental increase in entrainment timestamps as computational load increases. Scheer and Mathy, 2021 point to these results and claim that the earliest timestamp being evoked by memory retrieval indicates that this condition has the lowest amount of processing to be done. Conversely, the overt inflect condition with its tri-level processing demands the most of processing routines, and consequently exhibits the highest latency. In their behavioral study Scheer and Mathy, 2021 obtain reaction time (RT) measures for three tasks that similarly demand increasingly more of processing routines. Their results, likewise, reflect that

RT is lowest when the processing involves a simple memory retrieval, higher when it involves memory retrieval and concatenation but no phonological computations, and highest when the former two is followed a phonological component.

Scheer and Mathy, 2021, thus, argue that the study by Sahin et al., 2009, while not originally intended for this purpose, nonetheless provides a tool to disambiguate when phonological computations are actually invoked by an alternation – namely, it should evoke an ERP at ~450ms. Based on this argument Scheer, Jonge, and Chabot, 2020 conducted an EEG study of English velar softening, and report that the well-known [k] ~[s] alternation in English forms, such as electri[k] ~electri[s]ity, involve proper phonological computations, with an expected ~450ms entrainment. The design is identical to the study by Ned Sahin and colleagues, but utilizes both actual and nonce words. In read condition participants read English words like 'electric' (real) and 'nectic' (nonce) presented in isolation. In the null-inflect the carrier sentence 'This is really _' induced an identical phonological shape, while the over-inflect condition used the carrier 'They talk about _' which required adding the suffix '-ity' and results in a change in phonological shape – electri[k]+ity = electri[s]ity, or necti[k]+ity = necti[s]ity.

Scheer, Jonge, and Chabot, 2020 made use of three sets of data with eighty words each. Each word was presented three times in each of three experimental conditions, yielding a total of 80x3x3, or 720, trials per participant. The results are discussed by Scheer, Jonge, and Chabot, 2020, and in more detail by Chabot, 2021, 280ff who provides the following entrainment graphs.



(a) At 320 ms Null and Overt can be distinguished from Read (the P5 electrode).

(b) At 450 ms, Overt can be distinguished from Read and Null (electrodes Pz, P2, P4 from the central parietal cluster).

FIGURE 6.1: ERP entrainment for English velar-softening results. From Chabot, 2021, p. 281

Sub-figure (a) above, on the left, highlights the entrainment patterns at ~320ms. Recall that this timestamp correlates with concatenation, and as such both null (blue) and overt (green) pattern together, to the exclusion of read (pink). This is because in the read condition the only processing demand is memory retrieval, or lexical access. On the other hand, sub-figure (b), on the right, shows the entrainment patterns at ~450ms timestamp. This particular window is associated uniquely with phonological alternations, and consequently the read (pink) and null (blue) conditions, which do not involve any phonology, pattern together against the overt (green) condition where phonological computations are triggered.

These initial results replicate Sahin et al., 2009's findings, with overt activity observed at 450ms for both real and nonce words. Scheer, Jonge, and Chabot, 2020 and Chabot, 2021 take this as preliminary evidence that velar softening involves phonological computations, and more importantly that EEG/ERP can adequately replicate LFP findings from depth electrodes. The inclusion of nonce words in this study was driven by the observation that these items lack lexical recordings, and thus cannot realistically involve allomorphy in any meaningful sense.

However, this pilot study still leaves aside several uncontrolled factors that potentially confounds any eventual discussion of alternation-specific neurophysiology. First, nonce words do not rule out the possibility that participants might be engaging in analogy – resolving ambiguous alternations by pattern-matching closely matching lexical forms (necti[k] : necti[s]ity :: electri[k] : electri[s]ity) instead of performing any phonological computations per se. Closely related to this is the issue of neurophysiological signatures of analogy itself. While Sahin et al., 2009 (unintentionally) identified signatures of suppletion, concatenation and phonology, to our knowledge no study exists that has tried to contrast analogy with phonology. Since we do not know what the neural signatures of analogy look like, i.e. what happens in the functioning brain when participants do not know what to do with a form, we cannot reliably claim that the ~450ms entrainment is definitively a marker of phonology qua phonology. The following sections describes a set of studies to further refine this protocol in order to reliably gauge the neurophysiological signatures of phonology-proper using ERP.

6.3.2 Mind 2 Brain: disambiguating phonology from analogy

The primary motif behind the Mind 2 Brain project is to isolate the neurophysiological signatures associated with proper phonological computations. Ideally, if this project succeeds in its main goal, we would have constructed a neurocomputational tool that can be reliably re-used to immediately test whether any given contentious alternation (recall, for instance, the Malayalam velar palatalization phenomena discussed in chapter 6) actually invokes the neural routines necessary phonological processing. If it does, then it can be claimed to be within the purview of phonological theory, and consequently phonological theory must be properly equipped to deal with such alternations across languages. Of course, this is a rather lofty goal and one that cannot be realistically achieved through just one empirical study. For one, as mentioned previously, it is not just a matter of separating phonology from morphosyntactic concatenation (or allomorphy) and lexical access, but one also needs to be able to reliably identify the neurophysiological signatures of analogy. Furthermore, given the primary focus of SFP, one also needs to be wary of situations where alternations are attributed to phonology even though their causal roots lie in functional phonetic factors (Hale and Reiss, 2008; Odden, 2022; Chabot, 2021). As part of this larger project, Mind 2 Brain is currently implementing an experimental procedure designed to isolate the neurophysiological signatures of analogy, i.e. situations where feature-related changes are observed in UR ~SR alternations, but the subjects can be realiably claimed to lack any knowledge of what process to call upon. This first stage is described in detail below. Then, in the final section, I sketch out a preliminary outline of extending our study to disambiguating phonology (and analogy) from phonetics.

The target language for this first set of studies is French, and in particular we make use the process of pluralization in French. Forming of plurals in French can take one of two forms. First, plurals can be formed by attaching a silent -s] suffix (i.e. while orthographically this requires adding a -s] suffix, there is no phonetic realization of the same) to the base form – e.g. sac "bag" ~ sacs [sæk] "bags". These forms constitute the regular pluralisation in French. However, there is yet another, albeit less proliferate, way to pluralize in French. This involves the addition of the -aux] suffix to the base form – e.g. tribunal "tribunal" ~ tribunaux "tribunals". While usually native speakers have a pretty clear knowledge of when each suffix is required, French also contains a number of lexical entries where the plural form is ambiguous. Some of these forms, for instance *corporal*, are ambiguous because they are no longer widely utilized in everyday French vocabulary. In personal communications, Tobias Scheer points out that the world *corporal*, a reference to a specific military rank that is no longer actively utilized, only occurs in historical texts. As linguists we are not so much interested in the philology of a language as we are in the synchronic knowledge of modern day French speakers. If this word, and its referent, are no longer of synchronic significance, then it is to be expected that the only way to be familiar with this form, and its plural counterpart, is through academic or intellectual exercise. From the pedestrial native speakers' point of view however, the bread and butter of a working (generative) linguistic, these forms are ambiguous. The speaker has no knowledge of how to pluralize these forms, and will arbitrarily choose from between the regular -s] or the alternative -aux]. This constitutes our target scenario, or *analogy*.

While ambiguous words like *corporal* or *terminal* are good candidates for targeting analogy as a phenomena, they are not ideal candidates given the possibility that the participants might always end up choosing the regular suffix -s]. Recall that this suffix is silent and does not induce any pronunciation alternation in the surface form. A preliminary survey of speaker behavior using native speakers from Quebec confirmed this suspicion. In this survey thirty native speakers were given a list of nineteen words to pluralize, a mixture of real unambiguous, real ambiguous and nonce words. In real, but ambiguous, words (e.g. terminal) 55% of the responses collected were of the -aux] type, i.e. exhibiting analogy, while 45% involved the regular -s] ending. However, for the nonce words an overwhelming 76% of the responses collected exhibited analogy, or the -aux] ending. Given this trend, and the problems stemming from unusable data points (i.e. ambiguous words where participants opt to use -s] suffix), we opted to utilize only nonce words in order to elicit analogy from the participants.

Stimuli

The stimuli for this experiment will consist of a mixture thirty-six real and nonce French words – twelve real words that accept the regular -s] plural, twelve real words that accept the -aux] plural, twelve nonces that lack any lexical entries and can only be pluralized by analogy. Here we encounter a problem with the lexical stock of French words, namely the number of real words that accept -aux] plural is limited. Further, a number of words that do accept -aux] have ambiguous usage, in that they can be used either as a noun or as an adjective – e.g. musical. Given our experimental procedure that relies on eliciting singular and plural forms (see next section) these words cannot be reliably utilized as experimental items. Further restrictions are imposed by the fact that we opted to use only words of two-syllables/three-syllable lengths, and of these only those with high frequency in order to ensure that the participants have properly lexicalized them. These left us with a choice of only twelve -aux] type words, listed in the table below.

The total number of stimuli per run, across three word-types (-s], -aux], nonce), is thirty-six. In each run each word is going to be repeated three times, totalling one hundred and eight items per run. Each participant will undergo five runs in total, thus being exposed to a total of five hundred and forty stimuli items during the experiment.

Type: -aux]	**Frequency (Film; Out of Million)**	**Syllable Num.**
hôpital	133.15	3
tribunal	37.98	3
maréchal	2.84	3
arsenal	2.36	3
vassal	2.24	2
bocal	3.83	2
cheval	129.12	2
journal	110.8	2
cristal	10.22	2
canal	17.11	2
signal	39.61	2
métal	14.36	2

TABLE 6.2: List of -aux] type words

Procedure

EEG data collection is done through a sixty-four channel Compumedics Neuroscan device, located at the Bases, Corpus and Language Laboratory of the University of Côte d'Azur. Each participant will undergo five runs of EEG scan, and perform a silent articulation task under three distinct experimental conditions. Scanning will utilize 64 passive Ag/AgCl (Silver/Silver Chloride) electrodes, and EEG signal will be recorded at a 1000Hz sampling rate, with $< 5k\Omega$ impedance.

Our experimental procedure is derived largely from the protocol utilized by Sahin et al., 2009, and implemented in a pilot study previously by Scheer, Jonge, and Chabot, 2020. The experiment will be carried out with tri-level design – READ, NULL and OVERT. In each condition the participant will be exposed to a carrier sentence/instruction for 650ms, followed by a fixation cross for 1100ms. Following this a stimulus will be presented on-screen for a duration of 1250ms. Once the stimulus has been presented, the participants will see a second fixation cross for 1500ms. The duration of the carrier sentence/instruction plus the first fixation cross is termed the "cue epoch", while the duration of the stimulus presentation plus the second fixation cross is termed "response epoch". The participants will, thus, have a total of 2750ms to silently repeat the appropriate word form. The figure below provides a diagrammatic overview of the process.

