Modal-Dependence and Naturalness in Phonology: Confronting the Ontogenetic Question

LIN-3990

Joe Collins

Masters Thesis in Theoretical Linguistics
Faculty of Humanities, Social Sciences and Education
University of Tromsø
Spring 2013

Contents

Acknowledgments

I would like to thank my supervisor, Martin Krämer, for his guidance and support during my time as his student. His open mindedness and wealth of knowledge are the mark of a true scholar, and this dissertation has certainly benefited from his insightful feedback.

I would also like to thank the members of the Spring 2013 phonology reading group, Kevin, Olga and Elena. The discussions we had and the feedback I received were invaluable in the writing of this dissertation. Special thanks to Kevin for checking my arguments and letting me know when I'm not making sense.

Thanks to Eugene Buckley for providing me with a copy of Robert Oswalt's 1961 dissertation on the Kashaya language, and thanks for providing feedback on an early draft of my Kashaya analysis. I hope that in some small way this dissertation might contribute to the continued study of endangered American languages.

Finally, I would like to thank my wife Ine. I couldn't survive the real world without you.

Chapter 1

Introduction

Phonologists have long debated the role of phonetics in phonological theories. A key point in this debate is whether or not phonological computation is to be regarded as modally-independent, that is, whether or not facts about the articulatory and perceptual organs are regarded as external to the phonology. The modally-independent view is typified by *substance free* phonology (Hale & Reiss, 2008), where the phonology is perceived as a module of the mind performing computations over abstract symbols. Under this view, any apparent tendencies towards phonetic concerns, e.g. articulatory ease, are merely indirect products of the limits imposed by the equipment the phonology is 'plugged into', not properties of the phonology itself (Hale & Reiss, 2008; Buckley, 2000). On the side of the debate, modally-dependent approaches vary between attempts to incorporate phonetics directly into the phonology (e.g. Flemming, 2001), and *grounded phonology* which supposes that the phonology embodies certain phonetically motivated principles (e.g. markedness), but is nonetheless a form of self-contained, symbolic computation (Archangeli & Pulleyblank, 1994; Hayes, 1999; Bermudez-Otero & Börjars, 2006).

This dissertation is an extended argument for a form of grounded phonology. The argument depends on confronting the ontogenetic question of phonology, that is, the question of how phonology grows and develops in the brain. A review of neuroscientific literature demonstrates that a modally-dependent phonology is far more biologically plausible. Additionally, linguistic analyses in an Optimality Theoretic framework demonstrate that a phonetically grounded phonology is still capable of explaining phonetically unnatural patterns in natural language.

Document Structure

Chapter 2 addresses the question of the ontogeny of phonology, that is, how phonology grows and develops in the brain. By examining findings from neuroscience and developmental neurobiology, this chapter will present the argument that a modally-dependent phonology is more biologically plausible than a modally-independent phonology.

Chapter 3 will discuss some outstanding issues surrounding the function of phonology.

Chapter 4 sketches a model of language learning which reflects the 'autonomous but grounded' view of phonology, within an Optimality Theoretic framework (OT). This chapter will argue that both features and markedness constraints are phonetically grounded, but that exposure to primary linguistic data (PLD) drives constraint ranking and the construction of local conjunctions, without regard to phonetic concerns.

Chapter 5 follows from the conclusions of chapter 4, and shows how grounded markedness constraints can give rise to seemingly unnatural phonological patterns. This chapter will propose grounded analyses of patterns in Kashaya, Zuni, and Eastern Ojibwa, which are claimed to be unnatural in Buckley (2000).

The rest of the introductory chapter discusses some conceptual groundwork, as well as some of the arguments against phonetic grounding that this dissertation seeks to address.

1.1 Tinbergen's Four Questions

This dissertation is commited to a biolinguistic approach to phonology, where the study of phonology is regarded as the study of an evolved biological system (Fitch, 2010). As such, the capacity to learn a phonological grammar may be regarded as a phenotypic trait. This entails not only that phonology has evolved in the history of our species, but also that it must have grown and developed during the maturation of any organism which possesses a phonology.

Naturally enough, fully 'explaining' a phenotypic trait is a complex and difficult task. Simply determining what does or does not constitute an explanation can be a source of major confusion. In an attempt to avoid any pitfalls in this regard, this thesis will adopt the Four Questions model, first proposed by Tinbergen (1963) as a model for the study of animal behaviour, and subsequently applied biolinguistics by Fitch (2010).

The model outlines four distinct questions or levels of explanation (see figure 1.1). Crucially, a full explanation of a phenotypic trait requires that *all* of the questions be answered. The four explanations do not compete in any sense, nor does any one take precedence over any other. Thus, the model may be regarded as a pluralistic approach.

Example 1 demonstrates four potential answers to the question "Why do birds sing?" using Tinbergen's model.

(1) Why do birds sing?

Mechanistic:
Vibration in the syrinx controlled by specific neural mechanisms, and activated at certain times of year by hormone levels.

Functional:
In order to attract a mate.

	Proximate	Ultimate
Synchronic	Mechanistic: *How does it work?*	Functional: *What is it for?*
Diachronic	Ontogenetic: *How did it grow?*	Phylogenetic: *How did it evolve?*

Figure 1.1: Tinbergen's Four Questions. The distinction between *synchronic* and *diachronic* refers to the fact that ontogeny and phylogeny involve the study of changes over time, whereas mechanism and function do not. The distinction between *proximate* and *ultimate* refers to the fact that mechanism and ontogeny can be understood as causative at the level of the individual organism, whereas function and phylogeny are 'deeper' causes that can apply to whole species or clades.

Phylogenetic:
Genetic mutations enabling birdsong arose in the evolutionary history of the species and were passed on to descendants.

Ontogenetic:
Production of subsong during a sensorimotor learning phase. Vocal learning via exposure to adult songs.

- Adapted from Fitch (2010)

Note the mutual interdependence of the four answers in 1. The functional explanation provides context for the mechanistic explanation, which depends on the sequence of events given in the ontogenetic explanation, which in turn depends on the events in the phylogenetic explanation.

Applying The Four Questions to Phonology

How do we apply the four questions to phonology? Below are some of the types of questions and potential answers that arise when applying the four questions to phonology. A coherent approach to phonology should seek to provide consistent answers to all four of the questions. An argument running throughout the whole of this dissertation is that a grounded or modally-dependent phonology provides coherent, mutually interdependent answers to these questions, whereas substance-free phonology does not.

Mechanistic Question: Synchronic theories of grammar are all answers to the Mechanistic Question. Note that here the term 'mechanistic' is not understood as referring to purely physical or mechanical explanations. Certainly, explanations of the physical structure of the brain (functioning of neurons etc.) do fall into the mechanistic category, but the category need not be restricted to physical explanations. Neuroscientific theories and biolinguistic theories of grammar are all ultimately attempts to explain how the brain works. Their loci of explanation simply exist at different levels of abstraction or granularity.

Functional Question: Not to be conflated with functional theories of grammar (which fall into the mechanistic category). An explanation of this type seeks to answer the questions: "What is the purpose of phonological computation?" or "Why does phonology exist at all?". The answer assumed in this thesis is that the phonology takes underlying representations supplied by the lexicon, and computes an optimal output based on considerations of articulatory ease, perceptual salience and computational efficiency. Chapter 3 will present an argument that modally-independent approaches have difficulty providing an answer to this question.

Phylogenetic Question: What are the genetic differences between humans and other apes such that we can learn and compute phonology and they can not? Since the phylogenetic history of language remains shrouded in mystery, this dissertation will not say much about the phylogeny. A full answer to the Phylogenetic Question will have to make use of molecular biology, comparative genomics and other fields.

Ontogenetic Question: Assuming that the brain is, in some sense, a type of computer there are two distinct questions that can be asked regarding ontogeny: Firstly, how is the computer built? And secondly, how does the computer learn from the data it is exposed to? The second question is a well established issue in linguistics, and is the focus of theories of language acquisition. The first question is largely ignored in generative theories of grammar. In this dissertation I refer to the first question as 'The Ontogenetic Question'. It may be stated in general terms as : How does the nervous system develop from a fertilised egg to an initial state, capable of learning a phonological grammar from primary linguistic data? While this question does go far outside the scope of the field of linguistics, into developmental neurobiology and the like, I will argue that different approaches to phonology do entail different predictions for the ontogeny of the phonological capacity. Moreover it is not beyond modern neuroscience to assess the likelihood of these different predictions, and as such the ontogenetic question is relevant to any biolinguistic theory of phonology.

The four questions paradigm is implicit throughout this dissertation. Distinguishing precisely which question is under examination at any given moment allows for a degree of conceptual clarity that can otherwise be lost. Distinguishing the four questions is especially useful when engaging in interdisciplinary work. When focused on the mechanistic question, the issue of relating work in linguistics and neuroscience is fraught with a number of issues (Poeppel & Embick, 2005). The question of exactly how brains can do the things linguists say they must be doing is a question that could easily remain unresolved for years to come. Chapter 2 of this dissertation will review neuroscientific research, but only with regard to the ontogenetic question, i.e. how do brains grow into what linguists say they must grow into. In many ways this is a much simpler question. While the issue of how brains compute is largely still open, the issue of how they grow is arguably much better understood. By focusing on the ontogenetic question, the application of neuroscience research to linguistics (and vice versa) becomes a lot less problematic.

1.2 Arguments to be Addressed

This dissertation seeks to address two main arguments against phonetic grounding. The first argument is that phonetic factors are redundant or irrelevant to phonological explanations. This argument is expounded at length in Hale and Reiss (2008). The second argument is that languages can exhibit unnatural phonological patterns, which defy reduction to phonetic factors. This argument is raised in Buckley (2000) and elsewhere.

This section will briefly recount these arguments, and lay the groundwork for the counterarguments proposed in later chapters.

1.2.1 On the Irrelevance of Phonetics

Hale and Reiss (2008) make the argument that a model of phonological competence should make no reference to phonetic factors, ergo phonology should be modally-independent. This argument has several prongs[1]:

1. If phonological primitives are genetically encoded then they should be explained by reference to the genome, not phonetics.

2. Phonetically natural patterns in phonology can be explained as a product of diachronic sound change.

3. The function of language is not communication.

The first point is exemplified by Hale and Reiss's claim that "*Contrary, perhaps, to our intuitive sense of these matters, a scientific theory which presents two radically different explanations (the human genome and phonetic difficulty) for a single phenomenon is not twice as good as one which presents a single coherent account for the phenomenon in question - it is, in fact, not nearly as good.*" (p. 151) Implicit in this notion that genetics and phonetics are different explanations is a kind of dichotomy. That phonology must be *either* encoded in the genome *or* motivated by phonetics. This is a false dichotomy that works by neglecting to define exactly what is meant by "encoded in the human genome" or give an ontogenetic account of how such encoding results in a functioning phonology. As chapter 2 will argue, it is unlikely that the neural substrates underlying phonology are constructed wholly separately from articulatory and perceptual organs, i.e. the basis of phonetics. The neuroscientific findings discussed in chapter 2 make clear that development of cortical tissue is frequently driven by external organs, making the hypothesis that the development of phonology is driven by articulatory and perceptual organs seem

[1]Note: There are some additional arguments made by Hale and Reiss against approaches which incorporate gestural or acoustic representations into the grammar. Since I am not advocating such an approach here, I will not address these arguments.

very probable. Under this view, genetics and phonetics are not two different explanations but simply different facets of the same explanation.

Regarding the notions of redundancy and irrelevance, given that the genome must somehow construct both these things (phonetics and phonology), it is far more parsimonious for the genome to derive one from the other, than it is for the genome to rely on two separate ontogenetic processes. Hale and Reiss's substance free phonology places a far greater burden on the genome, since it requires the genome to encode the phonology wholly separately from the phonetics. Thus it is a far less parsimonious hypothesis than phonetic grounding, an insight made possible by confronting the ontogenetic question.

The second point is also made in Buckley (2000), though it is not necessarily an argument against phonetic grounding. That diachronic processes, subject to extralinguistic factors, can influence the sound patterns of languages is a relatively uncontroversial claim. But even if this observation were enough to explain all the phonetically natural sound patterns in language, this would not render phonetic grounding redundant. As already noted, a modally-independent phonology would still be a far less plausible answer to the ontogenetic question. Even if shifting some of the burden of explanation to diachrony makes for a more elegant theory of phonology, there is no reason to suppose that this produces a biologically accurate account of how phonology works. If one is truly committed to a biolinguistic approach, then biological plausibility must take precedence over formal elegance.

The third point addresses Tinbergen's functional question. Questioning the function of phonology is an entirely worthwhile discussion to have. It is not *a priori* true that the function of phonology is articulatory and perceptual ease. However, Hale and Reiss fail to provide a coherent alternative. While the oft repeated claim that "language is not for communication" is entirely fruitful and sensible when applied to syntax and semantics, when applied to phonology it undermines some of the basic architectural assumptions of modern generative grammar. This argument will be taken up in chapter 3.

1.2.2 On the Unnaturalness of Phonology

The existence of seemingly unnatural patterns in phonology has long been established (e.g. Anderson, 1981). Buckley (2000) presents some examples of unnatural processes as evidence against phonetic grounding. However, the extent to which one considers this evidence against phonetic grounding depends on one understanding of what is meant by phonetic grounding. For example, Archangeli and Pulleyblank (1994, p. 177) outlines two possible claims of what is meant by phonetic grounding, the first of which would make unnatural patterns strictly impossible, while the second would make phonetically natural patterns more likely[2]. Certainly, the first claim is falsified by the existence of unnatural patterns. The second claim might seem more tenable, but as Buckley

[2] I haven't quoted Archangeli and Pulleyblank verbatim here as their exact definition depends on framework-internal terminology which would be meaningless in this context.

notes, even seemingly unnatural patterns can persist in a language for a very long time, and children apparently find these patterns no more difficult to acquire than phonetically natural patterns. Thus, even the second claim may be too strong.

The form of phonetic grounding argued for in this dissertation, is the claim that the primitives of phonology, i.e. that which is present in the initial state, are phonetically motivated. In OT terms, this amounts to the claim that features and markedness constraints should be phonetically grounded. Note that there is nothing in this claim which makes unnatural phonological patterns impossible. Even assuming grounded primitives, it is not beyond the capacity of the phonology to arrange those primitives to produce unnatural phonological patterns. Therefore the existence of unnatural patterns does not falsify this definition of a phonetically grounded phonology. Chapter 5 will give OT analyses of the unnatural phenomena highlighted in Buckley (2000) and show how they arise through the interaction of grounded constraints.

1.3 The Initial State: What it is and what it isn't

Applying Tinbergen's Ontogenetic Question to phonology depended on distinguishing two separate issues: How is the computer built? And how does the computer learn? Regarding phonology, the first issue is what I refer to as the Ontogenetic Question, and is a neurobiological issue. The second question is the issue of language learning.

Notice that the second question depends on the first. It is trivially true that a computer can only start learning once it has been built, not before. My understanding of the initial state in this dissertation follows from this simple observation. I define the initial state as being the point at which the computer is built, but before any process of learning has taken place. Crucially, this 'point' is not assumed to be a literal point in time, it is quite possible that in practice the processes of building and learning overlap and interact to some degree. Rather I take it to be an epistemological divide, i.e. building and learning are only understandable at different levels of abstraction or granularity; learning is a linguistic problem, while building is a neurobiological problem.

I take the linguistic level of analysis to be *epistemologically emergent*. In simplified terms, this can be understood as the claim that events at the linguistic level of analysis are, in some sense, 'brought about' by events at a lower level of analysis, e.g. the neurological level, which are themselves brought about by events at the biological level (and so on until one reaches sub-atomic physics). Each epistemologically emergent level of analysis is, in principle, reducible to a lower level, though such reduction is neither desirable nor possible in practice. Fodor (1974) presents the analogy of the 'immortal econophysicist' (p. 104), who seeks to derive laws of economics from physical laws and a total knowledge of the state of the universe. While it is intuitively true that such a thing is possible in principle, the incomprehensible complexity involved makes it extremely unlikely that such a feat could ever be accomplished. Similarly, while it seems intuitively true that a total knowledge of every neuron in the brain could, in principle, provide one with a total understanding of phonological computation, few would suggest this as a serious research strategy. Thus, the linguistic level of analysis can be understood as a set of simplified principles which emerge at a point when neurological principles become incomprehensibly complex. This type of emergence is called epistemological, because it represents limits on human knowledge of complex systems, rather than claims about metaphysics (O'Connor & Wong, 2012).

Therefore, the initial state is the point at which there exists a 'language computer' which permits examination at the linguistic level of analysis, i.e. there exists a certain degree and type of neurological complexity from which linguistic principles emerge. Thus, the initial state is better understood as an epistemological divide, rather than an exact point in time[3]. A consequence of this, is that linguistic theories are restricted to explaining events after the initial state. They cannot, in principle, explain how the initial state was built by the genome. This has certain implications for claims of 'innateness' made by linguists.

Innateness and the Initial State

Generative theories of language learning depend on the notion of an initial state, or S_0. The initial state can be understood as those computational primitives which exist independently of learning (i.e. exposure to PLD), and which are themselves necessary for learning to take place. This notion of an initial state follows from a well established philosophical argument, expounded in Fodor (1975). In short, the argument states that because primitives are a prerequisite for learning, the primitives themselves cannot be learned. On this point I assume myself to be in agreement with Hale and Reiss (2008, p. 27) and perhaps most generative phonologists.

Hale and Reiss make an additional claim, that the primitives present in S_0 must be innate, since they cannot be learned. Hale and Reiss refer to this position as the Innateness of Primitives Principle (IoPP). I will argue that the IoPP is not necessary, nor indeed likely. While it is true that the computational primitives that make learning possible must exist independently of any learning in the computational sense, it does not automatically follow that these primitives must be innate in the strict biological sense. The IoPP relies on a false dichotomy between 'innate' and 'learned'. Fodor himself rejects this dichotomy in Fodor (2008, ch. 5).

(2) [*The Language of Thought argument*] does, I think, show that there is something radically incoherent about the thesis that concepts can be learned. But [...] it doesn't quite follow that any concepts are innate [...] 'learned' and 'innate' don't

[3]Logically, there must be a time before the initial state exists and a time after, but attempting to pinpoint exactly when the initial state comes into existence in a child's brain is likely a folly on a par with the immortal econophysicist.

> exhaust the options; God only knows how many ways a creature's genetic endowment may interact with its experience to affect the inventory of concepts at the creature's disposal, but surely there are lots. (Fodor, 2008, p. 130)

Fodor uses *learning* to mean some quite specific, and draws a distinction between *acquisition* and *learning*. Acquisition is used to refer to any process which results in the attainment of a concept, while learning is a specific type of acquisition which involves a process of inductive inference.[4] Fodor argues that acquisition could include everything from surgical implantation to banging your head on a hard surface, as long as it results in the attainment of a concept. But none of these forms of acquisition depend on the existence of primitives, since they are not processes of inductive inference. Answering the 'how the computer is built' question likely involves processes of acquisition, but it cannot involve learning, in principle.

Using Fodor's terminology then, the Language of Though argument says that the initial state must contain those things which cannot be learned, *not* those things which cannot be acquired. Ergo it does not follow that everything in the initial state must be innate.

Ontogeny and the Initial State

This poses an obvious question: if the primitives in the initial state cannot be learned, then how are they acquired? Ultimately this question is the same as the Ontogenetic Question. It asks how we get from a fertilised egg to a set of neural substrates capable of learning a grammar. Figure 1.2 shows this question in the form of a diagram. The line marked with a '?' is the subject of the Ontogenetic Question, or the 'how the computer is built' question. The arrow to the right of S_0 is the 'how the computer learns' question, i.e. learning by exposure to PLD.

The error implicit in the IoPP is to treat the question-marked line in figure 1.2 as a type of implication: *If* it is in the initial state *then* it is in the genome. But this is an oversimplification. In fact, there are any number of potential hypotheses about what the question-marked line might represent, as there are any number of ways in which the genome might build a phonological computer.

Moreover, whether or not we suppose phonology is modally-independent makes different predictions about what types of ontogenetic processes should be taking place, to get us from a fertilised egg to the initial state. Examining these hypotheses, and their biological and neurological plausibility, is the subject of chapter 2 and forms the core of the argument for rejecting a view of phonology as being wholly modally-independent.

The modally-dependent phonology I advocate here, views the primitives of the initial state as being neither innate nor learned, but acquired during ontogeny by feedback mechanisms with the articulatory and perceptual organs.

[4]Note: Somewhat confusingly, what linguists generally refer to as language acquisition is not synonymous with acquisition in the sense used by Fodor. Language acquisition (in the linguists sense) is a type of learning (in Fodor's sense). From this point on I adhere to Fodor's terminology.

$$\overbrace{Genome}^{Phylogeny}\overbrace{\xRightarrow[?]{\qquad\qquad} S_0 \xRightarrow[\substack{learning \\ \Uparrow \\ PLD}]{\qquad\qquad}}^{Ontogeny}\begin{pmatrix} English \\ Japanese \\ \ldots \end{pmatrix}$$

Figure 1.2: The Ontogenetic Question: How does the nervous system develop from a fertilised egg to an initial state, capable of learning a phonological grammar from primary linguistic data?

1.4 Previous References to Ontogeny

The notion of ontogeny in relation to phonology is not new. An important predecessor to this dissertation is Bermudez-Otero and Börjars (2006) which proposes the idea of markedness constraints emerging ontogenetically, rather than being strictly innate. However there are some important conceptual and terminological differences between their paper and this dissertation. Crucially, Bermudez-Otero and Börjars offer the ontogenetic account as an alternative to the phylogenetic account:

(3) There are, in principle, two possible accounts of the origin of markedness constraints. In a phylogenetic account, constraints are supplied innately by UG. Subject to nonpathological maturation, constraints become available to the infant in the absence of all experience. In this context, explaining the nonarbitrary character of constraints is a task for evolutionary biology. In an ontogenetic account, in contrast, constraints arise developmentally through the dynamic interaction between the grammar and related performance systems. In this account, the emergence of constraints requires experience; the relevant input, however, does not consist of a corpus of utterances, but rather of the childs active manipulation of the cognitive and physiological systems with which the grammar interacts: e.g. the conceptual-intentional system, the parser, systems of motor control, etc. From this viewpoint, accounts of grounding are to be sought in developmental cognitive psychology. (Bermudez-Otero & Börjars, 2006)

This dissertation agrees that the 'innateness' often referred to by generative linguists (i.e. UG) is an appeal to a phylogenetic explanation for some aspect of the language faculty. However, it is a mistake to argue that an ontogenetic explanation and a phylogenetic explanation are competing hypotheses (see section 1.1). It is trivially true that every phenotypic trait an organism possesses must emerge during the growth and development of that organism, be it innate or otherwise. The only logical alternative, to argue that an organism can possess phenotypic traits which are present *before* the organism has grown, is either incoherent or an appeal to supernatural forces. Therefore every trait is ontogetically emergent, and every ontogenetic process has a phylogenetic origin. The only questions are the nature of the ontogenetic processes, and the relative influence of genetics and the environment.

This point is significant because it emphasizes that any account of the origins of phonological primitives must be ontogenetically plausible. Simply claiming that phonological primitives are innate does not circumvent this fact, because innately specified things still have to grow and develop somehow. This is central to the argument in chapter 2, which compares the implied ontogenies of modally-independent, and modally-dependent phonology, and finds the ontogenetic account of a modally-dependent phonology to be more plausible.

A second point at which this dissertation departs from Bermudez-Otero and Börjars (2006) is exemplified by the use of the word 'experience'. I would argue that reference to intentional vocabulary like 'experience' is an unnecessary complication. Fodor's learning/acquisition distinction (section 1.3) provides us with two distinct problems with two distinct levels of analysis. The first problem is language learning, i.e. induction over PLD, and linguistic theories are more than adequate to address this problem. The second problem is the development or acquisition of phonological primitives, which chapter 2 focuses on. This is a problem for developmental neurobiology and related fields. I would argue that there is no point at which we need to turn to an intentional, or 'psychological' level of explanation when addressing the second problem.

Abandoning intentional vocabulary, the two accounts Bermudez-Otero and Börjars give can be restated thus: In the first account, there is an ontogenetic process which constructs the neural substrates underlying phonology without any interaction with related performance systems. In the second account, the ontogenetic process constructs the neural substrates by interaction with related performance systems. In this dissertation, I take 'related performance systems' to mean first and foremost the articulatory and perceptual organs.

This dissertation agrees with Bermudez-Otero and Börjars that the second account is a better hypothesis, and that this entails some degree of phonetic grounding.

1.5 Optimality Theory

The purely linguistic portion of this dissertation, chapters 4 and 5, work within the framework of Optimality Theory (Prince & Smolensky, 1993). However, the dissertation itself is not meant as an argument for OT. In principle, the broader argument for modal-dependence should be applicable to any phonological framework.

Nonetheless, OT does lend itself easily to a modally-dependent approach. OT constraints are able to explicitly express principles of articulatory and perceptual ease, making it comparatively simple to integrate these principles into the vocabulary of the framework itself. Additionally, the violable nature of OT constraints makes it possible to construct a model of phonology which is phonetically grounded, but still capable of phonetically unnatural outputs. This is the theme of chapter 5.

Clearly, the goal of designing a phonological framework which is both consistent with all linguistic and neuroscientific data goes way beyond the scope of a master's dissertation. Therefore this dissertation adopts the goal of remaining as close to standard, parallel OT as possible, while positing a minimal number of additional assumptions to satisfy the criterion of modal-dependence. Specifically, this dissertation works under the assumption that all features and markedness constraints must be phonetically grounded, since these are present in the initial state and are argued to be the product of interaction between the articulatory and perceptual organs, and the areas of the brain responsible for phonological processing.

The criterion that markedness constraints must be phonetically grounded sets strict limits on the number and type of constraints that can be posited by phonologists. I take this to be a positive thing as, arguably, some degree of restriction on constraints is necessary for OT to have any explanatory power at all.

Additionally, the analyses in chapter 5 make use of constraint local conjunctions. Chapter 4 will sketch an account of how local conjunctions may be learned from PLD, and the following section will present a hypothesis of how local conjunctions may be implemented at the neurological level. Finally, the last section of this chapter will discuss the issue surrounding the typological implications of local conjunction.

1.5.1 Local Conjunction

A local conjunction between two constraints effectively creates a single complex constraint from two simpler constraints. The constraint conjunction is only violated when both its conjuncts are violated within the same domain (e.g. by the same segment). A more formal definition can be given as follows (Smolensky & Legendre, 2006b, p. 503):

(4) a. A constraint $\mathbb{C}$ in *Con* may be the **local conjunction** of two simpler constraints in *Con*, $\mathbb{A}$ and $\mathbb{B}$: if $\mathbb{C} = \mathbb{A} \,\&_{\mathcal{D}}\mathbb{B}$, then $\mathbb{C}$ is violated whenever $\mathbb{A}$ and $\mathbb{B}$ are both violated within a common domain $\mathcal{D}$.

b. $\mathbb{C}$ may be viewed as implementing the **conjunctive interaction** of $\mathbb{A}$ and $\mathbb{B}$.

c. Universally, the conjunction dominates its conjuncts:
$\mathbb{A} \,\&_{\mathcal{D}}\mathbb{B} \gg \{\mathbb{A}, \mathbb{B}\}$.

To see how this works, consider a hypothetical language which spirantizes voiced onstruents in coda position. To capture this is an OT tableau, we need a constraint which is only violated if a segment is both voiced obstruent and appears in a coda position. This is accomplished with the local conjunction of the Voiced Obstruent Principle and *Coda:

(5) a. VOP

Assign one violation mark for every voiced obstruent present in the output.

b. *Coda

Assign one violation mark for every output segment in a coda position.

c. VOP&$_\varphi$*Coda

Assign one violation mark for every voiced obstruent in a coda position present in the output.