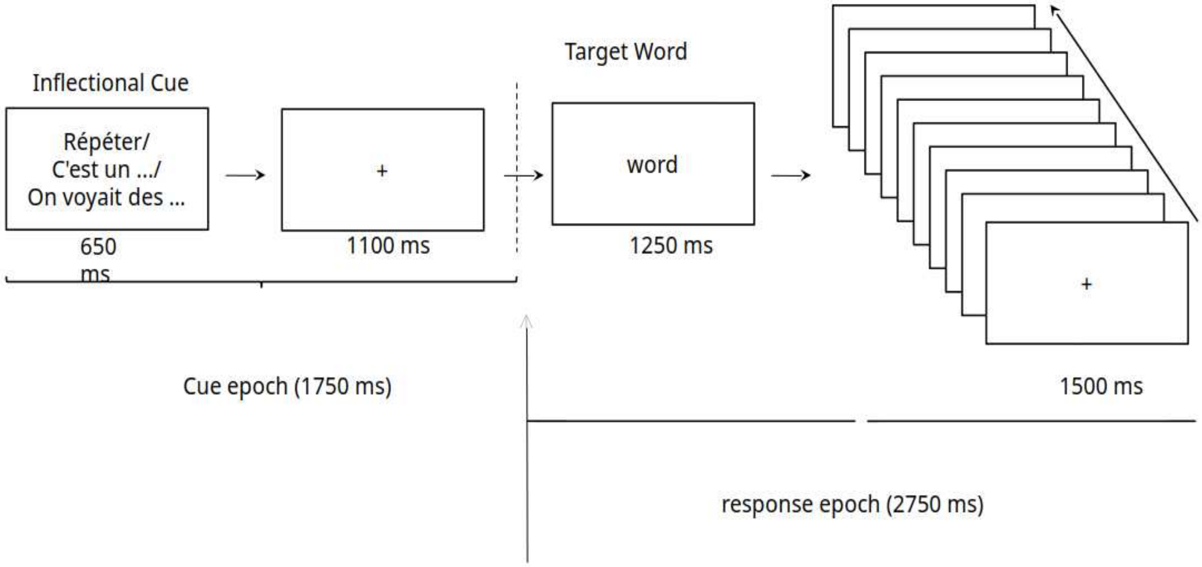


FIGURE 6.2: ERP experiment timeline for Mind 2 Brain

In the READ condition the participants are simply required to read the word presented on screen, and the instruction is *Répéter*. In the NULL condition the carrier sentence, *C'est un ...*, induces the singular form of the presented stimulus, while in the OVERT condition *On voyait des ...* induces the plurals.

Analyses

Analyses of EEG data will be performed using the Matlab package for EEG analyses, EEGLab. Response epochs, time-locked to the stimulus word onset, will be extracted and average ERP components will be calculated for each condition. The original signal, recorded at a 1000Hz sampling rate, will be resampled to 500Hz, and 0.1 to 50HZ Finite Impulse Response (FIR) filter will be applied. As part of the independent components analysis (ICA), we also plan to apply a 0.1-50Hz PICARD ICA algorithm for independent component decomposition. This separates the EEG signal into statistically independent components, which can be further analyzed to identify and remove artifacts from the data.

Expected outcomes

It is crucial to note that at this point Mind 2 Brain is a purely exploratory study. This is because, first, and to our knowledge, no one has so far attempted to isolate the neurophysiological signatures of analogy. Second, while our study is based on a protocol derived from Sahin et al., 2009 it must be noted that the original study was conducted on only three participants while they underwent cranial surgery. It is, therefore, not quite possible to predict with certainty what they outcomes will be, and surprises lark around every corner. But based on prior studies conducted within this program by Scheer, Jonge, and Chabot, 2020 and Scheer and Mathy, 2021 we have some (linguistic) analytical options available to us, and in this section I try to align them with anticipated neurophysiological evidence.

Our primary interest lies in recording the neurophysiological signatures of analogy – i.e. the nonce words in our stimuli. For the nonce trials we will take into account only those responses where the subjects responded with an -aux] plural. That is for, let's say farédal, we will only consider trials where the participant respond with farédaux (these, we refer to below as Nonce-aux). This is because the only way to produce farédaux is through analogy. We aim to make four distinct levels of comparison. First, and perhaps the most important comparison, is between Nonce-aux in the OVERT condition versus the NULL condition. In the NULL condition there is no analogy involved, but in OVERT analogy is the only available analytical option. The contrast, therefore, should provide us with the desired signature for analogy.

The next comparison of interest is between Nonce-aux and regular -s] across all three conditions. For the READ condition we expect them to be identical, and produce a ~200ms ERP. In NULL we expect them to be identical again – the singular forms are inserted, generating a ~200ms ERP for retrieval, but no further computations are performed. In OVERT we do not expect to see ~450ms entrainment for either word types. Regular -s] does not involve any phonological alternation, neither does Nonce-aux. However, we do expect Nonce-aux to trigger some ERP that differentiates it from regular -s], thus instantiating analogy.

The third set of comparisons involve the Nonce-aux versus the regular -aux] words. In READ both are identical, with ~200ms ERPs. In NULL, again, both should only trigger the aforementioned ERP for retrieval as no further computations are involved. In the OVERT condition, regular -aux] should trigger ~320ms ERP for allomorphy, but no ~450ms ERP as no phonological computations are involved. For Nonce-aux, again, we expect it to be different from regular -aux], with the difference in the plots yielding the difference between root-allomorphy and analogy.

Finally, we aim to compare regular -aux] with regular -s]. In READ condition both are expected to be identical, with a ~200ms ERP. In NULL, again, both only trigger the ERP for retrieval, but no further computations are triggered. In the OVERT condition we expect them to be identical as well. Regular -aux] triggers root-allomorphy while regular -s] triggers concatenation, both producing ~320ms ERPs.

To summarize, we expect to isolate at least three different ERP plots for French pluralization. First, for regular -s] words we expect ~320ms ERP for concatenation. Next, for the singular (NULL) versus plural (OVERT) comparison in regular -aux] words we expect ~320ms ERP for root-allomorphy (similar to concatenation in regular -s]). Finally, for Nonce-aux we expect to isolate an yet unknown ERP, and one that instantiates the neurophysiological signatures for analogy. The table below summarizes the expected ERPs for Nonce-aux across all three experimental conditions.

Condition	**Memory Retrieval**	**Concatenation**	**Phonology**	**Cue**	**Stimulus**	**Target**	**Entrainment**
Read	Yes	No	No	Répéter	farédal	farédal	200ms
Null	Yes	Yes	No	C'est le...	farédal	farédal	200ms
Overt	Yes	Yes	Yes	Ce sont les...	farédal	faréd-aux	200ms, ???

TABLE 6.3: Conditions in Scheer et al. forthcoming

6.3.3 What phonology is not

This section provides an initial sketch of where the Mind 2 Brain project expects to go in the near future. As a first point of departure, note that none of the alternations we have proposed to test thus far involve phonology in any sense. This issue is relevant because so far we have been assuming, based on Sahin et al., 2009 and Scheer, Jonge, and Chabot, 2020, that a ~450ms ERP is indicative of phonological computations. However, recall that a large number of phonological processes hew rather close to the phonetic bone. However, the SFP view, adopted throughout this work, is that such closeness is incidental and does not reflect any ontological phonetic grounding of phonology. Thus, continuing on with a line of discussion so far, in order to ascertain whether ~450ms ERP is really a marker of phonological activity we need to ascertain that the same is not associated with phonetic processing. Conversely, if distinct ERPs are obtained for phonological and phonetic alternations it can be reasonably argued that the two linguistic modules exhibit distinct types of neural computations, and are neurobiologically distinct. In order to probe such hypotheses it is necessary to investigate alternations of different types, in this case phonological and phonetic, and do so across more than one language.

In both Hindi and French underlying sequences of a nasalized vowel and voiced obstruent (/ṼC/) surface with a pre-nasalized obstruent – i.e. a short nasal consonant with identical place of articulation with the following obstruent is inserted between the vowel and the obstruent ($\tilde{V}^{N}$C]).

Ex 6.3.1. Epenthetic nasals in Hindi and French.

Hindi

/jəhãd̪ekʰo/ → [jãnd̪ekʰo] ’look here’

French

/sã bɛl/ →[sãᵐbɛl] (in ‘dit saint bel enfant’)

Ohala and Ohala, 1975 report that these epenthetic nasals, in both languages, display properties of phonetic co-articulatory effects – they are shorter in duration, between 60ms to 70ms, than nasals that are phonologically distinctive (~100ms), and exhibit reduced amplitude peaks. If we accept Ohala and Ohala, 1975’s arguments, then we have two instances of UR ~SR alternations that are purported to be entirely phonetic.

Hindi, likewise, also exhibits a process of schwa deletion; the schwa is deleted between VC and CV sequences, as long as the first C satisfies certain phonological conditions (denoted by the superscript numerals): ə→ ∅ /VC^{1,2} _CV (Pandey, 1990).

Ex 6.3.2. Hindi schwa deletion.

məcəl ’prance’ → məcla: ’pranced’

lətək ’hang’ → lətka: ’hanged’

pəkər ’catch’ → pəkra: ’caught’

What is important for our purpose, however, is that this process is unambiguously claimed to be phonological within the theory. A counterpart in French involves the /t/~∅ alternation in French verbal paradigm, called floating-C (Scheer and Mathy, 2021).

Ex 6.3.3. French floating-C

[gʁã] ’big, masculine’ ~[gʁãd] ’big, feminine’

[il paʁ] ’he leaves’ ~[paχt]ons ’we leave’

Since similar alternation types are attested in two unrelated languages, summarized in table 7.4 below, this offers us a possibility to comparatively probe the ERP signatures of phonetic versus phonological alternations.

	Hindi	**French**
Phonological	Schwa Deletion	Floating-C
Phonetic	Epenthetic Nasals	Epenthetic Nasals

TABLE 6.4: Phonetics and phonology in Hindi and French

Experimental procedure

Procedurally, such experimentation would involve obtaining the ERP signatures bound to each type of alternation, in each language, while participants perform tasks that induce the relevant alternations naturally. I propose two sets of experiments, one each adapted to the phonological and phonetic processes respectively. The phonological schwa deletion (Hindi) and Floating-C (French) alternations involve past~present and singular ~plural alternations respectively. These forms can be induced by appropriate carrier phrases. For instance, the carrier '/kal woh/ _ (Yesterday they _)' will induce the form [məcla:] 'pranced', while '/a:j ham/ _ (Today we _)' induces [məcəl] 'prance'. Likewise, for French the carrier 'Ils sont _' induces *grands (e.g. Ils sont grands. 'They (masc.) are big'/*, while 'Elle est _' induces *grande*.

The task conditions remain identical to our current study – READ, NULL and OVERT. Three task-conditions will be necessary in order to obtain the three sequential entrainments reported by earlier studies. The first READ condition will present a word to the participants and require them to read it out loud. This process involves a simple memory retrieval, or a 1° computation with a ~200ms ERP. The second NULL condition, induced by the present tense (Hindi) and mascular (French) carriers requires the participants to retrieve the proper words, and then apply appropriate grammatical inflections, which in both cases do not cause any overt change in pronunciation. The computations are, thus, 2° involving retrieval and (null) inflection. However, in the 3° Overt Inflect condition, things are as before except that the grammatical marker now is overt, that is, pronounced – In French, grande = [gʁãd] with a final [-d] in the feminine form, while in Hindi the induced past-tense form requires schwa deletion in the transition from məcəl ′prance′ to məcla: ′pranced .Here the computations involve retrieval, then inflection followed by sound alternation, and are thus 3° with an additional ERP expected at ~450ms.