(6)

/bad/			VOP&$_\varphi$*Coda	Ident([voice])	VOP	*Coda
a.	☞	baz			*	*
b.		bad	*!		**	*
c.		bas		*!	*	*
d.		bat		*!	*	*

In 6, the input /baz/ cannot be realized as-is because the segment [z] is both voiced and a continuant. Note that although candidate a. violates both *Voice and * Continuant, the violations do not co-occur in the same segment, so the local conjunction, whose domain is φ, is not violated.

1.5.2 Local Conjunction as Neural Gating

At the abstract level of discrete symbolic computation, local conjunctions can appear somewhat arbitrary. The formalism of OT by itself does not predict that constraints should be able to 'team up' and exert a greater influence. Indeed, it has even been argued that local conjunctions are a violation of OT's principle of strict dominance, and an unnecessary complication (e.g. Padgett, 2002).

However, it should be remembered that OT is designed as a high-level representation of the harmony maximization taking place in a neural network. As such it is one level of a split level architecture (Smolensky & Legendre, 2006b). Concepts at one level of analysis do not always translate neatly to concepts at another level. Indeed, at the sub-symbolic level, the principle of strict domination can appear to be an arbitrary stipulation, despite its obvious elegance and utility at high levels of analysis.

This section will present the hypothesis that local conjunctions, despite their alleged inelegance at the OT level of analysis, are a result of a simple fact about the functioning of neurons. Specifically, that they are a result of the gating effect of threshold potentials in cell membranes.

How Neurons Fire

A neuron's ability to fire, the *action potential*, is made possible by maintaining a difference in the polarity of the intracellular and extracellular fluids. This difference, the *resting potential* is typically around -70mv (Freberg, 2010). At the synapses, an excitatory signal from a presynaptic neuron will cause an exchange of ions between the intracellular and extracellular fluid, resulting in a depolarization of the postsynaptic neuron. If this depolarization is sufficient, it triggers a chain reaction in the exchange of ions, causing a voltage spike in the postsynaptic neuron. This spike is the action potential, which is propagated down the length of the axon. In most neurons in the nervous system, the action potential is always the same voltage or 'strength'. Thus an action potential is a discrete, or digital signal. The continuous values used in connectionist networks represent the variable firing rates of neurons and the strength of synaptic connections, not the strength of the action potential itself.

Neuronal Gating

Not all presynaptic excitatory potentials will result in an action potential in the postsynaptic neuron. To trigger an action potential, the cell must be depolarized past a certain point, the *threshold potential*, which is typically around -65mv (ibid). If a presynaptic excitatory potential only depolarizes the postsynaptic neuron by, say, 3mv (from -70mv to -67mv), then no postsynaptic action potential will occur.

The threshold potential gives rise to a peculiar property, namely the ability for a neuron to function like an AND-gate. If we imagine two excitatory presynaptic neurons, each of which depolarize the postsynaptic cell by 3mv, then the postsynaptic neuron will only fire when both presynaptic neurons are firing simultaneously. If we designate the two presynaptic neurons as A and B, and the postsynaptic neuron as P, then this relationship could be expressed as $A \wedge B \rightarrow P$. If A **and** B fire, then P will fire. This phenomenon is referred to as *gating* and has been posited as mechanism for controlling the flow of information between areas of the brain (Gisiger & Boukadoum, 2011).

Local Conjunctions

In principle, gating could be a means of producing local conjunction type effects. OT constraints represent the weighting of connections between units (i.e. synapses between neurons) in a neural network. The extent to which a given representation is well-formed depends on the extent to which the representation deviates from the weight matrix of the network. Thus, a network computes a well-formed representation simply by maximizing the harmony in the network. In OT, the higher ranked a constraint is, the higher the disharmony in the network when that constraint is violated (Smolensky & Legendre, 2006b, ch. 4).

A local conjunction between two constraints is always ranked higher than its individual conjuncts. In terms of harmony, this means that the total disharmony of violating two conjoined constraints simultaneously must be greater than the sum of the disharmony of violating both constraints individually. This point can be made somewhat clearer by using mathematical notation. Assuming x and

y are arbitrary values expressing the disharmony of violating the constraints $\mathbb{C}$ and $\mathbb{D}$, respectively:

$$\mathbb{C} = x$$
$$\mathbb{D} = y$$
$$\mathbb{C}\&\mathbb{D} > x + y$$

This is why violating a local conjunction of two constraints is worse than violating the same two constraints not in a conjunction. The local conjunction implies there is some additional disharmony incurred when both conjuncts are violated simultaneously. The question then is where does this additional disharmony come from? Neuronal gating provides a potential answer.

Suppose, within a network, we have a postsynaptic neuron P receiving input from two presynaptic neurons (figure 1.3). The synapses, A and B are weighted such that P will only fire if A and B both depolarize the cell simultaneously. This a neuronal AND-gate. Now suppose that the synapses A and B are are connected to the rest of the network such that A is only active when the constraint $\mathbb{C}$ is violated, and B is only active when $\mathbb{D}$ is violated. Because of the neuronal gate, P will only fire when $\mathbb{C}$ and $\mathbb{D}$ are both violated simultaneously. As long as the result of P firing is less harmony in the network, the effect here is a local conjunction of $\mathbb{C}$ and $\mathbb{D}$.

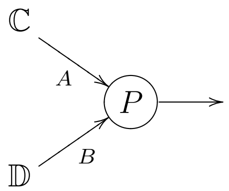


Figure 1.3: $\mathbb{C}\&\mathbb{D}$

Of course, this example is greatly oversimplified. Real neurons typically have thousands of synapses, not two. Plus, this is nothing more than a hypothesis. I know of no empirical evidence that this is actually what is taking place, and I suspect it will be beyond our means to adequately test the hypothesis for quite some time.

Despite this, the hypothesis shows something quite interesting, namely a way to implement local conjunctions without any additional machinery or complexity. It simply falls out of a basic fact about the functioning of neurons, the threshold potential. Constructing a local conjunction, at the level of the brain, could be accomplished by nothing more than the strengthening or weakening or synaptic connections. Since this is the exact same mechanism that allows constraints to be reranked during language learning, local conjunctions can be thought of as simply another type of constraint ranking. For this reason, chapter

4 proposes that local conjunctions are learned by induction over PLD. A simple mechanism is proposed which functions as an addendum to the constraint demotion algorithm of Tesar and Smolensky (2000), allowing a local conjunction to be constructed when the algorithm encounters a ranking paradox.

1.5.3 Typological Implications of Local Conjunction

Local Conjunction has been criticized for predicting implausible phonological processes, particularly when one allows for domains larger than the segment (syllable, p-word, etc.). For example, McCarthy (2002) gives the example of ID([BACK])$\&_\sigma$VOP which would cause the devoicing of obstruents before high vowels. However, I do not consider this to be a particularly strong argument against local conjunction, for two reasons.

Firstly, the judgment that a particular phonological pattern is implausible can never be based on anything other than phonologists' intuitions. This is a simple fact of the Induction Problem. Only a small percentage of the world's languages have been studied by phonologists. A pattern that is unattested in the languages studied so far could still exist among the languages which have not been studied. Moreover, even if a particular pattern were found to be completely unattested in the world's languages, this would not prove that the pattern is strictly impossible. The current state of the world's languages is as much a product of historical accident as anything else. There is every reason to suppose that phonological processes which are currently unattested are nonetheless entirely possible. If history had taken a different course, we might be studying those processes instead, and our intuitions about what is phonologically implausible might be wholly different. In principle, a typological argument can never tell us what is an impossible phonological process.

Secondly, languages *do* exhibit seemingly implausible phonological processes. Chapter 5 gives analyses of a number of these. McCarthy's example of obstruent devoicing before high vowels is certainly odd, but it is no more odd than the backing/rounding process in Kashaya or the palatalization patterns in Odawa. Were it not for the fact these processes are attested, we might think them implausible, yet there they are. With this in mind, the ability of local conjunctions to deal with crazy or unnatural processes may be considered an advantage rather than a disadvantage.

In the analyses in chapter 2, local conjunctions are an essential tool in demonstrating how a phonetically grounded phonology can produce the phonetically unnatural patterns attested in the world's languages. The combination of phonetically grounded constraints with local conjunction makes for an approach which is conceptually coherent, neurobiologically plausible, and capable of explaining the empirical data.

Chapter 2

Confronting the Ontogenetic Question

From the biolinguistic perspective, the study of phonology must ultimately be the study of a biological system. As such, fully explaining this system entails answering all of Tinbergen's Four Questions (section 1.1). While much work done by phonologists contributes to answering the mechanistic question, the other three remain relatively unexamined by linguists. A central claim of this dissertation is that confronting the ontogenetic question provides phonologists with insights which are relevant to the assumptions implicit in our synchronic theories of grammar. This claim can be split into two separate claims, which I will address in order:

1. Modally-independent and modally-dependent approaches to phonology make different predictions regarding the ontogeny of the phonological capacity.

2. The predictions made by a modally-dependent approach to phonology better fit the findings by neuroscientists regarding the growth and development of the nervous system.

Addressing the first claim means determining how these two approaches to phonology differ in their predictions regarding the ontogeny of phonology. Since the ontogenetic question is not addressed directly by Hale and Reiss (2008), or any other substance-free phonologists as far as I'm aware, the predictions of a modally-independent phonology are not made explicit and thus have to be deduced from other arguments.

The ontogentic predictions that I claim are made by modally-dependent phonology are based on my own understanding of grounded phonology. While I believe they are in accordance with similar views previously expressed in the literature (e.g. Bermudez-Otero & Börjars, 2006), I am not claiming that the ontogentic account given here is implicit in anyone's arguments except my own.

2.1 Two Ways Genes Can Build a Phonology

I have previously defined the ontogenetic question of phonology as: "How does the nervous system develop from a fertilised egg to an initial state, capable of learning a phonological grammar from primary linguistic data?". This question was represented in the diagram in figure 1.2 as the arrow with a question mark.

However we seek to answer this question, the genes must surely play a key role. However, there are presumably a multitude of ways in which the genes could accomplish such a task. I argue that modally-dependent and modally-independent phonology correspond to two broadly different ways that genes could construct the neural substrates underlying phonology. The key to distinguishing these two lies in the assumptions that modally-dependent and modally-independent phonology make about what is present in the initial state.

A modally-dependent approach to phonology, in the form I argue here, assumes that the initial state contains primitives acquired by some interaction with the perceptual and articulatory organs. In other words, the phonology develops under the influence of these external organs.

Conversely, a modally-independent approach to phonology assumes that the primitives in the initial state are substance free, and therefore must develop independently from the external organs. This would appear to be in line with Hale and Reiss's claim that the primitives in the initial state are innate (see section 1.3). It is also reflected in a thought experiment pursued by Hale and Reiss:

(7) We would like to present now a rather unsettling thought experiment. We apologize for its mildly disgusting nature. Imagine that, because of a nuclear power plant accident, a genetically transformed human-like species came into being. These humanoids are physiologically identical to humans in every way except the following: in their necks at the base of their oral cavities they have a thin but strong membrane which expands under sufficient pressure, not unlike the croaking membrane of the rana catesbeiana, or common bullfrog. In particular, imagine that the degree of pressure required to expand this membrane is sufficiently small that the well-known aerodynamic problems which arise in producing voiced stops (e.g.) disappear: voicing throughout the duration of the stop closure is no problem for these near-humans. (Hale & Reiss, 2008, p. 154)

Hale and Reiss dub this new species *homo collitumens*. The physiology of *homo collitumens* is such that there is no longer any articulatory basis for voicing constraints such as VOP or *VOICED-CODA. The question then is what would happen if *homo collitumens* were exposed to a language like German, which relies on both these constraints? Hale and Reiss examine two possible outcomes:

1. Assuming constraints are innate; the constraints VOP and *VOICED-CODA are specified by UG and thus still present in the grammars of *homo collitumens*, allowing them to learn German just as easily as a *homo sapiens*.
2. Assuming constraints are phonetically grounded; the constraints VOP and *VOICED-CODA cannot be derived from phonetic difficulty, but could be induced from PLD (as suggested by Hayes (1999)), allowing *homo colliumens* to learn German.

Hale and Reiss offer this thought experiment as an argument that the phonetic naturalness of constraints is irrelevant since we get the same outcome regardless. Leaving aside this issue for now[1], the argument reveals implicit assumptions about the ontogeny of innate constraints. If #1 were true, we could infer that whatever gene(s) resulted in *homo collitumens*' special membranes has no functional importance in the development of the neural substrates that underlie phonological processing. Ergo the neural substrates grow independently from the articulatory organs.

Since this is Hale and Reiss's definition of innateness, and they have argued that the primitives present in the initial state are innate, it follows that Hale and Reiss's definition of the initial state must be things which develop independently from the perceptual and articulatory organs. This appears to be further reinforced by Hale and Reiss's definition of UG:

(8) ... *by definition*, UG, the innate component of the language faculty, consists of the elements of linguistic representation which cannot be derived from anything else. (Hale & Reiss, 2008, p. 38)

I assume the "anything else" here would include things such as articulatory and perceptual factors. Thus, according to Hale and Reiss's definition, the genes must build the neural substrates underlying phonology wholly separately from the articulatory and perceptual organs.

In summary, a modally-dependent approach to phonology predicts that the development of the initial state, i.e. the capacity to learn grammar, depends on,

[1] In fact, the outcome would be quite different according to the version of grounded phonology I argue for here. As we shall see chapter 4 specifically rejects the creation of constraints by induction over PLD, except in the limited case of local conjunction creation, meaning *homo collitumens* would have no way of creating VOP or *VOICED-CODA. This would render German phonology strictly unlearnable by *homo collitumens*, except by some extralinguistic process of generalisation or the memorization of large numbers of allomorphs in the case of [rat]:[radəs] type alternations.

or somehow interacts with the development of the articulatory and perceptual organs. A modally-independent phonology predicts that the development of this capacity should happen independently from the articulatory and perceptual organs.

Returning again to the diagram from figure 1.2 (reprinted below as figure 2.1), we now have two ways of filling in the gap between the genome and the initial state. Figure 2.2 and figure 2.3 show the two ontogenetic accounts as diagrams. These diagrams can be thought of as an expansion of the question-marked arrow in figure 2.1.

It should be noted that these diagrams are conceptual aids and grossly oversimplify what is likely taking place in reality. Each arrow is likely a complex chain of events involving genetic cascades, synaptic pruning/strengthening and any other number of factors. They are however sufficient to illuminate a key difference between the predicted ontogenies of modally-dependent and modally-independent approaches to phonology, namely, whether or not the articulatory and perceptual organs play any role in the development of the initial state.

This leads us to the next claim, that the predictions of a modally-dependent approach to phonology are more plausible. The remainder of this chapter is dedicated to arguing this point.

$$\mathit{Genome} \xRightarrow[?]{\qquad\qquad} S_0 \xRightarrow[\substack{learning\\ \Uparrow\\ PLD}]{\qquad\qquad} \begin{pmatrix} \mathit{English} \\ \mathit{Japanese} \\ \dots \end{pmatrix}$$

Figure 2.1: The Ontogenetic Question: How does the nervous system develop from a fertilised egg to an initial state, capable of learning a phonological grammar from primary linguistic data?

$$\begin{tikzcd}
 & S_0 \mathit{Phonology} \\
\mathit{Genome} \arrow[ur, Rightarrow] \arrow[r, Rightarrow] \arrow[dr, Rightarrow] & \mathit{Articulatory\ Organs} \\
 & \mathit{Perceptual\ Organs}
\end{tikzcd}$$

Figure 2.2: Ontogeny of a Modally-Independent Phonology. No interaction between the development of the initial state and the external organs.

$$\begin{tikzcd}
\mathit{Genome} \arrow[r, Rightarrow] \arrow[dr, Rightarrow] \arrow[rr, Rightarrow, bend left] & \mathit{Articulatory\ Organs} \arrow[r, Rightarrow] & S_0 \mathit{Phonology} \\
 & \mathit{Perceptual\ Organs} \arrow[ur, Rightarrow] &
\end{tikzcd}$$

Figure 2.3: Ontogeny of a Modally-Dependent Phonology. The initial state develops under the influence of the external organs. The top arrow emphasizes the point that it is likely not *only* input from the external organs which influence the development of the phonology. There are almost certainly other factors as well.

2.2 Innateness in the Visual System

In the 1960s and 70s, David Hubel and Torsten Wiesel performed a series of experiments, measuring the activity of individual neurons in the primary visual cortex (henceforth V1) of kittens and macaque monkeys. The experiments were designed to test the extent to which the development of V1 is dependent on signals from the retina. This was accomplished by artificially restricting signals from the retina shortly after birth (e.g. by suturing one eye shut) and then comparing the resultant neural activity to that of a normal animal. As we shall see, their results indicate that the development of V1 is heavily dependent on signals from the retina during a critical period. I argue that this is analogous to the hypothesis of a modally-dependent phonology, in which the acquisition of the initial state would depend on signals from the articulatory and perceptual organs.

Note that this argument does not claim that language *learning*, i.e. induction over PLD, has any immediate analogy in the visual system. It merely claims an analogy between the ontogeny of the visual system and the ontogenetic of the initial state. The key point is whether or not there are signals being sent from external organs to areas of the brain. Therefore exposure to light should not be considered analogous to exposure to PLD, rather exposure to light should be considered analogous to babbling.

The following sections give a brief overview of the methods and results of Wiesel and Hubel (1963).

Method

The experiments described in Wiesel and Hubel (1963) used seven kittens and one adult cat. The animals had their vision in one eye deprived either by suturing the eye shut or by covering one eye with a translucent eye cover. The deprivation lasted for 1-4 months, after which the closed eye was opened and the response of neurons in V1 to visual stimuli was recorded. Four of the kittens had one eye deprived shortly after birth (from the time when the eye would normally open). The remaining three had some visual experience prior to deprivation.

After the period of deprivation, the responses of individual neurons to visual stimuli were recorded using electrodes. Stimuli were produced using a photo-stimulator held 6 inches from the animals' eyes. During stimulation, one eye would be covered with a patch of thick black rubber, allowing the experimenters to record the neural response to one eye at a time.

Results

Figure 2.4 shows the typical response of neurons in V1 of normal cats which have not been subject to any visual deprivation. Notice that most neurons responded to both eyes, although the contralateral eye appears to be slightly dominant. This histogram is the baseline to which the activation in visually deprived cats should be compared.

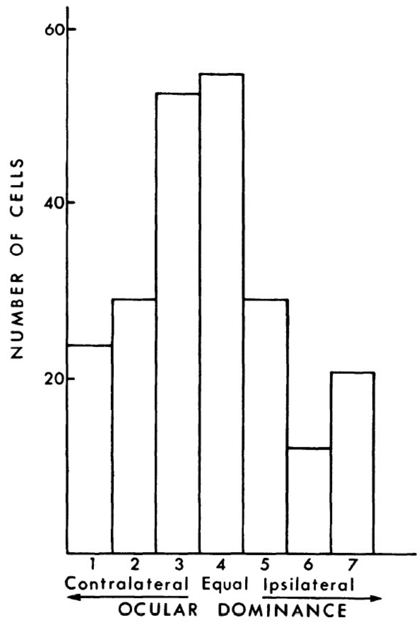


Figure 2.4: From Wiesel and Hubel (1963). Showing the response to visual stimuli of 223 cells recorded from the striate cortex of adult cats. The x-axis shows occular dominance in a discrete scale from contralateral dominance (driven by eye on opposite side) to ipsilateral dominance (driven by eye on same side). Group 1 is those cells which were only driven by the contralateral eye. Group 7 is those cells which were only driven by the ipsilateral eye. Group 4 is those cells which were driven equally by both eyes.

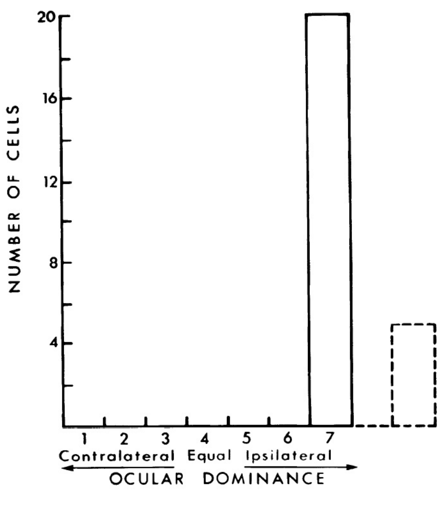


Figure 2.5: From Wiesel and Hubel (1963). Responses of 25 neurons in left cortex of 2.5 month old kitten whose right eye was sutured shut eight days after birth. Recordings taken once sutured eye was reopened. Of the 25 neurons recorded, 20 were only driven by the ipsilateral (open) eye, while 5 could not be driven by either eye (dotted lines). None of the neurons were driven by the contralateral (closed) eye.

Figure 2.5 shows the responses of 25 neurons in the left visual cortex of a 2.5 month old kitten after having its right eye sutured shut eight days after birth. Of the 25 neurons recorded, 20 were only driven by the eye which had remained open, and the remaining 5 could not be activated by either eye. None of the neurons recorded responded to stimulation of the eye which had been sutured shut. A similar response pattern was observed in kitten which had one eye covered by a translucent eye cover (see Wiesel and Hubel (1963) figure 6). For those kittens who had some visual experience prior to deprivation, the effects of the deprivation were less pronounced but nonetheless clearly evident (see Wiesel and Hubel (1963) figure 9).

Disucussion

The fact that the patterns of activation in V1 are so profoundly influenced by visual deprivation would seem to indicate that the development of V1 is heavily dependent on signals from the lateral geniculate nucleus and subsequently the retina. In visually deprived kittens, the total absence of activation driven by the closed eye indicates that those neurons which would normally develop to receive signals from the closed eye can be 'co-opted' and taken over by the open eye. The presence or absence of signals from the retina appears to be key. In effect, we can say that V1 develops into what input signals tell it to develop into. This makes an innate mechanism for specifying the exact wiring patterns in V1 redundant.

Note that innately specified structures do play a key role in this process. Hubel and Wiesel (1963) note that a great deal of the visual architecture appears to be present from birth, including the optic nerve, the geniculate receptive fields, the connections between the geniculate nuclei and V1, as well as the distinct, column-like organization of the visual cortex itself. All of these things appear to develop independently of visual experience. Despite this there is no genetically endowed wiring diagram for every synapse in V1.

2.2.1 Relevance to Phonology

The development of V1 is clearly analogous to proposed ontogeny of a modally-dependent phonology (shown in figure 2.3). The development of the cortical tissue depends on signals from external organs via innately specified pathways, rather an innate wiring diagram. The obvious question now is whether or not the evidence from the study of V1 is relevant to phonology. Certainly, it is not *a priori* true that the phonology must develop in a similar fashion to V1. However there are good reasons to think that it probably does.

Firstly, there is the issue of the 'critical period', which is common to both language learning and the development of V1. Subsequent experiments have shown that the effects of visual deprivation cannot be reversed once the animal has passed beyond the critical period (Wiesel & Hubel, 1965). The existence of a critical period for language learning is well established. This parallel hints at some fundamental similarity in the development of these two systems.

Secondly, like V1, phonological processing is likely localized somewhere in the cerebral cortex (e.g. the superior temporal sulcus (Hickok & Poeppel, 2007)). Thus we should expect that the phonology and V1 consist of broadly similar types of neurons, connected in a similar fashion to the rest of the central nervous system, and which have all undergone the same processes of migration, pruning and apoptosis (Freberg, 2010, ch. 5). In the absence of strong evidence to the contrary, there is every reason to suppose that different areas of the cortex develop along similar lines.

In principle, if it were discovered that phonological processing were taking place somewhere other than the cerebral cortex, then we could reasonably disregard the data on the growth of the visual system. Subcortical models of language have been proposed, for example Lieberman (2007) which proposes that the basal ganglia are central to language processing. However, nothing about Lieberman's model would support the notion of a modally-independent phonology. On the contrary, Lieberman's model proposes that grammatical competence is a product of the evolution of motor control. If anything, this model would predict that theories of phonology should be *more* phonetically motivated than even the proposal here.

Either way, the evidence so far does not come down in favour of a modally-independent phonology.

2.3 Cross-Modal Plasticity

While the experiments of Hubel and Wiesel discussed in section 2.2 give us some indication of the nature of cortical plasticity, the plasticity under examination is very limited. Specifically, it is plasticity within a single modality. In fact, cortical tissue can exhibit a far greater degree of plasticity under the right circumstances. In cases where one modality is entirely absent during the development of the cortex, such as congenital blindness, areas of the brain normally associated with the missing modality can adapt to process information from entirely different modalities. This phenomenon is particularly relevant to the sign-phonology argument, which claims that similarities between sign-phonology and spoken phonology are evidence of the modal-independence of phonology. This section will argue that this argument is incorrect, because sign-phonology can just as easily be understood as an example of cross-modal plasticity.

2.3.1 Cross-Modal Plasticity in Occipital Areas

The conclusion of the experimental data in section 2.2 was that V1 only develops into a visual centre because it receives signals from the retina. An obvious question then, is what happens to V1 in people who are born blind? In such cases V1 receives no signals from the retina at all, therefore we would not expect V1 to develop into a visual centre at all, but something else entirely. The truth of the matter is particularly interesting to linguists. In the congenitally blind, occipital areas are involved in language processing:

(9) In congenitally blind adults, occipital areas show a response profile that is similar to classic language areas: During sentence comprehension, occipital regions are sensitive to high-level linguistic information such as the meanings of words and compositional sentence structure [...] Like classic language areas, occipital regions respond most to sentences, less to lists of words and to Jabberwocky (non-word sentences), and still less to lists of nonsense words and backwards speech [...] There is also some evidence that occipital activation is functionally relevant to language processing: rTMS to the occipital pole leads to verb-generation errors and impairs Braille reading in congenitally blind individuals. (Bedny, Pascual-Leone, Dravida, & Saxe, 2011, p. 1)

Bedny et al. (2011) used neuroimaging techniques to compare the blood flow in the occipital areas of congenitally blind and late blind adults, when performing sentence comprehensions tasks. This allowed them determine whether or not the co-opting of occipital areas for language processing is dependent on a critical period, early in life. Their method and results are reviewed and discussed below.

Method

The study was conducted on a total of 42 participants. Of those, 22 were sighted, 11 were congenitally blind and 9 were late blind.

The participants heard short passages of English speech and were then required to answer true/false question. Unfortunately, Bedny et al. do not note whether or not the participants were native speakers of English.

As a control condition, participants heard two passages of speech played backwards, a long passage and a short passage. The participants then had to determine whether the short passage was a part of the longer passage or a novel string. The backwards speech is assumed to lack any English syntactic, semantic or phonological information, and therefore should not trigger any response in the language processing areas of the brain.

MRI scans of the whole brain were taken while the participants performed the tasks. This allowed the experimenters to monitor the flow of blood in the occipital areas.

Results

All three groups had comparable accuracy when performing the tasks. The linguistic tasks were accomplished with 82-85% accuracy, while the backwards speech tasks were accomplished with 52-56% accuracy (see Bedny et al. (2011) table 2).

The congenitally blind participants had a notably higher response in V1 during the story and questions portion of the trial than during the backwards speech portion. The difference in response was greater in the left hemisphere than in the right. No such difference was recorded in the V1 of late blind and sighted participants, who exhibited a similar response in V1 during all portions of the trial, and showed no evidence of lateralization.

Congenitally blind participants had a similar response in classic language areas to late blind and sighted participants.

Discussion

The high response in V1 of congenitally blind participants during linguistic tasks is strong evidence that V1 is involved in the language processing in the congenitally blind. This tells us that, in the absence of signals from the retina, V1 can be co-opted by non-visual modalities. The fact that there is no similar response in the late blind participants implies that this co-opting can only take place during a critical period, after which cross-modal plasticity is presumably not possible.

Clearly, this a much greater degree of plasticity than was observed in section 2.2. However it is entirely in keeping with Wiesel and Hubel's findings that the development of V1 is not determined by a genetic wiring diagram but rather by signals received from other areas of the central nervous system.