Next, consider the epenthetic nasals in French and Hindi, our purportedly purely phonetic alternations. The task conditions here are the same as before, however in order to ensure that the OVERT condition produces actually phonetic epenthesis we need to ensure that forms speakers do not contain any lexicalized effects. In this phase of the study, the READ condition will present the subjects with tokens of French and Hindi words, with instructions to just read them silently. However, for the NULL and OVERT conditions they will be required to perform a word-blending task. In psycholinguistic studies a word-blending task involves recognizing two words monosyllabic (preferrably CVC) words, followed by creating a nonce word by combining the first syllable of the first word with the final consonant of the second word. The relevance of this task for the proposed study lies in the fact that the resulting nonces lack any lexical entries, and thus any epenthetic nasals resulting from phonetics is much more likely to be purely co-articulatory. In order to contrast NULL with OVERT, however, we also need to ensure that any epenthesis only occurs in the OVERT condition. I will use Hindi to make this process clear. In the NULL condition subjects will be exposed to two orthographic representations of Hindi words with a fixation cross in between, as below. The first word contains a non-nasalized vowel, and the second ends in a voiceless stop. The speakers then combine them to create a nonce word where the first syllable has a non-nasalized vowel, and is followed by a voiceless stop. Per Ohala and Ohala, 1975 this mitigates any possibilities of an epenthetic nasal. In the OVERT condition, the first word contains a nasalized vowel, and the second ends in a voiced stop. The resulting nonce should produce epenthetic nasals here.

Condition	**Word 1**	**Fix**	**Word 2**	**Blend**
Null	sat̪ ′seven′	+	ruk ′stop′	sak
Overt	pãc ′five′	+	sud̪ ′interest′	pãⁿd̪

TABLE 6.5: Proposed word-blending task

Our primary interest in including this second phase in the Mind 2 Brain project is to reliably tease apart phonological alternations from all sources of ambiguity, namely analogy on the one hand and phonetics on the other. In the READ conditions here we expect identical ERPs with the analogy study, as subjects merely retrieve the word from the lexicon. For both phonological and phonetic alternations, the NULL condition is expected to trigger 2° computations. In the phonological alternation, in French this involves retrieval and grammatical inflection, but no phonological or phonetic change is observed. In Hindi, the NULL condition requires clear concatenation of the first syllable of the first word with the final consonant of the second. The interesting case, however, are the OVERT conditions in both types of alternations. In this condition, the French phonological task the induced feminine form requires the insertion of the final [-d], which requires phonological computations in addition to retrieval and concatenation. Likewise, in Hindi the schwa deletion requires a phonological alternation. The distinction between French and Hindi here should come from the absence of ERP for the concatenation process. In other words, for French we expect three ERPs – for retrieval at ~200ms, for concatenation at ~320ms and a third ERP following this for phonological condition. This third ERP could be the ~450ms stamp, if this is indeed associated with phonology proper. For Hindi, however, we only expect the first and third ERP, because schwa deletion does not involve any concatenation or allomorphy.

The crucial point of interest lies in the OVERT condition in the phonetic alternations in both Hindi and French. This is where we should be able to reliably decide on (a) whether ~450ms

ERP is indicative of phonological or phonetic processing, and (b) if it is phonological, then what the neurophysiological signatures of phonetics are. For both Hindi and French, then we expect three consecutive ERPs – a ~200ms ERP as the real words are read, a ~320ms ERP as concatenation is performed to create the nonce, and a third ERP that is indicative of phonetic nasal epenthesis. The comparison of this OVERT condition across both alternation types (phonological and phonetic) across Hindi and French should reliably yield the required ERP plots to tease apart phonology from phonetics, while comparisons with the previous stages of the study should allow us to separate both from analogy.

6.4 Final remarks

This final chapter has reported on an ongoing project that seeks to isolate phonology, and phonological alternations, from other types of grammar-internal and external processes using neurophysiological signatures. Parts of the proposed work is already underway, and parts are in development. Our objective is to reliably establish the central claim of substance-free phonology – that phonology is a module of its own, whose internal computations are distinct from other grammatical sub-modules and not motivated by functional substantive factors – and do so with neurophysiological evidence. As I have discussed, there exist good reason to believe that the Mind 2 Brain is on the right track. Our anticipated results are crucial for a number of reasons. First, from a purely theoretical perspective, they stand to contribute significantly to theory-pruning. The stated goal of the minimalist program in linguistic theory (Chomsky, 2014) is to set the boundary between UG-internal and external as low as possible, attaining a maximally sparse grammar. To this end, our anticipated results, and indeed the experimental apparatus that the Mind 2 Brain has created, stands to contribute a realist tool to determine just how much, or how little, should be attributed to phonological theory itself. The first step to determining this is being able to delineate those grammatical processes that fall within the purview of phonological theory from those that do not. In terms of an analogy created by

Tobias Scheer, our tool is intended to let phonologists reliably identify phonological stones from plastic. A realist scientific theory phonology should not be built around personal preferences for what the individual theorist considers phonological, nor what they might want phonological theory to look like. Rather, phonological theory should be treated as a proper sub-theory of the cognitive neurosciences, making testable predictions about real neurobiological events and procedures.

Likewise, the burgeoning biolinguistic enterprise seeks to firmly situate linguistic, and thus phonological, theory at the intersection of neurobiology and theories of computation(Chomsky, 2017; Chomsky, Gallego, and Ott, 2019; Samuels, 2015a). In chapters 4 and 6, while proposing a hybrid framework for the lower-interface with phonetics, I have reviewed and stressed the importance of neurophysiology for phonological theory. There is overwhelming evidence to suggest that phonological primes instantiated at the intersection of stable activation of neuronal maps and neural oscillations at the theta and gamma bands entraining with their activations (Poeppel, Idsardi, and Van Wassenhove, 2008; Poeppel and Idsardi, 2011; Hickok, 2022). The Mind 2 Brain further seeks to extend the importance of neurophysiological entrainments by attempting to use ERPs to reliably distinguish between analogy, phonology and phonetics. Scheer, Jonge, and Chabot, 2020 have already provided initial evidence that this hypothesis is on the right track, and we are currently well on our way to detect at least one phonology-external process in the brain – analogy. A successful three-way contrast between analogy, phonology and phonetics will complement prior neurobiological findings about phonological primes with evidence that, much like primes, phonological computations are also instantiated by real neurophysiological phenomena implemented by real neuronal wetware. Crucially, our project operates on the hypothesis that neurobiological wetware – such as neurons, or functionally connected activation of neuronal maps – and neurophysiological events – such as ERPs, or oscillation entrainments – themselves are neither features nor computations. Rather, the abstract and implicit knowledge endowed by UG is utilized by neural wetware and neurophysiology to resynthesize experience in a manner that both instantiates specific states of I-language in speakers' minds from minimal and impoverished experience, while simultaneously enabling generative productivity. Phonology is not reducible to neural cells or cortical regions, it is rather the ghost in the cells. This view lies at the heart of correlational behavioral neuroscience (Krakauer et al., 2017), and our project is squarely aimed utilizing this approach to identify distinct computations performed by two distinct modules of the faculty of language – phonology and phonetics.

Finally, the explicit attempt throughout this chapter, and indeed this work as a whole, has been to identify a proper division of labor between the faculty of language and the surrounding cognitive modules in whose context the faculty exists. Chomsky, 2005 has been explicit in his assertion that UG does not exist in vacuum, and indeed in recent writings has increasingly stressed the importance of deriving as much of language as possible from experience and third-factors (Watumull et al., 2014; Weerathunge et al., 2022; Roberts, Watumull, and Chomsky, 2023). I have attempted to apply this line of thinking throughout this work – for instance, in chapter 3 I have tried to reduce the complexity of phonological representations to flat strings and bring it closer to animal cognition, and in chapters 4 and 6 I have argued for a view of features and feature-interpretation that by and large make use of shared cognitive mechanisms that are not language-specific – and this final chapter has sought to extend this minimalist approach to a neurophysiological disambiguation of alternations, and a delineation of phonology from all that phonology is not.

6.5 Special acknowledgment

The Mind 2 Brain project is the brainchild of my supervisor, Prof. Tobias Scheer, and originated in its initial form as an attempt to determine whether phonological computations are indeed involved in velar palatalization or softening. The initial study was carried out in conjunction with Dr. Alexander Marx Chabot (University of Maryland, USA), and provided preliminary evidence for velar softening being possibly phonological kind. I am grateful to Prof. Scheer for the many hours of discussion, in person and over countless emails exchanged, that he has indebted me with. I am equally grateful to him for allowing me to join his project when my original fMRI study had to aborted due to, both, the COVID-19 pandemic and lack of resources at my home university. Without his help, I would not have a dissertation to write.

I am equally grateful to Dr. Chabot, to me *Alex*, whose penetrating insights and historical knowledge of phonological theory has been a source of both enlightenment and inspiration to me. Almost every chapter in this dissertation is partly or fully inspired by his works, but most of all this specific chapter.

Finally, my continuing participation in this project is enabled by a MITACS grant that has allowed me to visit France, and work on this project at the wonderful Bases, Corpus and Langage Laboratory at Université Côte d’Azur in Nice. I am grateful to my colleagues at the BCL Laboratory who have made my stay here a pleasant and fruitful experience. Particularly, I would like to thank Prof. Fabian Mathy and Dr. Seckin Arslan for providing me assistance with experimental design protocols, and Christophe Zimmer for the many hours he has spent helping me troubleshoot our EEG setup.

Chapter 7

Concluding Remarks

7.1 Ghost in the Cell

The title of this dissertation is a play on, both a phrase attributed to the philosopher of mind Gilbert Ryle – *Ghost in the Machine* – and a critically acclaimed Japanese manga by Masamune Shirow – *Ghost in the Shell*. It is, perhaps, then befitting to provide some elaboration on the analogues I had in mind. I will start with Gilber Ryle's phrase. Ryle coined this phrase in his 1949 book *The Concept of Mind* (Ryle and Tanney, [1949]2009) to critique the Cartesian dualist notion that the mind is a non-physical entity distinct from the body. Ryle argued that this idea mistakenly treats the mind as a separate "ghost" inhabiting a "machine" (the body, or more specifically the brain). He viewed the dualist perspective as a category mistake, meaning it incorrectly categorizes the mind as an independent substance rather than understanding mental processes as part of the physical workings of the body. To what extent Ryle had in mind what has, in recent years, come to be known as *embodied cognition* is a matter of some debate (Aizawa, 2015; Aronoff et al., 2021), and not one that I have attempted to settle here. But it is worth noting that a rejection of Cartesian dualism does not entail an adoption of embodiment, but rather the acknowledgment that while the mind is instantiated by the brain the computational properties of mental phenomena, such as Natural Language grammar, cannot be directly reduced to properties of neural cells. This position is perhaps most eloquently advanced by Noam Chomsky in

Cartesian Linguistics (Chomsky, [1966] 2009). In this seminal work on the history of rationalist epistemology, Chomsky reinterprets the term "Cartesian" to move beyond the strict separation of mind and body. Rather, Chomsky emphasizes the creative aspects of human language and cognition, arguing that crucial aspects of cognition, such as the capacity for language, are due to structural constraints imposed by the evolutionary history of the brain. This view offers a way out of Cartesian dualism by positing that though the mind's capabilities are not entirely separable from the body, they are nevertheless causally explainable only by referring to abstract computational rules that are not directly reducible to neural cells (see also Poeppel, 2012). Crucially, Chomsky is explicit about the fact that it is, in fact, the brain that instantiates the mind and implements its abstract computations. But Chomsky nonetheless insists on a separation of the wetware and the software that runs on it. A very similar idea runs through Shirow's manga (Shirow, 1995). Set in a cyberpunk future where cybernetic enhancements are common, the story follows Major Motoko Kusanagi, a highly skilled cyborg and leader of Public Security Section 9, an elite counter-terrorism unit. The plot revolves around Major Kusanagi and her team's pursuit of a mysterious hacker known as the Puppet Master, who is capable of "ghost-hacking" into the cyber-brains of cyborgs to control them. The title Ghost in the Shell reflects the core themes of the film, focusing on the philosophical exploration of identity, consciousness, and the nature of consciousness in a world where the human mind can exist within a mechanical body. The "ghost" represents the human spirit or consciousness, while the "shell" represents the artificial body. It is not hard to see the same wetware/hardware $\sim$ software dichotomy that Chomsky has in mind in Shirow's works.

In the context of phonological theory my intent in alluding to Shirow and Ryle has been simply to emphasize a similar dichotomy between primes of phonological computation (software) and the neural events and structures (hardware/wetware) that instantiate them. Phonological features are much like "ghosts," are abstract entities that govern the behavior of the phonological wetware in the brain. Generative phonology posits that phonological features are the fundamental units that define the distinct linguistically relevant properties of speech sounds. These features are not directly observable in the acoustic signal; rather, they are abstract and underlying, and must be distilled by featurally driven computations implemented by neural cells. In this framework, phonological features determine the way sounds interact and change in different phonological environments, akin to how Shirow's "ghost" might affect the operations of the "shell". Just as the ghost is not part of the shell's physical structure, phonological features are not directly tied to the phonetic form but exist as an underlying level of representation that guides the production and perception of speech. Phonological theory is a theory of the abstract, of the "ghosts" in neural cells – it is a theory of representation and of computations, and not of behavior. It is a theory of *competence*, in Chomskyan terminology, and not of "performance". A typical problem in the study of Natural Language based on the only observable external data that it generates, speech forms, is that the data is typically noisy and impoverished, and being many layers removed from the real object of study typically tends to underdetermine the theory. As much as my intent in this *ghost in the cell* approach has been to be mindful of this problem, here at the end it occurs to me that issues perhaps remain. What follows, then, is a brief attempt to summarize some thoughts that have occured to me on what has been presented so far, and how I now feel in retrospect about what I have worked on over the past five years.

7.2 Competence, performance, and relating verbal behavior to phonological features

The first issue I would like to revisit here concerns the dichotic listening study presented in Chapter 4. The reader might recall that based on a pattern of perceptual intrusions observed in adult Malayalam-Australian English bilinguals I argued for a privative feature theory. This argument was driven primarily by the observation that a purely substantive account, one based on the harmonics-to-noise ratio (HNR) of the stimuli used proved insufficient in accounting for the observed asymmetries. Particularly, we reported the unattended voiced segments ([+Voice]) were more likely to intrude into the attended voiceless segments ([-Voice]) than vice versa. Given that a purely substantive account was not feasible, the reasoning presented was that a privative feature system (i.e. [Voice] vs. [ ]) appeared to be necessary. In such a system a voiced segment is literally "marked" by the presence of the [Voice] feature, while the voiceless counterparts simply lack this laryngeal feature. It was, therefore, argued that a signal containing more detectable features is harder to ignore, implying that perceptual salience is likely to be correlated with the sheer number of detectable features in the signal.

In retrospect, however, it occurs to me that perhaps this conclusion is reached a little to hastily. There exist compelling formal reasons to assume a privative feature system. Odden, 2022, for instance, points out that a privative feature theory is formally simpler. In a binary system a representation of the form, say, +Voice, includes three primitive notions – a value (+ or -), an attribute (Voice) and a feature as conjunction of the former two (+Voice or -Voice). Further, we also need a postulation that forbids attribute-free, or floating, values. However, one might also wonder, given a privative feature theory, whether there exist some universal argument for the concerned feature to be [Voice], rather than [Voiceless]? Given the argument forwarded in chapter 4 this is a non-trivial issue. If feature theory is indeed to be privative, but with the privative feature for voicing distinction being [Voiceless], the proposal advanced in this chapter no longer works. Indeed, one would expect exactly the opposite pattern of intrusions to occur. I must admit that at the time of writing the concerned chapter this was not an argument I had considered. As I went looking through the literature for some possible solution to this conundrum a tempting urge was to refer to Odden, 2022, p. 506 who points out that

> [...] nothing [...] precludes having [Voice] and [Voiceless] coexist in a language. When voiceless segments act as a class under a rule, that behavior motivates the existence of a feature [Voiceless]. Nothing prevents a language from having a fact pattern motivating a feature [Voice] as well as a fact pattern motivating a feature [Voiceless].

Notice, however, that Odden's arguments crucially concern *phonological rules*, not perceptual patterns. For Odden's arguments to save the proposal advanced in chapter 4 there would need to be independent arguments, from both English and Malayalam, to assume that [Voice] is, indeed, the necessary feature in both languages. To what extent the assumption that the presence of additional features makes a signal more perceptually salient is an admissible phonological observation is unclear to me, though studies in the neural responses to vowel undespecification by Lahiri and colleagues (Lahiri and Reetz, 2002; Monahan, Pérez, and Schertz, n.d.) appear to suggest that this might not be too wild of an idea. Either way, I think there is a lesson to be learnt here, and one that is almost a twice-told tale – competence and performance are layers of abstraction apart, and how one reasons about the former based on the latter must necessarily be a function of one's theory of competence and not one's corpus of performative data.

7.3 On the ghosts

Next, I want to briefly revisit the Malayalam velar palatalization data presented in chapter 5, because it relates to both phonological theory and its relation to its neurobiological implementation, particularly the interface with the sensory-motor systems. The reader might recall that in Malayalam velar consonants undergo palatalization – a process by which they surface with a secondary palatalized articulation – when preceded by the vowel set /i,e,a/. Based on empirical data drawn from a set of behavioural experiments it was argued that the process is phonologically best explained as the spreading of the Designated Articulator node from the preceding vowel (Halle, 2002; Halle, 2005). I would like to take this opportunity to provide some further clarifications here. First, my primary motivation behind adopting the DAX features proposed by Halle, 2005 is related primarily to the fact that I conceive of features as providing dimensions of contrast along which phonetic categorization occurs over the critical period. These dimensions, that is the set of possible features, I have argued in chapters 3 and 5, are a priori limited by the properties of the articulatory channel. However, it is important to acknowledge that this is by no means the only possible syntax of phonological rules that is capable of accounting for the Malayalam process. An alternative account is made possible by referring to the framework of Logical Phonology (Bale and Reiss, 2018).

In Logical Phonology phonological rules are conceived of as set-theoretic operations. One consequence of this is that all rules must be stated in terms of natural classes, defined as generalized intersection of segments which are themselves defines as sets of sets of valued features. Logical Phonology also assumes a deterministic relationship between phonological primes and their phonetic correlates. Avoiding this notion was also my primary motivation behind not appealing to Logical Phonology. In this framework, the Malayalam vowel /a/ must be specified with +Back because it bears all the phonetic qualities of back vowels. A rule for palatalization stated in terms of front vowels, that is the class of [-Back] vowels, will thus fail to trigger palatalization in the context of [a]. However, note that /i,e,a/ still share one feature in common; namely they are the class of [-Round] vowels. Given that Malayalam has a five vowel inventory – /i,e,a,o,u/ – [-Round] effectively groups the three triggering vowels together, to the exclusion of /o,u/. We can, thus, readily formulate the Malayalam palatalization rule as:

Ex 7.3.1. [-Syllabic, +Back] → [-Back] / [-Round]__

While this rule correctly derives all the necessary surface forms, the theory of phonological representations it necessitates is generally incompatible with the theory of the lower interface I have adopted here. Any theory of phonological rules and representations must interface with a theory of transduction to substantive information. This theory, likewise, must be able to account for, in biolinguistically plausible terms, how abstract features are related to physical substance. Given that this has been my focus throughout the dissertation, I will briefly turn to the issue of transduction in the next section.

7.4 On the cells

The primary concern of this dissertation has been the nature of the lower interface of phonology, the one with the sensory-motor systems, and the one that links the abstract with the physical. There are two intricately related components to this story, as I see it. First, the story of the ghost in the cell crucially hinges on a realist conception of the "ghost". Features, in the view adopted here, are not useful methodological labels constructed for the theorist's convenience. They are biologically real units that allow for the operations of the phonological computational system. In contemporary phonological theory this position is perhaps made explicit in the strongest terms by Hale and Reiss, 2008. The original conception of Substance-Free Phonology by Hale and Reiss, 2008, and one that has been the primary driver of almost all of my academic interests, conceivde of features as symbolic entities with intrinsic substantive content, much like SPE. The crucial contribution, however, lay in Hale and Reiss' bold re-assertion that this intrinsic content of features have no implications for phonological computations, which are blind to all such content. As I look back on what I have learned about phonology, and phonetics, and how they interface, I must admit that I have very little to argue against this original conception. As far as I can tell, and as Poeppel and Idsardi, 2022 point out, given our current state of understanding of how the brain represents information the original Hale and Reiss conception might very well be true. What I have taken issue with is the altered conception presented in later works by Volenec and Reiss, 2017. The best that I can tell, and as I have attempted to argue in chapters 3 and 5, my primary objection stems from the fact that Volenec and Reiss provide very little in the way of actual neurobiological explanations. In fact, it seems to me now that the only purpose that Cognitive Phonetics serves is to take the "intrinsic content" of SPE (and Hale and Reiss, 2008), and attempt to push it one level outward, creating an extra level of representation and two extra sets of algorithms. Regarding these algorithms, too, very little is elaborated, and it seems odd to me how little of the literature on the neural mechanisms underlying transduction is actually appealed to. Nevertheless, my interpretations are my own, and my only motivation in these discussions has been to retain what I see as the primary contribution of Hale and Reiss, 2008 – the assertion that phonological representations and computations are devoid of subtantive concerns – and combine it with what I take to be a "more" neurobiologically plausible account of the phonetic interface.

7.5 Parting remarks

This entire dissertation has been an exercise in substance-freedom, though the efforts therein have taken various forms. I have focused primarily on exploring the nature of the interface with phonetics, driven by the desire to rid phonological theory of all that it can be rid of. The ultimate objective being to have as lean a phonological UG as is computationally feasible. Yet, what remains elusive is a definition of what, if anything, is "phonological computation". To some regret of mine the dissertation includes a promissory note, in the form of chapter 6. Here I have laid out the design and procedure of a study, delayed repeatedly by circumstances beyond my control, that seeks to create a pre-theoretical toolkit that might be utilized to disambiguate the phonological from everything that is not phonological, and to identify that which is unambiguously phonetic. Given the focus of the rest of the dissertation I had hoped for this chapter to contain some insights, by the time the occasion to write it down came about, that would perhaps help elucidate some of the issues I have attempted to grapple with. This has not happened, yet. But hope remains, as I now have the time and opportunity to finally devote myself to the path laid out in chapter 6, without the burden of having to complete a dissertation. This one. I should very much like to think that the story of the ghost in the cell has only just begun, and I hope very much to stand corrected in all the errors that I have allowed to remain here as it progresses.

Bibliography

Aarsleff, Hans (1970). “The history of linguistics and Professor Chomsky”. In: *Language*, pp. 570–585.

Abrams, Kenneth and Thomas G Bever (1969). “Syntactic structure modifies attention during speech perception and recognition”. In: *The Quarterly journal of experimental psychology* 21.3, pp. 280–290.

Aizawa, Ken (2015). “What is this cognition that is supposed to be embodied?” In: *Philosophical psychology* 28.6, pp. 755–775.

Anderson, Stephen R and David W Lightfoot (2000). “The human language faculty as an organ”. In: *Annual Review of Physiology* 62.1, pp. 697–722.

Antoniou, Mark, Catherine T Best, and Michael D Tyler (2013). “Focusing the lens of language experience: Perception of Ma’di stops by Greek and English bilinguals and monolinguals”. In: *The Journal of the Acoustical Society of America* 133.4, pp. 2397–2411.

Antoniou, Mark, Michael D Tyler, and Catherine T Best (2012). “Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode?” In: *Journal of phonetics* 40.4, pp. 582–594.