2.3.2 Relevance to Sign-Phonology

The similarities between sign-phonology and spoken phonology have been cited as evidence that phonology is modally independent. If the same phonology can work across multiple modalities, it is argued, then the phonology must be modally independent. However, the existence of cross-modal plasticity presents a serious problem for this argument. If a region of the brain so obviously modally-dependent as the visual system can be co-opted by other modalities in the congenitally blind, then it would seem entirely plausible that a modally-dependent phonology could be co-opted by other modalities in the deaf.

But is there any evidence that sign-phonology is an instance of cross-modal plasticity? Bedny et al. (2011) discovered a clear difference in the V1 response between the congenitally blind and the late blind/sighted, which presents a means of determining when an area of the brain has been co-opted cross-modally. In the congenitally blind, V1 was active during language processing, and thus this area had been co-opted. In the late blind/sighted, no such activation was recorded, and thus V1 had *not* been co-opted cross-modally.

By way of analogy then, if sign-phonology is an instance of cross-modal plasticity, we should see a clear difference in the activation patterns of congenitally deaf signers from those of hearing signers. Ideally, we should see a difference in those areas of the brain normally associated with phonology.

Such differences may well have already been discovered. MacSweeney et al. (2002) used fMRI to compare the responses of congenitally deaf signers to hearing signers, while performing sentence acceptability tasks in British Sign Language. They discovered a greater response to in the superior temporal gyri (STG) of congenitally deaf signers compared to hearing signers. This would seem to indicate some degree of cross-modal plasticity in the absence of auditory input.

The remaining question is whether or not the STG is associated with phonological processing. Certainly, this area is normally associated with auditory processing. Hickok and Poeppel (2007) propose the STG is involved in spectrotemporal analysis, which is clearly not a form of modally-independent processing. However, Hickok and Poeppel propose that phonological level processing itself takes place in the nearby superior temporal sulcus (STS).

Whatever the exact distribution of labour between the STG and the STS, there is a clear difference in the activation patterns of congenitally deaf signers and hearing signers, and this difference is localized to areas of the brain associated with modally-dependent aspects of language. It was precisely the difference between the congenitally blind and late-blind that formed the evidence for cross-modal plasticity in Bedny et al. (2011). Thus there is undeniably some degree of cross-modal plasticity in congenitally deaf signers. The remaining question is the extent to which this plasticity correlates with the observable differences and similarities between sign-phonology and spoken phonology. Whatever the answer, it us currently untenable to claim that any similarities between sign-phonology and spoken phonology *must* be evidence of modal independence in phonology, since the cross-modal plasticity hypothesis is at least an equally tenable hypothesis.

A Note on Hearing Signers

Despite lacking the activation in the STG indicative of cross-modal plasticity, hearing individuals are nonetheless fully capable of acquiring a sign-phonological grammar. Does this refute the claim that similarities between sign-phonology are an instance of cross-modal plasticity? Not necessarily. Firstly it should be noted that hearing individuals are on average less competent than deaf individuals at signing. MacSweeney et al. report that the hearing signers performed the sentence identification task less accurately than deaf signers.[2]

Additionally, it should be pointed out that hearing signers have succeeded in learning a language created, maintained and propagated chiefly by deaf individuals. Humans are nothing if not expert pattern learners. It is entirely conceivable that hearing individuals learning sign may be able to overcome a handicap, stemming from lack of STG recruitment, by means of some alternate mechanism, for example, by general purpose learning mechanisms or by some plasticity elsewhere in the brain driven by exposure to PLD.

[2]MacSweeney et al. point out that this disparity may be in part due to the fact that hearing children of deaf parents may interact with their parents differently to deaf children of deaf parents. Additionally, hearing signers may use sign language less frequently in adult life compared to deaf signers. Exactly how important these factors might be remains an open question.

Because these languages are created by deaf signers, the structure of sign-phonology should be regarded as indicative of the linguistic capacity of the congenitally deaf. There is no evidence that a community of hearing individuals could spontaneously create and use a sign language with the complexity and richness of British Sign Language. Indeed, in this case, absence of evidence may well be evidence of absence, given that we have the recorded histories of hearing communities going back literally thousands of years, and have not one instance of a hearing community ever developing a full sign language. While functional arguments have been offered to explain the prevalence spoken language over sign, such as the ability to communicate in the dark, these arguments don't hold up to much scrutiny. It is just as easy to think up functional advantages for sign-language, such as the ability to communicate in silence (Fitch, 2010, p. 442).

Moreover, if modal-independence were true, then there would be no clear reason for people to restrict themselves to either spoken language or sign-language. If the whole of grammar were modally-independent, then using the same grammar in multiple modalities should be a comparatively simple task. There could easily be a signed *and* a spoken version of English, which speakers could switch between at will. And yet all natural languages, both signed and spoken, appear to be confined to a single modality[3].

We can follow this line of reasoning even further. It is quite easy to conceive of a language which utilises multiple modalities simultaneously. For example, lexical items could be expressed with speech, while tense and aspect could be expressed using sign. As long as the grammar itself were modally-independent, nothing would seem to forbid this. And yet nothing like this appears to exist. Despite the fact that people instinctively gesticulate when speaking, this gesticulation never seems to exhibit any kind of grammar comparable to sign language. This should be deeply puzzling to anyone advocating a modally-independent phonology.

If we assume a modally-dependent phonology however, the puzzle disappears. Even assuming a modally-independent syntax/semantics[4], if the phonological component of grammar is modally-dependent, then there would be no easy way of expressing the same language in different modalities, or using multiple modalities in the same language. A different modality would entail a different phonological grammar, an entirely different proposition from simply transducing the same grammar between different modalities.

2.4 Pluripotentiality

The findings discussed in sections 2.2 and 2.3 present an understanding of cortical development in which areas of the cortex adapt to a particular task, not because of an innately specified wiring diagram, but in response to signals from

[3] With the obvious exception of written language, which is hardly a natural language phenomenon in the same sense as sign language and spoken language.

[4] Which I do.

other areas of the central nervous system. But this does not entail that cortical tissue is a kind of blank slate. Certainly, there are limits to cortical plasticity. Rather we can think of cortical tissue as having the potential to adapt to a relatively limited number of roles, depending on the input it receives. This view has been termed *pluripotentiality*:

(10) Although few neurobiologists would argue in favor of *equipotentiality*, i.e., the idea that all areas of cortex are created equal [. . .], there is now overwhelming evidence in favor of *pluripotentiality*, i.e., the idea that cortical tissue is capable of taking on a wide array of representations, with varying degrees of success, depending on the timing, nature and extent of the input to which that tissue is exposed. (Bates, 1999, p. 9)

Pluripotentiality represents a kind of middle ground between blank slate *equipotentially* and rigid *innate predetermination* (ibid). This view is supported, not only by the findings discussed in sections 2.2 and 2.3, but also by lesion studies. Specifically, patients ability to recover from seemingly irreparable brain damage implies that damaged cognitive faculties must be able to reemerge elsewhere in the brain, a fact that is irreconcilable with innate predetermination. Bates (1999) gives an overview of lesion studies, as an argument for pluripotentiality. While it would be redundant to recount the whole of Bates (1999) here, a key discovery regarding lesion recovery bears repeating in the context of the current discussiony.

Lack of Long-Term Symptom/Lesion Correlations

Bates notes that lesions incurred early in life in specific locations can correlate with a specific loss of function in the language faculty. For example, damage in the frontal lobes between 19-31 months may result in a delay in word production. However, this particular loss of function does not persist later in life. After 5-7 years of age, while the child may show impairment in global measures like IQ, the correlation between the location of the lesion and the symptoms has disappeared:

(11) To summarize, our prospective studies of language development in children with early focal brain injury have provided evidence for specific delays, correlated with specific lesion sites. However, the nature of these lesion/symptom correlations differs markedly from those that we would expect based on the adult aphasia literature. Furthermore, these correlations are only observed within specific windows of development, followed by evidence for recovery and (by implication) reorganization. (Bates, 1999, p. 7)

The conclusion is that specific functions which are knocked out by a lesion early in life can later reemerge elsewhere in the cortex, at the expense of non-specific competence like IQ. If we accept that the development of cognitive functions involves pluripotential cortical tissue whose development is driven by input from external organs, it is fairly clear how and why this happens. In the case of phonology, even if an area of the cortex normally dedicated to phonology were knocked out by a lesion early in life, the signals into the cortex from the articulatory and perceptual organs would still persist. And these signals could still influence the development of other areas of the cortex, thereby allowing phonology to reemerge elsewhere.

Conversely, if phonology were predetermined by some innate wiring diagram, it is not at all clear how this reemergence could take place. If genes really did build cortical structure on a synapse-by-synapse basis, much as genes regulate the growth of limbs and organs, then damaging cortical tissue should be much the same as cutting of a finger. We wouldn't expect a replacement to grow of its own accord, either in the same place or somewhere else entirely. It would seem the only way an innately predetermined cerebral cortex could recover from damage is if there were some kind of 'executive function' which oversaw the whole development, and responded to damage in one area by reassigning tissue in another area. Aside from the fact that it is not at all clear where and what such an executive function could be, and how it could work, it is a deeply unparsimonious hypothesis when compared to the hypothesis of pluripotential tissue.

Moreover, the pluripotential account of lesion recovery appears to be far more congruous with the known biology of neuronal growth and development. It is a well established fact that the development of the nervous system appears to depend on the production of *too many* neurons and synapses, which are subsequently lost to cell death and synaptic pruning (Freberg, 2010, ch. 5). It is easy to see how this excess of neurons and synaptic connections could give rise to pluripotential tissue. The sheer number of potential circuits contained in young cortical tissue could, potentially, perform any number of different tasks. Input to the tissue from other areas determines which of these potential circuits are used, and thus win the ensuing competition for the neurotrophins that stave off cell death (ibid), resulting in specialization of the tissue. This notion of competition among neurons is well established in the literature, and is expounded in Edelman's theory of Neural Darwinism (Edelman, 1993).

If we were to assume the existence of an innate wiring diagram, it is not clear why this proliferation of redundant neurons and synapses should take place at all. Nor indeed, how it relates to the supposed executive function required to account for neural plasticity if one assumes the existence of innate wiring diagrams.

Phonology and Pluripotentiality

Bates (1999) does not discuss phonology directly. However the findings from lesion studies do support a modally-dependent approach to phonology. The hypothesis that the initial state of the phonology develops under the influence of the articulatory and perceptual organs closely parallels the notion of pluripotential cortical tissue developing according the nature and timing of the signals it receives.

A modally-independent approach to phonology would seem to depend on some form of innate prespecification, a hypothesis which is unable to account for the reemergence of damaged faculties on other areas of the cortex. Certainly, Hale and Reiss's Innateness of Primitives Principle appears to implicitly assume some form of innate prespecification (see section 1.3).

In principle, it might be possible to give an ontogenetic account of a modally-independent phonology which does not rely on innate specification. If one supposed that the signals that drive cortical tissue toward phonological processing come not from the articulatory and perceptual organs, but instead from some other area of the nervous system, one could perhaps argue that a modally-independent phonology might arise from pluripotential tissue. This is a rather vague hypothesis, but it would appear to be the most obvious line of inquiry open to phonologists wishing to advocate a modally-independent phonology.

2.5 Conclusion

This chapter began with two claims:

1. Modally-independent and modally-dependent approaches to phonology make different predictions regarding the ontogeny of the phonological capacity.
2. The predictions made by a modally-dependent approach to phonology better fit the findings by neuroscientists regarding the growth and development of the nervous system.

If both these claims are true, then Hale and Reiss's argument (section 1.2) is moot. It could well be true that, from the mechanistic perspective, a modally-independent phonological module would be far more elegant than a modally-dependent phonology which relied on the redundant inclusion of phonetic factors. But if there is no way for such a thing to emerge during ontogeny, then its hypothetical elegance is wholly irrelevant to the issue of whether or not it actually exists.

If one is committed to the study of language as a biological system, then this is simply a principle to which one is beholden. For example, biologists have long observed the lack of mammals with green fur. The absence is somewhat puzzling, when one considers the abundance of green flora, it would seem that green fur could provide a significant survival advantage in the form of effective camouflage. And yet apparently no mammal has evolved to fill this gap. The reaction of biologists is not to argue that, in fact, mammals *are* green[5], but instead to explain this gap between ideal design and biological reality.

It could well be that phonologists find themselves in a similar situation. The findings discussed in this chapter suggest that, however elegant and well

[5](at some underlying level of representation)

designed a modally-independent phonology might be, it is simply unlikely that such a thing could grow in the cerebral cortex. The evidence suggests that the development of complex cognitive capacities in the cortex is driven by signals from the external organs, influencing pluripotential tissue. In many ways this is a clever trick by mother nature. The human genome consists of only 20-25,000 genes, yet the adult nervous system consists of trillions of synapses (Freberg, 2010). It is only the fact that the genome can build cortical structures indirectly, without an innate wiring diagram, that makes this seem possible.

Given that the phonology has to interact with the articulatory and perceptual organs, it is entirely reasonable to assume that the development of the phonology in the cortex is driven by the interaction these organs, much as the development of V1 is driven by interaction with the eyes. This entails a modally-dependent phonology. If we replaced the articulatory and perceptual organs with other organs in a different modality, as in the case of sign-phonology, we would expect a somewhat different kind of phonology to emerge.

Conversely, a modally-independent phonology seems to imply some form of innate predetermination. This leaves substance-free phonologists with a number of unresolved issues: the vast gulf between the number of genes and synapses, how plasticity is possible, and why phonology should be so fundamentally different to other, older, more complex cognitives systems like the visual system.

The findings discussed in this chapter are not conclusive. There is still a great deal that remains unknown about the ontogeny of the phonological capacity. Nonetheless, confronting the ontogenetic question makes apparent that the hypothesis of a modally-independent phonology, at least as expounded by Hale and Reiss (2008), entails assumptions which seem implausible and which raise more issues than they resolve. Thus it falls to advocates of modal-independence to provide an answer to the ontogenetic question congruent with the facts, which could allow for a modally-independent phonology.