Arbib, Michael A (2005). “From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics”. In: *Behavioral and Brain Sciences* 28.2, pp. 105–167.

Archangeli, Diana (1988). *Underspecification in phonology*. Cambridge University Press.

Archangeli, Diana B and Douglas George Pulleyblank (1994). *Grounded phonology*. Vol. 25. MIT Press.

Aronoff, Mark et al. (2021). “Talk Isn’t So Cheap”. In: *Inference: International Review of Science* 6.1.

Bale, Alan and Charles Reiss (2018). *Phonology: A formal introduction*. MIT Press.

Beckman, Jill, Michael Jessen, and Catherine Ringen (2013). "Empirical evidence for laryngeal features: Aspirating vs. true voice languages1". In: *Journal of linguistics* 49.2, pp. 259–284.

Beckman, Jill N (1997). "Positional faithfulness, positional neutralisation and Shona vowel harmony". In: *Phonology* 14.1, pp. 1–46.

Belin, Pascal, Shirley Fecteau, and Catherine Bedard (2004). "Thinking the voice: neural correlates of voice perception". In: *Trends in cognitive sciences* 8.3, pp. 129–135.

Bell, Alexander Melville (1867). *Visible speech: The science of universal alphabetics: Or, self-interpreting physiological letters, for the writing of all languages in one alphabet. Illustrated by tables, diagrams, and examples*. Simpkin, Marshall.

Berent, Iris (2013). "The phonological mind". In: *Trends in cognitive sciences* 17.7, pp. 319–327.

— (2021). "Can we get human nature right?" In: *Proceedings of the National Academy of Sciences* 118.39, e2108274118.

Berko, Jean (1958). "The child's learning of English morphology". In: *Word* 14.2-3, pp. 150–177.

Bermúdez-Otero, Ricardo and April McMahon (2006). "English phonology and morphology". In: *The handbook of English linguistics*, pp. 382–410.

Best, Catherine T (1995a). "A direct realist view of cross-language speech perception". In: *Speech perception and linguistic experience* 171.

— (1995b). "Learning to perceive the sound pattern of English". In: *Advances in infancy research* 9, pp. 217–217.

Best, Catherine T, Gerald W McRoberts, and Elizabeth Goodell (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system". In: *The Journal of the Acoustical Society of America* 109.2, pp. 775–794.

Best, Catherine T et al. (1994). "The emergence of native-language phonological influences in infants: A perceptual assimilation model". In: *The development of speech perception: The transition from speech sounds to spoken words* 167.224, pp. 233–277.

Best, Catherine T et al. (2007). “Nonnative and second-language speech perception”. In: *Language experience in second language speech learning*, pp. 13–34.

Best, Cathy et al. (2015a). “From Newcastle MOUTH to Aussie ears: Australians’ perceptual assimilation and adaptation for Newcastle UK vowels”. In: *Interspeech 2015*.

Best, Cathy et al. (2015b). “Perceiving and adapting to regional accent differences among vowel subsystems”. In: *18th International Congress of Phonetic Sciences (ICPhS)*. Newcastle University.

Bever, Thomas G (2009). “Remarks on the individual basis for linguistic structures”. In: *Of minds and language: A dialogue with Noam Chomsky in the Basque country*, pp. 278–299.

Bever, Thomas G, Jerry A Fodor, and W Weksel (1965a). “Is linguistics empirical?” In.

Bever, Thomas G, Jerry A Fodor, and William Weksel (1965b). “On the acquisition of syntax: A critique of" contextual generalization."” In.

Bever, Thomas G and David Poeppel (2010). “Analysis by synthesis: a (re-) emerging program of research for language and vision”. In: *Biolinguistics* 4.2-3, pp. 174–200.

Binder, Jeffrey R et al. (1997). “Conceptual processing during the conscious resting state: A functional MRI study”. In: *Journal of cognitive neuroscience* 9.1, pp. 133–143.

Binder, Jeffrey R et al. (2000). “Human temporal lobe activation by speech and nonspeech sounds”. In: *Cerebral cortex* 10.5, pp. 512–528.

Blaho, Sylvia (2008). “The syntax of phonology: A radically substance-free approach”. In.

Blank, S Catrin et al. (2002). “Speech production: Wernicke, Broca and beyond”. In: *Brain* 125.8, pp. 1829–1838.

Blevins, Juliette (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Boersma, Paul, Kateřina Chládková, and Titia Benders (2022). “Phonological features emerge substance-freely from the phonetics and the morphology”. In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 611–669.

Brecht, Michael (2017). “The body model theory of somatosensory cortex”. In: *Neuron* 94.5, pp. 985–992.

Broca, Paul et al. (1861). “Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech)”. In: *Bulletin de la Société Anatomique* 6, pp. 330–357.

Bromberger, Sylvain and Morris Halle (2000). “The ontology of phonology (revised)”. In: *Phonological knowledge: conceptual & empirical issues, ed. N. Burton-Roberts, P. Carr, & G. Docherty*, pp. 19–37.

Browman, Catherine P and Louis Goldstein (1989). “Articulatory gestures as phonological units”. In: *Phonology* 6.2, pp. 201–251.

— (1992). “Articulatory phonology: An overview”. In: *Phonetica* 49.3-4, pp. 155–180.

Buckner, Cameron and James Garson (1997). “Connectionism”. In.

Bundgaard-Nielsen, Rikke L, Catherine T Best, and Michael D Tyler (2011). “Vocabulary size is associated with second-language vowel perception performance in adult learners”. In: *Studies in second language acquisition* 33.3, pp. 433–461.

Bürki, Audrey, Mirjam Ernestus, and Ulrich H Frauenfelder (2010). “Is there only one “fenêtre” in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words”. In: *Journal of Memory and Language* 62.4, pp. 421–437.

Butcher, Andrew et al. (2004). “On the back of the tongue: Dorsal sounds in Australian languages”. In: *Phonetica* 61.1, pp. 22–52.

Caldwell, R. (1998). *A Comparative Grammar of the Dravidian Or South-Indian Family of Languages*. Asian Educational Services. ISBN: 9788120601178. URL: `https://books.google.co.in/books?id=5PPCYBApSnIC`.

Caramazza, Alfonso et al. (2001). “A crosslinguistic investigation of determiner production”. In: *Language, brain, and cognitive development: Essays in honor of Jacques Mehler*, pp. 209–226.

Carey, Susan and Carol Smith (1993). “On understanding the nature of scientific knowledge”. In: *Educational psychologist* 28.3, pp. 235–251.

Carnie, Andrew and David Medeiros (2005). “Tree maximization and the generalized Extended Projection Principle”. In.

Chabot, Alex (2022a). “On substance and Substance-Free Phonology: Where we are at and where we are going”. In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 429–443.

— (2022b). “On substance and Substance-Free Phonology: Where we are at and where we are going”. In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 429–443.

Chabot, Alexander Marx (2021). “Langues possibles et impossibles: Naturalité, facteurs thiers, et phonologie sans substance à la lumière des règles folles”. PhD thesis. Université Côte d’Azur.

Chabot, Alexander Marx. (2023). *What phonology is and why it should be*. Lingbuzz.

Cherry, E Colin (1953). “Some experiments on the recognition of speech, with one and with two ears”. In: *The Journal of the acoustical society of America* 25.5, pp. 975–979.

Chomsky, N and R Dinozzi (1972). “Language and mind: Harcourt Brace Jovanovich”. In: *New York*.

Chomsky, N., M. Halle, and P. Encrevé (1973). *Principes de phonologie générative*. Éditions du Seuil.

Chomsky, Noam (1953). “Systems of syntactic analysis”. In: *The Journal of Symbolic Logic* 18.3, pp. 242–256.

— (1956). “Three models for the description of language”. In: *IRE Transactions on information theory* 2.3, pp. 113–124.

— (1959). “A review of BF Skinner’s" verbal behavior"”. In: *Language* 35.1, pp. 26–58.

— ([1966] 2009). *Cartesian linguistics: A chapter in the history of rationalist thought*. Cambridge University Press.

— (1980). “Rules and representations”. In: *Behavioral and brain sciences* 3.1, pp. 1–15.

Chomsky, Noam (1985). "The logical structure of linguistic theory". In.

— (1995). "Language and nature". In: *Mind* 104.413, pp. 1–61.

— (2000). "Language as a natural object". In.

— (2002). *On nature and language*. Cambridge University Press.

— (2002[1957]). *Syntactic structures*. Mouton de Gruyter.

— (2005). "Three factors in language design". In: *Linguistic inquiry* 36.1, pp. 1–22.

— (2014). *The minimalist program*. MIT press.

— (2014[1965]). *Aspects of the Theory of Syntax*. 11. MIT press.

— (2017). "Two notions of modularity". In: *On concepts, modules, and language: Cognitive science at its core*, pp. 25–40.

Chomsky, Noam, Ángel J Gallego, and Dennis Ott (2019). "Generative grammar and the faculty of language: Insights, questions, and challenges". In: *Catalan Journal of Linguistics*, pp. 229–261.

Chomsky, Noam and Morris Halle (1965). "Some controversial questions in phonological theory". In: *Journal of linguistics* 1.2, pp. 97–138.

— (1968). "The sound pattern of English." In.

Chomsky, Noam and Marcel P Schützenberger (1959). "The algebraic theory of context-free languages". In: *Studies in Logic and the Foundations of Mathematics*. Vol. 26. Elsevier, pp. 118–161.

Chomsky, Noam et al. (1963). "Introduction to the formal analysis of natural languages". In: *1963*, pp. 269–321.

Clements, George N (1993). "Lieu d'articulation des consonnes et des voyelles: une théorie unifiée". In: *Architecture des représentations phonologiques*, pp. 101–145.

Coltheart, Max (1999). "Modularity and cognition". In: *Trends in cognitive sciences* 3.3, pp. 115–120.

— (2011). "Methods for modular modelling: Additive factors and cognitive neuropsychology". In: *Cognitive neuropsychology* 28.3-4, pp. 224–240.

Courtenay, Baudouin de (1870[1972]). *Some general remarks on linguistics and language*. Bloomington London: Indiana University Press, pp. 49–80.

Cutler, Anne (1992). "Why not abolish psycholinguistics?" In: *Phonologica 1988*. Cambridge University Press, pp. 77–87.

— (1994). "Segmentation problems, rhythmic solutions". In: *Lingua* 92, pp. 81–104.

— (2012). *Native listening: Language experience and the recognition of spoken words*. Mit Press.

— (2017). *Twenty-first century psycholinguistics: Four cornerstones*. Routledge.

Cutler, Anne and David A Fay (1982). "One mental lexicon, phonologically arranged: comments on Hurford's comments". In: *Linguistic Inquiry*, pp. 107–113.

Dabbous, Rim et al. (2021). "Satisfying long-distance relationships (without tiers)". In.

Dennett, Daniel C (2002). "Evolution, error, and intentionality". In: *Contemporary Materialism*. Routledge, pp. 265–296.

Donegan, Patricia J and David Stampe (1979). "The study of natural phonology". In: *Current approaches to phonological theory* 126173.

Doupe, Allison J and Patricia K Kuhl (1999). "Birdsong and human speech: Common themes and mechanisms". In: *Annual Review of Neuroscience* 22, pp. 567–631.

Dresher, B Elan (2014). "The arch not the stones: Universal feature theory without universal features". In: *Nordlyd* 41.2, pp. 165–181.

Dresher, B Elan and Daniel Currie Hall (2022). "18 Developments Leading Toward Generative Phonology". In.

Duanmu, San (2000). *Phonology of Standard Chinese*. Oxford University Press Oxford:

Dunbar, Ewan, Brian Dillon, and William J Idsardi (2013). "A Bayesian evaluation of the cost of abstractness". In: *Language down the garden path: The cognitive and biological basis for linguistic structures*, pp. 360–384.

Eggermont, Jos J (1998). "Is there a neural code?" In: *Neuroscience & Biobehavioral Reviews* 22.2, pp. 355–370.

Embick, David and David Poeppel (2015). “Towards a computational (ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics”. In: *Language, cognition and neuroscience* 30.4, pp. 357–366.

Faris, Mona M, Catherine T Best, and Michael D Tyler (2016). “An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized”. In: *The Journal of the Acoustical Society of America* 139.1, EL1–EL5.

Fechner, Gustav Theodor (1860). *Elemente der psychophysik*. Vol. 2. Breitkopf u. Härtel.

Fitch, W Tecumseh (2000). “The evolution of speech: a comparative review”. In: *Trends in cognitive sciences* 4.7, pp. 258–267.

— (2010). *The evolution of language*. Cambridge University Press.

Fitch, W Tecumseh and A Cutler (2005). “Computation and cognition: Four distinctions and their implications”. In: *Twenty-first century psycholinguistics: Four cornerstones*, pp. 381–400.

Fitch, W Tecumseh, Marc D Hauser, and Noam Chomsky (2005). “The evolution of the language faculty: Clarifications and implications”. In: *Cognition* 97.2, pp. 179–210.

Flemming, Edward S (2013). *Auditory representations in phonology*. Routledge.

Fodor, Jerry, A Bever, TG Garrett, et al. (1974). “The psychology of language: An introduction to psycholinguistics and generative grammar”. In.

Fodor, Jerry and Massimo Piattelli-Palmarini (2011). *What Darwin got wrong*. Profile books.

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

— (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. MIT press.

Fodor, Jerry A and Thomas G Bever (1965). “The psychological reality of linguistic segments”. In: *Journal of verbal learning and verbal behavior* 4.5, pp. 414–420.

Fodor, Jerry A and Zenon W Pylyshyn (1988). “Connectionism and cognitive architecture: A critical analysis”. In: *Cognition* 28.1-2, pp. 3–71.

Fudge, Eric C (1967). “The nature of phonological primes”. In: *Journal of Linguistics* 3.1, pp. 1–36.

Gallagher, Gillian (2011). “Acoustic and articulatory features in phonology–the case for [long VOT]”. In.

Gallistel, Charles R and Adam Philip King (2011). *Memory and the computational brain: Why cognitive science will transform neuroscience*. John Wiley & Sons.

Gallistel, CR (1981). “Matters of principle: Hierarchies, representations, and action”. In: *Behavioral and Brain Sciences* 4.4, pp. 639–650.

— (2001). “Psychology of mental representation”. In: *International encyclopedia of the social and behavioral sciences*, pp. 9691–9695.

Gervain, Judit, Iris Berent, and Janet F Werker (2012). “Binding at birth: The newborn brain detects identity relations and sequential position in speech”. In: *Journal of Cognitive Neuroscience* 24.3, pp. 564–574.

Gick, Bryan, Ian Wilson, and Donald Derrick (2013). *Articulatory phonetics*. John Wiley & Sons.

Giraud, Anne-Lise and David Poeppel (2012). “Cortical oscillations and speech processing: emerging computational principles and operations”. In: *Nature neuroscience* 15.4, pp. 511–517.

Gleason, Jean Berko (2019). *The Wug Test*. Larchwood Press.

Goldsmith, John (1976a). “Autosegmental phonology”. PhD thesis. MIT.

Goldsmith, John Anton (1976b). “Autosegmental phonology”. PhD thesis. Massachusetts Institute of Technology.

Golston, Chris (1996). “Direct optimality theory: Representation as pure markedness”. In: *Language*, pp. 713–748.

Graf, Thomas (2022). “Subregular linguistics: bridging theoretical linguistics and formal grammar”. In: *Theoretical Linguistics* 48.3-4, pp. 145–184.

Green, Alexander A et al. (2017). “Complex cellular logic computation using ribocomputing devices”. In: *Nature* 548.7665, pp. 117–121.

Green, Christopher and John Vervaeke (1997). “But what have you done for us lately”. In: *The future of the cognitive revolution*, pp. 149–163.

Grosjean, François (1988). “Exploring the recognition of guest words in bilingual speech”. In: *Language and cognitive processes* 3.3, pp. 233–274.

Guenther, Frank H and Gregory Hickok (2016). “Neural models of motor speech control”. In: *Neurobiology of language*. Elsevier, pp. 725–740.

Guignard Guion, Susan (1998). “The role of perception in the sound change of velar palatalization”. In: *Phonetica* 55.1-2, pp. 18–52.

Guion, Susan Guignard (1996). *Velar palatalization: Coarticulation, perception, and sound change*. The University of Texas at Austin.

Gwilliams, Laura et al. (2018). “In spoken word recognition, the future predicts the past”. In: *Journal of Neuroscience* 38.35, pp. 7585–7599.

Gwilliams, Laura et al. (2022). “Neural dynamics of phoneme sequences reveal position-invariant code for content and order”. In: *Nature communications* 13.1, p. 6606.

Hackett, Troy A, Todd M Preuss, and Jon H Kaas (2001). “Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans”. In: *Journal of Comparative Neurology* 441.3, pp. 197–222.

Hale, Mark and Charles Reiss (2000a). “" Substance abuse" and" dysfunctionalism": Current trends in phonology”. In: *Linguistic inquiry* 31.1, pp. 157–169.

— (2000b). “Phonology as cognition”. In: *Phonological knowledge: Conceptual and empirical issues*, pp. 161–184.

— (2003). “The Subset Principle in phonology: why the tabula can’t be rasa”. In: *Journal of Linguistics* 39.2, pp. 219–244.

— (2008). *The phonological enterprise*. OUP Oxford.

Hall, T Alan (2001). “Introduction: Phonological representations and phonetic implementation of distinctive features”. In: *Distinctive feature theory*, pp. 1–40.

Halle, Morris (1992). “Phonological features”. In: *International encyclopedia of linguistics* 3, pp. 207–212.

— (2002). “Speculations about the representations of words in memory”. In: *Reprinted in Halle*, pp. 122–136.

— (2005). “Palatalization/velar softening: What it is and what it tells us about the nature of language”. In: *Linguistic inquiry* 36.1, pp. 23–41.

Halle, Morris, Bert Vaux, and Andrew Wolfe (2000). “On feature spreading and the representation of place of articulation”. In: *Linguistic inquiry* 31.3, pp. 387–444.

Hammarberg, Robert (1981). “The cooked and the raw”. In: *Journal of information science* 3.6, pp. 261–267.

Hauser, Marc D, Noam Chomsky, and W Tecumseh Fitch (2002). “The faculty of language: what is it, who has it, and how did it evolve?” In: *science* 298.5598, pp. 1569–1579.

Hickok, Gregory (2022). “The dual stream model of speech and language processing”. In: *Handbook of Clinical Neurology* 185, pp. 57–69.

Hickok, Gregory, John Houde, and Feng Rong (2011). “Sensorimotor integration in speech processing: computational basis and neural organization”. In: *Neuron* 69.3, pp. 407–422.

Hickok, Gregory and David Poeppel (2004). “Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language”. In: *Cognition* 92.1-2, pp. 67–99.

— (2007). “The cortical organization of speech processing”. In: *Nature reviews neuroscience* 8.5, pp. 393–402.

Hirschfeld, Lawrence A and Susan A Gelman (1994). *Mapping the mind: Domain specificity in cognition and culture*. Cambridge University Press.

Honeybone, Patrick (2005). “Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications”. In: *The internal organization of phonological segments* 319, p. 54.

Hornstein, Norbert and Cedric Boeckx (2009). “Approaching universals from below: I-universals in light of a minimalist program for linguistic theory”. In: *Language universals, ed. MH Christiansen, C. Collins, & S. Edelman*, pp. 79–98.

Hyman, Larry M. (1985). *A theory of phonological weight*. Dordrecht: Foris.

Hyman, Larry M (2013). “Enlarging the scope of phonologization”. In: *Origins of sound change: Approaches to phonologization*, pp. 3–28.

Iacoboni, Marco and Mirella Dapretto (2006). “The mirror neuron system and the consequences of its dysfunction”. In: *Nature Reviews Neuroscience* 7.12, pp. 942–951.

Idsardi, William (2022). “Underspecification in time”. In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 670–682.

Indefrey, Peter and Willem JM Levelt (2004). “The spatial and temporal signatures of word production components”. In: *Cognition* 92.1-2, pp. 101–144.

Inkelas, Sharon and Young-mee Yu Cho (1993). “Inalterability as prespecification”. In: *Language*, pp. 529–574.

Isac, Daniela and Charles Reiss (2013). *I-language: An introduction to linguistics as cognitive science*. Oxford University Press.

Iverson, Gregory K and Sang-Cheol Ahn (2007). “English voicing in dimensional theory”. In: *Language Sciences* 29.2-3, pp. 247–269.

Iverson, Gregory K and Joseph C Salmons (1995). “Aspiration and laryngeal representation in Germanic”. In: *Phonology* 12.3, pp. 369–396.

— (2003). “Laryngeal enhancement in early Germanic”. In: *Phonology* 20.1, pp. 43–74.

Iverson, Gregory K, Joseph C Salmons, et al. (1999). “Glottal spreading bias in Germanic”. In: *Linguistische Berichte*, pp. 135–151.

Jacobs, Haike and Janine Berns (2012). “Perception, production and markedness in sound change: French velar palatalization”. In: *Research on Old French: The state of the art*. Springer, pp. 107–122.

Jakobson, Roman (1949). *On the identification of phonemic entities*. Nordisk Sprog-og Kulturforlag.

— (1971). “The Kazan’school of Polish linguistics and its place in the international development of phonology”. In: *Roman Jakobson, Selected writings II. Word and language*, pp. 394–428.

— (1973). *Essais de linguistique générale*. traduit en français par Nicolas Ruwet. Paris: Éditions de Minuit.

James, William (1890). “The consciousness of self.” In.

— (2007[1890]). *The principles of psychology*. Vol. 1. Cosimo, Inc.

Jessen, Michael and Catherine Ringen (2002). “Laryngeal features in German”. In: *Phonology* 19.2, pp. 189–218.

Jiang, Haowen (2010). “Malayalam: A grammatical sketch and a text”. In: *Department of Linguistics, Rice University*.

Just, Marcel Adam et al. (2004). “Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity”. In: *Brain* 127.8, pp. 1811–1821.

Kager, René et al. (2007). “Representations of [voice]: Evidence from acquisition”. In: URL: `https://api.semanticscholar.org/CorpusID:61644289`.

Keating, Patricia and Aditi Lahiri (1993). “Fronted velars, palatalized velars, and palatals”. In: *Phonetica* 50.2, pp. 73–101.

Keating, Patricia A (1984). “Phonetic and phonological representation of stop consonant voicing”. In: *Language*, pp. 286–319.

Kehrein, Wolfgang and Chris Golston (2004). “A prosodic theory of laryngeal contrasts”. In: *Phonology* 21.3, pp. 325–357.

Kemmerer, David (2014). *Cognitive neuroscience of language*. Psychology Press.

Kenstowicz, Michael J (1994). *Phonology in generative grammar*. Vol. 7. Blackwell Cambridge, MA.

Keyser, ANTOINE (1994). “Carl Wernicke”. In: *Reader in the history of aphasia: From Franz Gall to Norman Geschwind* 4, p. 59.