Chapter 3

The Function of Phonology

This chapter will discuss some outstanding issues relating to Tinbergen's Functional Question, in the context of arguments raised by Hale and Reiss (2008). This chapter will argue that, while it is not *a priori* true that the function of phonology is articulatory and perceptual ease, any bioloinguistic study of phonology must posit some function for phonology.

Function and Dysfunction

Hale and Reiss (2008, section 7.8) present an argument against functionalism in phonology. 'Functionalism' here can be understood as the claim that the purpose or goal of phonology is articulatory and perceptual ease. Or, in more general terms, to minimize effort while making things easier for the listener. Hale and Reiss point out that functional approaches to phonology typically assume that the phonology a struggle or balance between these two seemingly opposing forces. This is certainly true of the approach advocated in this dissertation.

Hale and Reiss argue against functionalism on the grounds that the functional principles of articulatory and perceptual ease could be replaced with their dysfunctional opposites (i.e. maximize effort and minimize perceptibility) without affecting the set of predicted grammars. This would imply that functional principles have no explanatory power. However, there are two flaws in this argument.

Firstly, there exist phonological patterns which serve both articulatory and perceptual ease. The Agree constraints used in the Zuni analysis in chapter 5 are an example of this. Vowel epenthesis would be another example. If a vowel is inserted between two stops it has the benefit of breaking up the consonant cluster, making the stops easier to pronounce. But it also creates a space for the formants which provide information about the place of articulation of the first stop, making the stop easier to perceive. Assuming functional principles, patterns like this are to be expected, since it satisfies both articulatory and perceptual ease. If we were to assume 'dysfunctional' principles, then phenomena like vowel epenthesis should be strictly forbidden, since it violates both the principle of maximizing effort and the principle of minimizing perceptibility. Therefore the set of grammars predicted by functional and dysfunctional principles are not the same, and there is good reason to think that the world's languages belong to the set predicted by functional principles.

Secondly, the functional principles of articulatory and perceptual ease are not principles specific to phonology *per se*. Nor are they, as Hale and Reiss say, the claim that "*human beings are fundamentally lazy, but helpful*"(Hale & Reiss, 2008, p. 185). Rather they are simply general principles of good design, i.e. that a machine or system should perform its job as efficiently and effectively as possible. We would expect these principles to be evident in any evolved system. Thus, even if functionalism and dysfunctionalism did produce the same predictions, functionalism would still retain a degree of explanatory power which dysfunctionalism lacked, simply because functionalism unifies linguistic phenomena with general biological principles.

Of course, evolution frequently fails to produce good design. A prime example is the giraffe's laryngeal nerve, which traverses the entire length of a giraffe's neck twice, despite only needing to cover a distance of a few inches (Dawkins, 2009). Cases like the laryngeal nerve are interesting precisely because they violate principles of good design, which tells biologists something about their evolutionary history. They are not, however, evidence that evolution specifically produces systems which are as ineffective and inefficient as possible. Such a conclusion would be quite absurd, and wholly at odds with the notion of adaptation by natural selection, which assume that adaptations are selected on the basis of the advantage they provide. Naturally, this tends toward efficiency and effectiveness. If phonology does contain elements of bad design, then these should be interesting to phonologists for the same reason that the laryngeal nerve is interesting to biologists. At no point however, would bad design be an argument for dysfunctionalism.

Phonology as an Epiphenomenon

Functionalist principles in phonology follow from general principles of good design, that a system should do its job as effectively and efficiently as possible. The assumption here is that phonology is an evolved system whose 'job' has something to do with speech and communication. This point is an assumption, not an *a priori* truth. It is, in principle, entirely conceivable that phonology evolved for some purpose other than speech and communication. This would entail a very different answer to Tinbergen's functional question.

Note however that rejecting the functionalist principles of articulatory and perceptual ease does not entail that phonology is entirely function*less*. The claim that phonology has no function whatsoever is the claim that phonology is simply a *spandrel*, a peculiar accident or evolutionary byproduct of something else (Fitch, 2010). While this is certainly a logical possibility, it is not a very interesting (or likely) possibility to phonologists; substance free, functionalist or otherwise. The claim that phonology is a spandrel would be incompatible with the claim that phonology is a distinct module of the mind, complete with its own specification in the genome. Phonology would simply be an epiphenomenon of something else.

Interestingly, Hale and Reiss express concern that the 'phonology as an epiphenomenon' view results from functionalist principles:

(12) Given a sufficiently rich and explicit theory of the human personality (giving us principles such as "be lazy" and "be helpful to the listener"), the human articulatory and perceptual systems ("phonetic" substance), phonology itself will turn out to be epiphenomenal. While this seems considerably less promising to us, it has clear implications for the research strategy which phonologists should adopt. Phonologists, under such a view, should focus their energies in two domains: phonetics and the empirical explication of fundamental features of the human personality ("laziness", "helpfulness", etc.). (Hale & Reiss, 2008, p. 185)

This line of reasoning assumes that the functional principles of articulatory and perceptual ease are products of human psychology. But, as I have argued, these need not be claims about human psychology but rather general principles of design, a motivating factor in all evolved biological systems. While many phonologists might despair at the prospect of attempting to reduce phonology to psychology[1], any phonologist who is committed to the biolinguistic perspective should be committed to the unification of linguistic phenomena with biology.

Within the biolinguistic perspective, avoiding the 'phonology as an epiphenomenon' view means positing a plausible answer to Tinbergen's functional question, i.e. one must define what the job of phonology is. The claim that the job of phonology has something to do with speech and communication is one such answer. This answer, combined with general principles of good design, give us the functionalist principles of articulatory and perceptual ease. Other answers are possible, though Hale and Reiss do not provide such an answer, leaving the function of a modally-independent phonology somewhat unclear.

Generative Phonology as Externalization

While Hale and Reiss do not posit a clear function for phonology, they do relate their 'anti-functionalist' stance to a long-standing claim in generative linguistics, namely the claim that language *per se* is not an adaptation for communication. This claim goes at least as far back as Chomsky (1966/2009). Under this view, language *qua* language exists as a tool for thought, with communication being regarded as a later adaptation. While this is a plausible argument for syntax and semantics, attempting to define phonology as a tool for thought poses a

[1] And rightly so.

$$
\begin{tikzcd}
 & DS \arrow[d] & \\
 & SS \arrow[dl] \arrow[dr] & \\
PF & & LF
\end{tikzcd}
$$

Figure 3.1: Standard Theory involving Deep Structure, Surface Structure and the Phonetic Form and Logical Form interfaces. More recent versions of the model vary significantly, but retain the distinction between the phonological/phonetic branch and the semantic branch (Boeckx & Uriagereka, 2007).

number of conceptual problems, and even contradicts some basic assumptions in generative linguistics about the architecture of the language faculty.

Generative approaches to language generally assume some separation between the semantic interface and the phonological interface (e.g. Chomsky, 1965, 1995; Boeckx & Uriagereka, 2007). This separation if often shown in the form of an inverted-Y diagram, shown in figure 3.1. There are good reasons for this separation. The spoken form of a sentence is a linear, ordered string, while the syntax/semantics appears to be non-linear and hierarchical. Moreover, by allowing the semantics to interpret what is supplied by the syntax, rather than the phonological/phonetic form, we allow for significant variation between the spoken form of a sentence and its semantic interpretation. This would seem to be a necessary condition to account for ambiguity in the case of sentences like "I saw a man with a telescope", in which the same phonological form has multiple semantic interpretations.

A consequence of this architecture is that phonological processes have no influence on semantic interpretation, and thus no effect on the interface with conceptual/intentional systems. As a result, the 'language as a tool for thought' hypothesis only makes sense when applied to the syntax/semantics part of the derivation, since these are the only parts of language which interact with 'thought' in some broader sense.

Upon reflection, this conclusion is quite sensible. While the benefit that syntactic recursion might provide to thought or general reasoning is fairly obvious, it is not at all obvious what benefit could be provided by phonological processes like spirantization, schwa epenthesis or stress assignment. And even if we could come up with some explanation of how stress assignment functions as a tool for thought (which seems dubious), we would still be left with the issue of how phonological processes can interact with conceptual/intentional systems, while still maintaining a distinction between the phonetic and logical forms of a sentence.

The simplest answer to these problems, is simply to posit that phonology exists for the purpose of the externalization of the internal structures produced by syntax. This is the answer which is most congruous with the standard assumptions in generative grammar, and the simplest answer to Tinbergen's functional question.

Phonology as a Mathematical Game

Hale and Reiss (2008) do not advocate explicitly that phonology should be regarded as a tool for thought (though it is a plausible implication of their rejection that phonology is a tool for communication). Instead they advocate that the study of language be regarded as a "mathematical game" (p. 186). While there is nothing inherently wrong with this approach, it is deeply antithetical to the notion that the study of language, and hence phonology, is the study of a biological system. Studying a biological system entails answering Tinbergen's Four Questions. Any biological system has evolved (phylogeny), grown (ontogeny), and works (mechanism) to accomplish some particular task (function). Simply studying how a system works is meaningless without some understanding of what task the system is performing.

By way of analogy, consider an organ such as the heart. Certainly, we could give a complete mechanistic description of the heart's workings without ever referencing the notion of function. We could imagine a rich and detailed account of every muscle fiber, nerve and blood vessel, as well as accurate measurements and models of its performance. But without reference to the fact that the heart exists *for pumping blood*, none of these descriptions would constitute an explanation. It is only by defining the function of the organ that we can understand how the organ's physical structure and design work to accomplish this function. Positing an answer to Tinbergen's Functional Question drives research in the Mechanistic Question, because it is only by positing *what* the organ does that we can begin to explain *how* the organ does it. Phonology should prove no exception here. Theories of grammar are an attempt to explain how the phonology does what it does. To do this we first need some notion of what the phonology is doing. We need to define its function.

If the claim that the task of phonology is maximizing articulatory and perceptual ease is rejected, advocates of substance free or modally-independent phonology still face the question of what the function of phonology is. Why is there a phonology at all? Why aren't lexical items just fed directly to the phonetics? Why bother with all that extra computation? These are not peripheral issues, these questions are the context that make the study of phonology meaningful. A grounded, modally-dependent approach to phonology provides clear answer to these questions. In the case of substance free phonology, the answers are considerably less clear.

Chapter 4

Learning a Phonological Grammar

4.1 The Origins of Features and Constraints

Any computational theory of learning must assume some set of computational primitives which are themselves unlearnable. In phonology, these primitives compose the initial state as argued for in section 1.3. This dissertation argues that phonological primitives are acquired during ontogeny by some interaction between articulatory and perceptual organs, and the areas of the cerebral cortex associated with phonological processing.

If we accept this argument, we are still left with the question of precisely what the phonological primitives in the initial state are. Standard OT proposes that the primitives are a universal set of features and constraints (Prince & Smolensky, 1993). It follows that if constraints and features arise through some interaction with articulatory and perceptual organs, then constraints and features should reflect articulatory and perceptual concerns, i.e. they should be phonetically grounded. The analysis in chapter 5 will work from this assumption.

However the standard OT assumption is not the only plausible theory of what is present in the initial state. Archangeli, Mohanan, and Pulleyblank (2009) and subsequently Archangeli and Pulleyblank (2012) propose that features are not themselves universal but learned via exposure to PLD. This entails that constraints are also not in the initial state but must also be learned by induction (e.g. Hayes, 1999). This proposal does not circumvent the argument that primitives are unlearnable, implicitly it must assume a more fundamental set of primitives which guide the induction of features. In principle this proposal in entirely compatible with the arguments from chapters 1 and 2, provided the primitives in question arise through interaction with the articulatory and perceptual organs.

The following section give a brief discussion of the 'learned features' and standard OT proposals, and justify the decision to adopt the standard OT proposal for this dissertation.

4.1.1 Learnable Features and Constraints?

Archangeli and Pulleyblank (2012) propose that phonological features, and subsequently constraints, can be learned by induction over PLD. Under this view features and constraints would not be part of the initial state. Archangeli and Pulleyblank label this a form of *Emergent Grammar*, a proposed alternative to the concept of UG.

Despite this, nothing about the proposal refutes Fodor's arguments regarding computational learning (section 1.3), Emergent Grammar must still assume something like an initial state. Archangeli and Pulleyblank argue that featural categories are generalizations driven by the data, that is, the child is exposed to data and forms symbolic categories by inducing patterns in the data. But for this inductive process to take place, there must still be some pre-existing notion of what to look for in the data. In practice, induction cannot work over an infinite hypothesis space. Thus, there have to be some limitations to the hypotheses that can be considered, even if the only limitations are apparently obvious or trivial, such as the limitations imposed by the articulatory and perceptual organs.

Archangeli and Pulleyblank propose that categories can arise by identifying acoustic and articulatory properties in the data, such as short vs. long VOT, lips vs tongue etc. But the ability to identify long and short VOT still assumes a pre-existing concept of VOT. At the very least, the learner must be able to associate the patterns in the data with their own knowledge of the kinds of sounds they can make with their articulatory organs. Such knowledge is, in principle, as much a theory of the initial state as the standard OT assumptions.

Of course, concepts such as VOT make ideal candidates for the kinds of things that could arise by some interaction between articulatory and perceptual organs with puripotential tissue in the cortex. In principle then, the Emergent Grammar approach would seem to be seem to be compatible with the neuroscientific literature discussed in chapter 2.

However, there are some potential issues with Emergent Grammar regarding OT and explanatory adequacy. OT seeks to explain the differences between languages as different rankings of a set of universal constraints (Prince & Smolensky, 1993). But according to Emergent Grammar, features and constraints are language specific, not universal. If this were the case then the differences between languages cannot be explained as simply differences in constraint ranking. In fact, one might even go so far as to say that if Emergent Grammar is true, then positing rankings of violable constraints is wholly redundant. One could simply attempt to explain differences between languages as differences in their features, plus whatever rules or declarative constraints could be induced from PLD.