Kimura, Doreen (1961a). “Cerebral dominance and the perception of verbal stimuli.” In: *Canadian Journal of Psychology/Revue canadienne de psychologie* 15.3, p. 166.

— (1961b). “Some effects of temporal-lobe damage on auditory perception.” In: *Canadian Journal of Psychology/Revue canadienne de psychologie* 15.3, p. 156.

— (1967). “Functional asymmetry of the brain in dichotic listening”. In: *Cortex* 3.2, pp. 163–178.

Kochetov, Alexei and John Alderete (2011). “Patterns and scales of expressive palatalization: Experimental evidence from Japanese”. In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 56.3, pp. 345–376.

Kotzor, Sandra, Allison Wetterlin, and Aditi Lahiri (2017). “Symmetry or asymmetry: Evidence for underspecification in the mental lexicon”. In: *The Speech Processing Lexicon: Neurocognitive and Behavioural Approaches* 22, p. 85.

Krakauer, John W et al. (2017). “Neuroscience needs behavior: correcting a reductionist bias”. In: *Neuron* 93.3, pp. 480–490.

Krivochen, Diego and Douglas Saddy (2018). “Towards a classification of Lindenmayer systems”. In: *arXiv preprint arXiv:1809.10542*.

Kuhl, Patricia K (1991). “Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not”. In: *Perception & psychophysics* 50.2, pp. 93–107.

— (2004). “Early language acquisition: cracking the speech code”. In: *Nature reviews neuroscience* 5.11, pp. 831–843.

— (2010). “Brain mechanisms in early language acquisition”. In: *Neuron* 67.5, pp. 713–727.

Kuhl, Patricia K and Andrew N Meltzoff (1996). “Infant vocalizations in response to speech: Vocal imitation and developmental change”. In: *Journal of the Acoustical Society of America* 100.4, pp. 2425–2438.

Kuhl, Patricia K and James D Miller (1978). “Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli”. In: *The Journal of the Acoustical Society of America* 63.3, pp. 905–917.

Kuhn, Thomas S (2012[1967]). *The structure of scientific revolutions*. University of Chicago press.

Ladefoged, Peter and Ian Maddieson (1996). “The sounds of the world’s languages”. In: *(No Title)*.

Lahiri, Aditi (2012). “Neural correlates of underspecification in phonology”. In: *Journal of Phonetics* 45.3, pp. 123–135.

Lahiri, Aditi and Henning Reetz (2002). "Underspecified recognition". In: *Laboratory phonology* 7, pp. 637–675.

— (2010). "Distinctive features: Phonological underspecification in representation and processing". In: *Journal of Phonetics* 38.1, pp. 44–59.

Lawyer, Laurel and David Corina (2014). "An investigation of place and voice features using fMRI-adaptation". In: *Journal of neurolinguistics* 27.1, pp. 18–30.

Lenneberg, Eric H (1953). "Cognition in ethnolinguistics". In: *Language* 29.4, pp. 463–471.

— (1967). "The biological foundations of language". In: *Hospital Practice* 2.12, pp. 59–67.

— (1969). "On Explaining Language: The development of language in children can best be understood in the context of developmental biology." In: *Science* 164.3880, pp. 635–643.

Levelt, Willem JM (2013). *A history of psycholinguistics: The pre-Chomskyan era*. Oxford University Press.

Lewis, Shevaun and Colin Phillips (2015). "Aligning grammatical theories and language processing models". In: *Journal of Psycholinguistic Research* 44, pp. 27–46.

Liberman, Alvin M and Ignatius G Mattingly (1985). "The motor theory of speech perception revised". In: *Cognition* 21.1, pp. 1–36.

Liberman, Alvin M et al. (1967). "Perception of the speech code." In: *Psychological review* 74.6, p. 431.

Lichtheim, Ludwig (1885). "On aphasia". In: *Brain* 7, pp. 433–484.

Lisker, Leigh and Arthur S Abramson (1964). "A cross-language study of voicing in initial stops: Acoustical measurements". In: *Word* 20.3, pp. 384–422.

Lombardi, Linda (2018). *Laryngeal features and laryngeal neutralization*. Routledge.

Maddieson, I (1984). "Patterns of Sounds (Cambridge UP, Cambridge)". In.

Mailhot, Frédéric and Charles Reiss (2007). "Computing long-distance dependencies in vowel harmony". In: *Biolinguistics* 1, pp. 028–048.

Mandal, Savantan et al. (2020). "Bilingual phonology in dichotic perception: A case study of Malayalam and English voicing". In: *Glossa: a journal of general linguistics* 5.1.

Mandal, Sayantan (2021). "Cognitive Revolution, The". In: *Encyclopedia of Evolutionary Psychological Science*, pp. 1167–1177.

— (2023). "Cognitive Phonetics of Velar Palatalization, The". In: *Proceedings of Western Conference of Linguitics*, pp. 8–19.

Marcus, Gary F (2003). *The algebraic mind: Integrating connectionism and cognitive science*. MIT press.

Marr, David (1970). "A theory for cerebral neocortex". In: *Proceedings of the Royal society of London. Series B. Biological sciences* 176.1043, pp. 161–234.

— (1977). "Artificial intelligence—a personal view". In: *Artificial Intelligence* 9.1, pp. 37–48.

— (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

Marslen-Wilson, William D (1975). "Sentence perception as an interactive parallel process". In: *Science* 189.4198, pp. 226–228.

— (1987). "Functional parallelism in spoken word-recognition". In: *Cognition* 25.1-2, pp. 71–102.

Martins, Pedro Tiago and Cedric Boeckx (2016). "What we talk about when we talk about biolinguistics". In: *Linguistics Vanguard* 2.1, p. 20160007.

Mayr, Ernst (1961). "Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist." In: *Science* 134.3489, pp. 1501–1506.

— (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.

McClelland, James L and Jeffrey L Elman (1986). "The TRACE model of speech perception". In: *Cognitive psychology* 18.1, pp. 1–86.

Medeiros, David P, Massimo Piattelli-Palmarini, and Thomas G Bever (2016). "Many important language universals are not reducible to processing or cognition". In: *Behavioral and Brain Sciences* 39, e86.

Mesgarani, Nima et al. (2014). “Phonetic feature encoding in human superior temporal gyrus”. In: *Science* 343.6174, pp. 1006–1010.

Mielke, Jeff (2008). *The emergence of distinctive features*. Oxford University Press.

Miller, George A (1956). “The magical number seven, plus or minus two: Some limits on our capacity for processing information.” In: *Psychological review* 63.2, p. 81.

— (2003). “The cognitive revolution: a historical perspective”. In: *Trends in cognitive sciences* 7.3, pp. 141–144.

Miller, George A and Noam Chomsky (1963). “Finitary models of language users”. In.

Miller, George Armitage (1951). “Language and communication”. In.

Minsky, Marvin (1961). “Steps toward artificial intelligence”. In: *Proceedings of the IRE* 49.1, pp. 8–30.

Minsky, Marvin and Seymour A Papert (2017). *Perceptrons, reissue of the 1988 expanded edition with a new foreword by Léon Bottou: an introduction to computational geometry*. MIT press.

Mohanan, Karuvannur P and Tara Mohanan (1984). “Lexical phonology of the consonant system in Malayalam”. In: *Linguistic inquiry*, pp. 575–602.

Mohanan, Karuvannur Puthanveettil (1982). “Lexical phonology”. PhD thesis. Massachusetts Institute of Technology.

Mohanan, KP and KP Mohanan (1986). “Lexical Phonology and Psychological Reality”. In: *The Theory of Lexical Phonology*, pp. 182–204.

Monahan, Philip J, Ellen F Lau, and William J Idsardi (2013). “Computational primitives in phonology and their neural correlates”. In: *The Cambridge Handbook of Biolinguistics*, pp. 233–256.

Monahan, Philip J, Alejandro Pérez, and Jessamyn Schertz (n.d.). “Abstract Phonological Features: EEG Evidence from English Voicing”. In: ().

Monahan, Philip J et al. (2022). “Unified Coding of Spectral and Temporal Phonetic Cues: Electrophysiological Evidence for Abstract Phonological Features”. In: *Journal of Cognitive Neuroscience* 34.4, pp. 618–638.

Moreton, Elliott (2007). "Analytic bias as a factor in phonological typology". In: *Proceedings of WCCFL*. Vol. 26, pp. 393–401.

— (2008). "Analytic bias and phonological typology". In: *Phonology* 25.1, pp. 83–127.

Moro, Andrea et al. (2001). "Syntax and the brain: disentangling grammar by selective anomalies". In: *Neuroimage* 13.1, pp. 110–118.

Newell, Allen and Herbert Simon (1956). "The logic theory machine–A complex information processing system". In: *IRE Transactions on information theory* 2.3, pp. 61–79.

Newmeyer, Frederick J (1986/2014). *Linguistic theory in America*. Elsevier.

Norris, Dennis (1994). "Shortlist: A connectionist model of continuous speech recognition". In: *Cognition* 52.3, pp. 189–234.

Norris, Dennis, James M McQueen, and Anne Cutler (2003). "Perceptual learning in speech". In: *Cognitive psychology* 47.2, pp. 204–238.

Odden, David (2022). "Radical substance-free phonology and feature learning". In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 500–551.

Ohala, Manjari and John J Ohala (1991). "Nasal epenthesis in Hindi". In: *Phonetica* 48.2-4, pp. 207–220.

Ohala, Manjari and Manjari Ohala (1975). "Nasals and nasalization in Hindi". In: *Nasalfest*, pp. 317–32.

Pa, Judy and Gregory Hickok (2008). "A parietal–temporal sensory–motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians". In: *Neuropsychologia* 46.1, pp. 362–368.

Pandey, Pramod Kumar (1990). "Hindi schwa deletion". In: *Lingua* 82.4, pp. 277–311.

Peterson, Steven E et al. (1988). "Positron emission tomographic studies of the processing of single words". In: *Journal of cognitive neuroscience* 1.2, pp. 153–170.

Petkov, Christopher I and Erich D Jarvis (2012). "Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates". In: *Frontiers in Evolutionary Neuroscience* 4, p. 12.

Petrova, Olga et al. (2006). "Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish". In.

Piattelli-Palmarini, Massimo and Juan Uriagereka (2004). "The immune syntax: the evolution of the language virus". In: *Variation and universals in biolinguistics*. Brill, pp. 341–377.

— (2008). "Still a bridge too far? Biolinguistic questions for grounding language on brains". In: *Physics of Life Reviews* 5.4, pp. 207–224.

Piattelli-Palmarini, Massimo and Giuseppe Vitiello (2017). "Quantum field theory and the linguistic Minimalist Program: a remarkable isomorphism". In: *Journal of Physics: Conference Series*. Vol. 880. 1. IOP Publishing, p. 012016.

Pierrehumbert, Janet B (2003). "Phonetic diversity, statistical learning, and acquisition of phonology". In: *Language and speech* 46.2-3, pp. 115–154.

Pinker, Steven (2003). *How the mind works*. Penguin UK.

— (2005). "So how does the mind work?" In: *Mind & language* 20.1, pp. 1–24.

Plotkin, Henry (1997). *Darwin machines and the nature of knowledge*. Harvard University Press.

Poeppel, D et al. (2004). "FM sweeps, syllables, and word stimuli differentially modulate left and right non-primary auditory areas". In: *Neuropsychologia* 42.2, pp. 183–200.

Poeppel, David (2012). "The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language". In: *Cognitive neuropsychology* 29.1-2, pp. 34–55.