This approach could well be a worthwhile research program, but since this dissertation uses an OT framework, it is not a research program I will pursue here. Instead I will adopt the standard OT assumption that different grammars are simply different rankings of the same universal constraints, with the additional caveat that all features and markedness constraints should be phonetically grounded, owing to the assumption that they have arisen through some interaction with the articulatory and perceptual organs.

Standard OT and Constraint Induction

Nowhere in this dissertation do I propose anything like a formal theory of how grounded constraints are acquired. This is because the standard OT assumption, that constraints and features are present in the initial state, is in principle incompatible with theories of constraint induction (e.g. Hayes, 1999). Anything present in the initial state is not learned as it is a prerequisite for any inductive process of learning. Under the standard OT assumption, the origin of constraints is strictly a neurobiological problem, not a linguistic one. It is a problem of 'How the computer is built' not 'How the computer learns' (chapter 1).

This does not mean that the problem of how primitive constraints arise is explained away or swept under the rug, rather it means that linguistic theories alone cannot answer this problem. A thorough answer will depend on collaboration with biologists and neuroscientists. It will require an extension of the kind of research examined in chapter 2.

4.1.2 What is a Grounded Constraint?

The argument that constraints should be grounded depends on some coherent understanding of what is meant by a grounded constraint. My understanding here is quite simple: All other things being equal, a candidate that violates a grounded constraint should be more articulatorially or perceptually difficult than a candidate which does not.

Of course, this raises a number of other questions, such as what is meant by articulatory and perceptual difficulty. To some extent this an open question since it depends on our understanding of articulatory and acoustic phonetics, which is very much an ongoing field of study. Nonetheless it is still possible to have sensible conversations about these subjects with our current level of understanding.

It should also be pointed out that I only advocate that markedness constraints be phonetically grounded, not faithfulness constraints. There are reasons for this. Firstly, in the approach advocated here, articulatory and perceptual difficulty define markedness. That's just what markedness is. Secondly, faithfulness constraints are not conceptually difficult or puzzling in the same way as markedness constraints. Faithfulness, i.e. the tendency to *not* change, is really just a local instance of Newton's First Law. It is the claim that things don't happen unless something causes them to happen. It is not so much a fact about language as a fact about the universe we live in. Faithfulness constraints are simply a way of capturing this fact within the formalism of OT.

There is no scientific problem to be answered regarding their origins since it is literally impossible for anything in the universe not to obey this principle[1]. Consequently, faithfulness constraints are not said to arise through any interaction with articulatory and perceptual organs, and thus are not phonetically grounded.

The Limits of Phonetic Grounding

Finally, it should be noted that the claim that all markedness constraints should be grounded in phonetics is likely far too strong a claim in practice. The phonetic grounding claim is a conclusion that follows from the hypothesis that the development of the initial state is driven by signals from external organs. But it is entirely plausible that ontogeny of phonology depends on other factors than simply the external organs. For example, phonology must presumably develop in the cortex alongside other cognitive faculties, including a syntax module. It is quite conceivable that these areas interact and influence one another as they develop, which could allow for the possibility of markedness constraints which would reflect facts about the phonology-syntax interface rather than the phonology-phonetics interface. In addition, we could imagine there might be constraints which simply represent facts about neural computation such as computational efficiency.

However, it would be far more problematic to judge how well grounded constraints of this sort are. What would make a good syntactically grounded constraint? This is an entire discussion in itself. Since none of the analyses in chapter 5 depend on any sort of grounding other than phonetic grounding, this dissertation simply adopts phonetic grounding as its stated goal.

Moreover, phonetic grounding makes for a very clear research strategy, since it posits a single cause for many phenomena. Thus the goal of the researcher is to attempt to reduce as many phenomena to this single cause as possible. This principle is a cornerstone of the scientific method. For this reason, I propose that under a modally-dependent approach, speculations about other sorts of grounding should be saved until one encounters problems which seem wholly irreducible to phonetically grounded constraints. The analyses of unnatural patterns in chapter 5 never reach this point, which would seem to suggest that a great deal can be explained with phonetically grounded constraints alone.

[1] Excepting perhaps the weird and wonderful world of quantum mechanics.

4.2 Constraint Ranking

In OT, languages vary from one another in their individual rankings of constraints. Given a set of universal constraints present in the initial state, learning a phonological grammar must involve a process of ranking these universal constraints, on the basis of the PLD the learner is exposed to. Naturally, OT phonologists bear the burden of explaining exactly how this takes place.

One explanation for how constraints are ranked during learning is constraint demotion, proposed in Tesar and Smolensky (2000). This dissertation adopts constraint demotion as an explanation of language learning.

Tesar and Smolensky assume constraints are initially unranked and subsequently ranked by a process of demotion, informed by PLD. Demotion operates by comparing the 'correct' form of a word, provided by PLD, with competing incorrect forms, provided by *Gen*:

(13) Each piece of positive evidence, a grammatical structural description, brings with it a body of implicit negative evidence in the form of the competing descriptions. Given access to *Gen* (which is universal) and the underlying form (contained in the given structural description), the learner has access to these competitors. Any competing candidate, along with the grammatical structure, determines a data pair related to the correct ranking: the correct ranking must make the grammatical structure more harmonic than the ungrammatical competitor. (Tesar & Smolensky, 2000, p. 33)

This process can be demonstrated by taking a relatively simple phonological problem, such as coda devoicing in Dutch. Explaining this patern in OT requires only three constraints (Kager, 1999):

(14) a. IDENT([VOICE])

Assign one violation mark for every output vowel that differs from its input correspondent in the feature [voice].

b. *VOICED-CODA

Assign one violation mark for every voiced segment in a coda position.

c. MAX-IO

Assign one violation mark for every input segment which does not have a corresponding output segment.

These three constraints will produce coda devoicing. As shown in the tableau below with the derivation of [bɛt] from /bɛd/ 'bed':

(15) /bɛd/ → [bɛt] 'bed'

/bɛd/			*Voiced-Coda	Max-IO	Ident([voice])
a.		bɛd	*!		
b.	☞	bɛt			*
c.		pɛt			**!
d.		bɛ		*!	

But how does the learner arrive at the correct ranking? In the initial state, all the constraints are assumed to be unranked. The learner constructs winner/loser pairs by pairing the correct form provided by PLD with a loser candidate provided by *Gen*. This is shown in the tableau below:

(16) bɛt - 'bed'

win/lose pairs			Ident([voice])	*Voiced-Coda	Max-IO
a.	✓	bɛt	✱		
b.		bɛd		*	
a.	✓	bɛt	✱		
c.		pɛt	**		
a.	✓	bɛt	✱		
d.		bɛ			*

The ✓ symbol indicates the winner candidate i.e. the candidate that should be the most harmonic in the final constraint ranking. The ✱ symbol indicates violations by the winner candidate, which should not be fatal in the final ranking.

Marks are canceled when both the winner and the loser violate a given constraint. Looking at the winner/loser pairs, we can see that both a. and c. violate Ident([voice]). So, the violation mark of candidate a. is canceled.

Having assessed the winner/loser pairs, constraint demotion then proceeds by the following principle:

(17) For any constraint ℂ assessing an uncanceled winner mark, if ℂ is not dominated by a constraint assessing an uncanceled loser mark, demote ℂ to immediately below the highest ranked constraint assessing an uncanceled loser mark. (Tesar & Smolensky, 2000, p. 36)

Note that only one of the violation marks for candidate a. was canceled. There are still two uncanceled marks, one for each of the remaining winner/loser pairs. This means that Ident([voice]) still assess two uncanceled winner marks, which are not dominated by any constraints assessing an uncanceled loser mark. So, Ident([voice]) is demoted, creating a new stratum in the hierarchy below *Voiced-Coda and Max-IO.

In a more complex grammar, the algorithm would proceed like this in a cyclic fashion, until there are no more undominated winner marks. However, in this case, the algorithm has already arrived at the target grammar, i.e. the ranking in 15. Thus, by using constraint demotion, all the non-fatal violations are now ranked lower than the fatal ones.

4.3 Local Conjunctions: Induction over Primary Linguistic Data

I have proposed that the only markedness constraints present in the initial state are those constraints which are phonetically natural. This entails that local conjunctions, which are capable of producing phonetically unnatural patterns, are not present in the initial state and must be learned by exposure to PLD. However, the constraint demotion algorithm has no way of producing local conjunctions.

In this section I propose a process by which local conjunctions can be created, in cases where the demotion algorithm is unable to arrive at a stable ranking. This process is conceived as a kind of 'subroutine' of the demotion algorithm, which is invoked whenever constraint demotion fails.

To demonstrate how this process works, let us return to the example given in section 1.5 of how local conjunction could give rise to the spirantization of voiced obstruents in codas. The tableau below is a reprint of the one in section 1.5:

(18) /bad/ → [baz]

/bad/			VOP&$_{\varphi}$*Coda	Ident([voice])	VOP	*Coda
a.	☞	baz			*	*
b.		bad	*!		**	*
c.		bas		*!	*	*
d.		bat		*!	*	*

The local conjunction VOP&$_{\varphi}$*Coda correctly produces the spirantization of voiced obstruents in coda positions. Since this conjunction must be learned, only the individual conjuncts can be present in the initial state, not the local conjunction itself. Thus, the demotion algorithm must begin by attempting to rank the constraints in their unconjoined form. The tableau in 19 shows the constraints assessing winner/loser pairs:

(19) winner/loser pairs

win/lose pairs			Ident([voice])	VOP	*Coda
a.	✓	baz		✱	✱
b.		bad		**	*
a.	✓	baz		✱	✱
c.		ba		*	
a.	✓	baz		✱	✱
d.		bat	*	*	*
a.	✓	baz		✱	✱
e.		bas	*	*	*
a.	✓	baz		✱	✱
f.		zaz			*

The constraint demotion algorithm is unable to arrive at the correct ranking in this case. The algorithm proceeds by demoting constraints assessing uncanceled winner marks until they are ranked lower than constraints assessing uncanceled loser marks. Since both VOP and *CODA assess uncanceled winner marks, these two constraint would be ranked lower than IDENT([VOICE]) which assesses uncanceled loser marks. However this ranking would still not produce the correct output. At this point the algorithm is stuck, and the phonology must turn to local conjunction.

The algorithm for conjunction learning I propose has two parts: conjunction creation and conjunction selection.

Conjunction Creation

First, the phonology proliferates a large number of possible constraint conjunctions, by the following principle:

(20) For every two constraints, $\mathbb{C}$ and $\mathbb{D}$, which are violated by a winning candidate, create one local conjunction $\mathbb{C}\&_{\mathcal{D}}\mathbb{D}$ for every possible domain $\mathcal{D}$.

This will produce all possible logical combinations of constraints which are violated by a winning candidate, in every possible conjunction domain (segment, syllable etc). In the case of the constraints in 19, this will produce the conjunction VOP$\&_{\varphi}$*CODA, as well as the same conjunction in different domains e.g. VOP$\&_{\sigma}$*CODA, and all the possible conjunctions of violated constraints which are not listed in 19 but nonetheless assumed to be present in the initial state e.g. VOP$\&_{\varphi}$IDENT([CONT]), *CODA$\&_{\varphi}$IDENT([CONT]) etc.

Constraint Selection

Having produced a number of potential conjunctions, the correct conjunction is selected by checking all the potential conjunctions against the winner/loser pairs:

(21) winner/loser pairs

win/lose pairs			VOP&$_\varphi$*CODA	VOP&$_\sigma$*CODA	*CODA&$_\varphi$IDENT([CONT])	VOP&$_\varphi$IDENT([CONT])
a.	✓	baz		✱	✱	
b.		bad	*	*		
a.	✓	baz		✱	✱	
c.		ba				
a.	✓	baz		✱	✱	
d.		bat		*		
a.	✓	baz		✱	✱	
e.		bas		*	*	
a.	✓	baz		✱	✱	
f.		zaz			*	

Any conjunction which assesses an unchecked loser mark is selected. The rest are rejected. In 21, VOP&$_\varphi$*CODA is the only conjunction assessing an unchecked loser mark (pair a. & b.), so this conjunction would be fed back into the constraint demotion algorithm. With the local conjunction in place, the demotion algorithm can proceed as normal, this time arriving at the correct ranking.

It is worth noting that, like the demotion algorithm, the process of conjunction creation and selection is assumed to work with all the PLD the learner is exposed to, not just a single lexical item as shown in the examples. Thus, a conjunction will only be selected if it consistently assess uncanceled loser marks, and no uncanceled winner marks, throughout the lexicon.

Implications for Local Conjunction

One distinct advantage of the proposed method of conjunction learning is that it rejects strictly impossible conjunctions by default. Hypothetical conjunctions such as MAX&$_\varphi$DEP or VOP&$_\varphi$*[-VOICE] can never be selected, since there is no possible loser candidate which could violate them, thus they could never assess an uncanceled loser mark. This means that there are no additional stipulations required to explain away these impossible conjunctions, making for a somewhat more elegant theory of local conjunction.

Chapter 5

Unnatural Rules

5.1 Introduction

The final argument to be addressed is the claim that languages exhibit phonetically 'unnatural' phonological patterns (Anderson, 1981; Buckley, 2000). In principle, nothing discussed in the previous chapters explicitly forbids strange or unnatural phonological patterns in language. Chapter 4 regards features and markedness constraints as being phonetically grounded. But, as this chapter will show, the processes of constraint ranking and local conjunction are sufficiently powerful that they can derive all manner of phonological patterns. Thus, the existence of seemingly unnatural patterns in phonology is not itself an argument against phonetic grounding.

The chapter proposes analyses for apparently unnatural patterns in Kashaya, Zuni and Odawa. All three of these patterns are expounded in Buckley (2000), which this chapter takes as its starting point. The analyses will show how the patterns in these languages can be derived from the interaction of constraints which are themselves phonetically natural. Constraint local conjunction will play a key role in the Kashaya and Odawa analyses.

At the end of each analysis, a discussion of the markedness constraints used is given. Since I have argued that all markedness constraints must be phonetically grounded, these discussions will focus on the phonetic plausibility of the constraints in question.

The analyses in this chapter are deliberately restricted to segmental phonology. How issues of phonetic grounding may relate to auto-segmental phonology and alignment constraints is a large subject on its own, and thus have been omitted for lack of space. Tackling these issues would represent the next logical step in the approach advocated here.