Poeppel, David and M Florencia Assaneo (2020). "Speech rhythms and their neural foundations". In: *Nature reviews neuroscience* 21.6, pp. 322–334.

Poeppel, David and David Embick (2017). "Defining the relation between linguistics and neuroscience". In: *Twenty-first century psycholinguistics: Four cornerstones*. Routledge, pp. 103–118.

Poeppel, David and William Idsardi (2011). "Recognizing words from speech: the perception-action-memory loop". In: *Lexical representation: A multidisciplinary approach* 17, p. 171.

— (2022). "We don't know how the brain stores anything, let alone words". In: *Trends in Cognitive Sciences*.

Poeppel, David, William J Idsardi, and Virginie Van Wassenhove (2008). “Speech perception at the interface of neurobiology and linguistics”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1493, pp. 1071–1086.

Poeppel, David et al. (2012). “Towards a new neurobiology of language”. In: *Journal of Neuroscience* 32.41, pp. 14125–14131.

Pöppel, Ernst (1997). “A hierarchical model of temporal perception”. In: *Trends in cognitive sciences* 1.2, pp. 56–61.

Price, Cathy J, Richard J Wise, and Richard S Frackowiak (1996). “Demonstrating the implicit processing of visually presented words and pseudowords”. In: *Cerebral cortex* 6.1, pp. 62–70.

Prince, Alan and Paul Smolensky (2004). “Optimality Theory: Constraint interaction in generative grammar”. In: *Optimality Theory in phonology: A reader*, pp. 1–71.

Pylyshyn, Zenon W (1980). “Computation and cognition: Issues in the foundations of cognitive science”. In: *Behavioral and Brain sciences* 3.1, pp. 111–132.

Pylyshyn, Zenon Walter (1984). “Computation and cognition”. In.

Ramaswami, Ayyar (1936). *The Evolution of Malayalam Morphology*. Cochin Government Press, pp. 1–37.

Redcay, Elizabeth and Eric Courchesne (2008). “Deviant functional magnetic resonance imaging patterns of brain activity to speech in 2-3-year-old children with autism spectrum disorder”. In: *Biological psychiatry* 64.7, pp. 589–598.

Reiss, Charles (1998). “Should Output-Output Correspondence be Invoked to Account for Analogy?” In.

— (2003). “Quantification in structural descriptions: Attested and unattested patterns”. In.

— (2007). “Modularity in the “Sound” domain: Implications for the purview of universal grammar”. In.

— (2017). “Substance free phonology”. In: *The Routledge handbook of phonological theory*. Routledge, pp. 425–452.

— (To Appear). “Research Methods in Armchair Linguistics”. In: *To Appear*.

Reiss, Charles and Veno Volenec (2021). "Naturalism, Internalism, and Nativism:< What> The Legacy of The Sound Pattern of English< Should Be>". In: *A Companion to Chomsky*, pp. 96–108.

— (2022b). "Conquer primal fear: Phonological features are innate and substance-free". In: *Canadian Journal of Linguistics/Revue canadienne de linguistique*, pp. 1–30.

— (2022a). "Conquer primal fear: Phonological features are innate and substance-free". In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 581–610.

Reiss, Diana and Brenda McCowan (1993). "Spontaneous vocal mimicry and production by bottlenose dolphins (Tursiops truncatus): Evidence for vocal learning". In: *Journal of Comparative Psychology* 107.3, pp. 301–312.

Roberts, Ian G, Jeffrey Watumull, and Noam Chomsky (2023). "Universal Grammar". In: *Xenolinguistics*. Routledge, pp. 165–181.

Rubach, Jerzy (2003). "Polish palatalization in derivational optimality theory". In: *Lingua* 113.3, pp. 197–237.

Rumelhart, David E, Geoffrey E Hinton, James L McClelland, et al. (1986). "A general framework for parallel distributed processing". In: *Parallel distributed processing: Explorations in the microstructure of cognition* 1.45-76, p. 26.

Ryle, Gilbert and Julia Tanney ([1949]2009). *The concept of mind*. Routledge.

Sagey, Elizabeth Caroline (1986). "The representation of features and relations in non-linear phonology". PhD thesis. Massachusetts Institute of Technology.

Sahin, Ned T et al. (2009). "Sequential processing of lexical, grammatical, and phonological information within Broca's area". In: *Science* 326.5951, pp. 445–449.

Samuels, Bridget (2009). "The third factor in phonology". In: *Biolinguistics* 3.2-3, pp. 355–382.

— (2015a). "Biolinguistics in phonology: a prospectus". In: *Phonological Studies* 18, pp. 161–171.

Samuels, Bridget D (2011). *Phonological architecture: A biolinguistic perspective*. Vol. 2. Oxford University Press.

— (2015b). "Can a bird brain do phonology?" In: *Frontiers in psychology* 6, p. 1082.

Samuels, Bridget D et al. (2022). "Getting ready for primetime: Paths to acquiring substance-free phonology". In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 552–580.

Sandler, Wendy (2014). "The emergence of the phonetic and phonological features in sign language". In: *Nordlyd* 41.2, pp. 183–212.

Sanyal, Paroma (2012). "Phonology to morphophonology: Reanalyzing Bangla verbs". In: *English and Foreign Languages Journal* 3.2, pp. 65–84.

Sapir, Edward (1925). "Sound patterns in language". In: *Language* 1.2, pp. 37–51.

Saur, Dorothee et al. (2006). "Dynamics of language reorganization after stroke". In: *Brain* 129.6, pp. 1371–1384.

Saussure, Ferdinand de. (1916[1967]). *Cours de linguistique général*. Paris: Grand bibliothèque Payot.

Scheer, T., M. de Jonge, and A.M. Chabot (2020). "Velar palatalization is phonological. EEG evidence". In.

Scheer, Tobias (2009). "External sandhi: what the initial CV is initial of". In: *Studi e Saggi Linguistici* 47, pp. 43–82.

— (2010a). *A guide to morphosyntax-phonology interface theories: how extra-phonological information is treated in phonology since Trubetzkoy's Grenzsignale*. Berlin: Mouton De Gruyter.

— (2010b). "What is initial CV initial of". In: Communication présentée au colloque du Réseau Français de Phonologie, Orléans.

— (2022). "3 x Phonology". In: *Canadian Journal of Linguistics/Revue canadienne de linguistique* 67.4, pp. 444–499.

— (n.d). "Phonetic arbitrariness and its consequences". In: *Phonological Studies. Available at https://ling. auf. net/lingbuzz* 4554.

Scheer, Tobias and Fabien Mathy (2021). "Neglected factors bearing on reaction time in language production". In: *Cognitive Science* 45.10, p. 13050.

Schein, Barry and Donca Steriade (1986). “On geminates”. In: *Linguistic inquiry* 17.4, pp. 691–744.

Schiller, Francis (1992). *Paul Broca: Founder of French anthropology, explorer of the brain*. Oxford University Press, USA.

Shanmugam, SV (1976). “Formation and development of Malayalam”. In: *Indian Literature* 19.3, pp. 5–30.

Shaw, Jason et al. (2018). “Resilience of English vowel perception across regional accent variation”. In: *Laboratory Phonology* 9.1.

Shirow, Masamune (1995). “Ghost in the Shell”. In.

Skinner Burrhus, F. (1957). *Verbal Behavior*. Copley Publishing Group.

Spivey, Michael J and Viorica Marian (1999). “Cross talk between native and second languages: Partial activation of an irrelevant lexicon”. In: *Psychological science* 10.3, pp. 281–284.

Stein, Barry E and M Alex Meredith (1993). *The merging of the senses*. MIT press.

Steriade, Donca and John A Goldsmith (1995). “Underspecification and markedness”. In: *1995*, pp. 114–174.

Stevens, Kenneth N (1989). “On the quantal nature of speech”. In: *Journal of phonetics* 17.1, pp. 3–45.

— (2000). *Acoustic phonetics*. Vol. 30. MIT press.

— (2002). “Toward a model for lexical access based on acoustic landmarks and distinctive features”. In: *The Journal of the Acoustical Society of America* 111.4, pp. 1872–1891.

Stevens, Kenneth N and Sheila E Blumstein (1981). “The search for invariant acoustic correlates of phonetic features”. In: *Perspectives on the study of speech*, pp. 1–38.

Stokoe, William C (1980). “Sign language structure”. In: *Annual review of anthropology* 9.1, pp. 365–390.

Thulborn, Keith R, Paul A Carpenter, and Marcel A Just (1999). “Plasticity of language-related brain function during recovery from stroke”. In: *Stroke* 30.4, pp. 749–754.

Troubetzkoy, Nikolaï (1967). *Principes de phonologie*. Paris: Klincksieck.

Turing, Alan M (2009). *Computing machinery and intelligence*. Springer.

Turing, Alan Mathison (2004). *The essential turing*. Oxford University Press.

Van Oostendorp, Marc (2006). “Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), Phonetically based phonology. Cambridge: Cambridge University Press, 2004. Pp. viii+ 375.” In: *Journal of Linguistics* 42.2, pp. 467–473.

Volenec, Veno and Charles Reiss (2017). “Cognitive phonetics: The transduction of distinctive features at the phonology-phonetics interface”. In: *Biolinguistics* 11, pp. 251–294.

— (2020). “Formal generative phonology”. In: *Radical: A Journal of Phonology* 2, pp. 1–148.

Ward, Lawrence M (2003). “Synchronous neural oscillations and cognitive processes”. In: *Trends in cognitive sciences* 7.12, pp. 553–559.

Warrier, Tara (1976). “The Phonetics and Phonology of Malayalam and Its Pedagogical Implications-A Generative Phonological Study”. In: *Diss. CIEFL: Hyderabad*.

Watumull, Jeffrey and Noam Chomsky (2020). “Rethinking universality”. In: *Syntactic architecture and its consequences II*, p. 3.

Watumull, Jeffrey et al. (2014). “On recursion”. In: *Frontiers in Psychology* 4, p. 1017.

Weber, Andrea and Anne Cutler (2004). “Lexical competition in non-native spoken-word recognition”. In: *Journal of memory and language* 50.1, pp. 1–25.

Weerathunge, Hasini R et al. (2022). “LaDIVA: A neurocomputational model providing laryngeal motor control for speech acquisition and production”. In: *PLoS computational biology* 18.6, e1010159.

Werker, Janet F and Richard C Tees (1984). “Cross-language speech perception: Evidence for perceptual reorganization during the first year of life”. In: *Infant behavior and development* 7.1, pp. 49–63.

Whitehead, Alfred N and Bertrand Russell (1926[1912]). “Principia mathematica”. In: *Nueva York, Anchor Books*.

Wilson, Colin (2006). “Learning phonology with substantive bias: An experimental and computational study of velar palatalization”. In: *Cognitive science* 30.5, pp. 945–982.

Wilson, Edward O (2000). *Sociobiology: The new synthesis*. Harvard University Press.

Wilson, Stephen M et al. (2004). “Listening to speech activates motor areas involved in speech production”. In: *Nature neuroscience* 7.7, pp. 701–702.

Wundt, Wilhelm Max (1893). *Grundzüge der physiologischen Psychologie*. Vol. 2. W. Engelmann.

Yang, Charles et al. (2017). “The growth of language: Universal Grammar, experience, and principles of computation”. In: *Neuroscience & Biobehavioral Reviews* 81, pp. 103–119.

Yip, Moira (2003). “What phonology has learnt from Chinese”. In: *Glot International* 7.1/2, pp. 26–35.

Yu, Xinchi and Ellen Lau (2023). “The binding problem 2.0: beyond perceptual features”. In: *Cognitive Science* 47.2, e13244.