5.2 Kashaya

Buckley (2000) cites the following synchronic rule in Kashaya, as an exmaple of an “unnatural” phonological rule:

(22) i → u / d__

That is, /i/ is realised as [u] following [d]. Examples of this rule in action are given below in 23.

(23) /i/ backing and rounding (Buckley, 2000)

a.	/cad-i/	→	[cadu]	‘look!’
b.	/wa-ad-i/	→	[waːdu]	‘come here!’
c.	/mahsad-in/	→	[mahsadun]	‘while taking away’
d.	/mo-aq-ad-in/	→	[moːqadun]	‘while running out from here’
e.	/cad-ins’/	→	[caduns’]	‘I wonder if he saw it’
f.	/cic’-id-iʔba/	→	[cic’iːduʔba]	‘would do’

The realisation of /i/ as [u] only happens following [d]. There are three other realisations of /i/ in Kashaya. Firstly, all vowels are realised as [a] following [q], so /qi/ would be realised as [qa]. As Buckley notes, this process is more obviously natural than /di/ → [du], so I won’t deal with it here. Secondly, /i/ lowers to [a] following [m] (Buckley, 1994). Although this is another good candidate for an unnatural rule, it is not obviously related to the /di/ → [du] pattern raised in Buckley (2000). As such I have opted to put the /mi/ → [ma] process to one side and focus solely on the data presented in Buckley (2000). Finally, underlying /i/ is realised as [i] in all other contexts. This is shown in the data below.

(24) Faithful /i/ (Buckley, 2000)

a.	/ʃu-q’aːṭ-i/	→	[ʃuq’aːṭi]	‘groan!’
b.	/du-s’eːk̇-i/	→	[dus’eːk̇i]	‘pleat it!’
c.	/mo-mul-in/	→	[momuːlin]	‘while running around’
d.	/ʔ-ins’-eː/	→	[ʔins’eː]	‘I suppose’

The simplest analysis of the change from [i] to [u] involves two feature changes. A change from [+back] to [-back] and a change from [-round] to [+round], that is, [i] is front and unrounded while [u] is back and rounded. Since [d] is neither back nor rounded there does not appear to be any place feature of [d] which could spread, thereby triggering a process of backing and rounding. Since the rule cannot be motivated by feature spreading, a process plausibly linked to ease of articulation, it can be argued to be phonetically unnatural.

The theme of this chapter is to demonstrate how seemingly unnatural patterns can arise through the interactions of constraints which are themselves entirely natural. The following section gives an analysis of Kashaya which demonstrates how the rule i → u / d__ can result from the interaction of palatalization constraints and faithfulness constraints. Thus, even though the surface alternation between [i] and [u] seems to run counter to phonetic ease, it is nonetheless the product of a phonetically grounded phonology.

5.2.1 Analysis

The analysis outlined here derives the u/i alternation as a repair strategy for avoiding palatalization. That is, [i] should cause the palatalization of the preceding consonant, but rather than change [d] to [d^j], Kashaya instead resolves the conflict by changing [i] to [u], thus removing the trigger for palatalization.

This effect can be derived using only three constraints:

(25) a. ONSET-IDENT

Assign one violation mark for each output segment in an onset position which differs from its input correspondent in its feature specification.

b. IDENT([HIGH])

Assign one violation mark for each output segment which differs from its input correspondent in its specification for the feature [high].

c. PALATALIZECORONAL

Assign one violation mark for each instance of an unpalatalized coronal before [i].[1]

These three constraints are ranked as shown below:

(26) ONSET-IDENT ≫ PALATALIZECORONAL ≫ IDENT([HIGH])

The tableau in (27) demonstrates how these constraints work to produce [ca.du] from /cad-i/ 'look!' (an instance of the i→u/d__ rule in effect):

(27) /di/ ⟶ [du]

/cad-i/			ONS-ID	PALCOR	ID([HI])
a.	☞	ca.du			
b.		ca.d^ji	*!		
c.		ca.di		*!	
d.		ca.da			*!

[1]This constraint is a deliberate oversimplification. See section 5.2.3.

The completely faithful candidate c. is ruled out because the adjacent [d] and [i] violate PALATALIZECORONAL. Normally we would expect this constraint to trigger the palatalization of [d], however, the palatalized candidate b. is ruled out because is violates ONSET-IDENT. This leaves only candidates a. and d., which both avoid violating PALATALIZECORONAL by changing the final vowel. Of the two, candidate a. is the optimal candidate because candidate d. violates IDENT([HIGH]) by realising /i/ as [a].

At this point there is a loophole in the grammar whereby /cad-i/ could be realised as *[cadʲ.i]. By realising the [dʲ] as a coda rather than an onset, ONSET-IDENT would not be violated. This loophole can be closed by introducing a constraint which forbids empty onsets:

(28) ONSET

Assign one violation mark for each syllable which has an empty onset.

A highly ranked ONSET constraint seems to be well motivated in Kashaya, as onsets are obligatory in the language (Oswalt, 1961; Buckley, 1994). Therefore this constraint can be placed at the top of the constraint ranking:

(29) ONSET ≫ ONSET-IDENT ≫ PALATALIZECORONAL ≫ IDENT([HIGH])

This ranking forbids the realisation of /cad-i/ as *[cadʲ.i]. This is demonstrated in the tableau below:

(30) /di/ ⟶ [du]

/cad-i/			ONS	ONS-ID	PALCOR	ID([HI])
a.	☞	ca.du				
b.		ca.dʲi		*!		
c.		ca.di			*!	
d.		ca.da				*!
e.		cadʲ.i	*!			

Candidate e. palatalizes the /d/ without violating ONSET-IDENT by realizing the palatalized consonant in the coda of the first syllable. However, this leaves the final syllable without an onset, fatally violating the ONSET constraint. Therefore this candidate is rejected leaving candidate a. as the optimal candidate.

This constraint ranking accurately predicts the output [ca.du] from the input /cad-i/. However, in its current form this grammar is much too powerful to accurately model Kashaya. Notice what happens when we feed /ʃu-qaːt-i/ 'groan!' into this grammar:

(31) Flase Predicition: /ti/ ⟶ *[tu]

/ʃu-qaːt-i/			Ons	Ons-ID	PalCor	ID([Hi])
a.	✓	ʃu.qaː.ti			✱!	
b.	✗	ʃu.qaː.tu				
c.		ʃu.qaː.tʲi		*!		
d.		ʃu.qaː.ta				*!
e.		ʃu.qaːt.i	*!			

The grammar incorrectly predicts the output *[ʃu.qaː.tu] (candidate b.) instead of the correct [ʃu.qaː.ti] (candidate a.). This is because candidate a. violates the PalatalizeCoronal constraint, whereas candidate b. does not. In Kashaya, /i/ is only realised as [u] after [d] and not after [t]. Therefore what is required is a way of restricting the PalatalizeCoronal constraint such that it will only apply to [d], and no other stops in the language.

Achieving this effect requires a local conjunction between PalatalizeCoronal and VOP, as well as a general faithfulness constraint. Definitions are given below:

(32) a. VOP

Assign one violation mark for every voiced obstruent present in the output.

b. VOP&PalatalizeCoronal

Assign one violation mark for each instance of an unpalatalized, voiced, coronal obstruent before [i].

c. Ident-IO

Any output segment must have identical feature specifications to the corresponding segment in the input.

The local conjunction VOP&PalatalizeCoronal replaces PalatalizeCoronal in the ranking, while Ident-IO is ranked low, giving the following ranking:

(33) Ons ≫ Ons-ID ≫ VOP&PalCor ≫ ID([hi])≫ID-IO

Implicitly, the conjuncts VOP and PalatalizeCoronal are present at the very bottom of the ranking. They are ranked so low as to have no effect on the grammar and thus need not be shown in the tableaux.

This new grammar is demonstrated correctly predicting [ʃu.qaː.ti] from /ʃu-qaːt-i/ in the tableau below:

(34) /ti/ ⟶ [ti]

		/ʃu-qaːt-i/	Ons	Ons-ID	VOP&PalCor	ID([Hi])	ID-IO
a.	☞	ʃu.qaː.ti					
b.		ʃu.qaː.tu					*!
c.		ʃu.qaː.tʲi		*!			*
d.		ʃu.qaː.ta				*!	*
e.		ʃu.qaːt.i	*!				

The grammar now correctly predicts candidate a. as the output. While candidate a. did violate PalatalizeCoronal it does *not* violate VOP and thus does not violate the local conjunction of these two constraints. Candidate b. is rejected because the change from /i/ to [u] violates Ident-IO.

Returning to the example of /cad-i/ [ca.du] 'look!':

(35) /di/ ⟶ [du]

		/cad-i/	Onset	Ons-ID[Pl]	VOP&PalCor	ID[Hi]	ID-IO
a.	☞	ca.du					*
b.		ca.dʲi		*!			*
c.		ca.di			*!		
d.		ca.da				*!	*
e.		cadʲ.i	*!				*

The improved grammar still makes the correct prediction. Unlike the [ti] in [ʃu.qaː.ti], the [di] in *[ca.di] *does* violate *VOP&PalCor, owing to the fact that [d] is a voiced coronal and therefore violates both constraints of the local conjunction.

The local conjunction ensures that /i/ will surface as [u] only following [d] i.e. it correctly produces the i→u/d__ pattern cited by Buckley as an example of an unnatural rule.

5.2.2 How Natural Are VOP and Ons?

The central claim of this chapter is that unnatural phonological patterns can be reduced to interactions between faithfulness constraints and phonetically grounded markedness constraints. The Kashaya analysis only supports this claim if indeed the markedness constraints used can be plausibly argued to be phonetically grounded.

The VOP constraint is practically the poster child for phonetic grounding. So much so that Hale and Reiss in part centre their critique of phonetic grounding around the discussion of this constraint (Hale & Reiss, 2008, ch. 6). Therefore I take it as given that, irrespective of one's stance on phonetic grounding, voiced obstruents are recognised as being phonetically more difficult than voiceless obstruents.

Similarly, the constraint Onset captures a fact of phonology (the naturalness of CV syllables) that I assume to be relatively uncontroversial. Consonants which immediately precede a vowel are more easily perceptible owing to the fact that the consonant's place of articulation is reflected in the formant frequencies of the following vowel (Joanisse, 1999).

5.2.3 How Natural is PALATALIZECORONAL?

The final markedness constraint used in this analysis, PALATALIZECORONAL, is a deliberate oversimplification. While palatalization is a widespread phenomenon, it takes many forms and can be analysed in any number of different ways, depending on which approach to feature representation one wishes to use. I consider issues of feature representation to be tangential to the central issue of phonetic grounding under examination in this dissertation. Thus I have attempted to capture the fact of palatalization with a constraint which remains agnostic regarding the representation of features. I assume that the Kashaya analysis could be adapted to many different theories of feature representation, simply by replacing PALATALIZECORONAL with theory specific constraints. It is quite possible, for example, that PALATALIZECORONAL could be reinterpreted as a variant of the AGREE constraint used in the Zuni analysis in section 5.3.

The downside of such a loosely defined constraint is a loss of detail regarding the exact mechanics of palatalization. Nonetheless, irrespective of how one wishes work out the details, I assume the claim that palatalization is phonetically natural to be fairly uncontroversial. As a form of assimilation/coarticulation, palatalization would seem to be clearly rooted in articulatory ease.

5.2.4 Local Summary

The analysis shows how the /di/→ pattern can be produced by a local conjunction of two grounded constraints, VOP&PALATALIZECORONAL, interacting with faithfulness constraints and ONSET, itself a grounded markedness constraint.

Therefore the existence of the unnatural pattern in Kashaya cannot be considered evidence against phonetic grounding.

5.3 Zuni

Velar stops in Zuni are subject to both palatalization and labialization processes. Additionally, the low vowel /a/ is fronted to [æ] after palatalized stops. The result is that underlying /ka/ will always be realised as [kʲæ] (except in loanwords, discussed in section 5.3.2). Buckley (2000) dubs this process 'hyperpalatalization' and argues that it is unnatural on the grounds that [kʲæ] is more cross-linguistically marked than [ka].

The following section will demonstrate how both hyperpalatalization and labialization can be produced by a high-ranking constraint banning prevocalic velars, interacting with agreement constraints for rounding and backness.

5.3.1 Analysis

Zuni has a total of four velar phonemes. They are, a plain velar and a labialized velar /k/ and /kʷ/, and their non-pulmonic equivalents /k'/ and /kʷ'/. Alternatively, the non-pulmonics have been analysed as the realisation of an underlying velar-glottal cluster, i.e. /kʔ/ and /kʷʔ/ (Newman, 1996). Whether the non-pulmonics are underlying clusters or separate phonemes is not directly relevant to this analysis, however. For simplicity I assume underlying /k'/ and /kʷ'/ throughout.

The phonology of Zuni is such that plain velars are never found directly preceding a vowel. Instead, preceding vowels we find the palatal [kʲ] and labiovelar [kʷ]. The data in 36 demonstrate the distribution of palatal stops in Zuni (both the pulmonic [kʲ] and non-pulmonic [kʲ']).

(36) Zuni Palatalization (Buckley, 2000; Walker, 1972)

a.	[ʔakʲætoːla]	'a type of poppy'
b.	[jaɬapkʲæ]	'they asked him'
c.	[ɬatkʲæ]	'he killed them'
d.	[siʔkʲæneː]	'person with tousled hair'
e.	[kʲ'æweʔ]	'water'
f.	[kʲemeː]	'leather'
g.	[ʔakʲe]	'large metate'
h.	[laːkʲ'i]	'today'

The segment [kʲ], and its non-pulmonic equivalent [kʲ'], only occur directly preceding the front vowels [i, e, æ]. This implies a process of palatalization. Conversely, when directly preceding the round vowels [u, o], velars are always labialized, as shown in the data in 37.[2]

[2]Buckley notes that the rounding effect is not transcribed in older analyses of Zuni phonology. Consequently, the only available data attesting this phenomenon are those given in Buckley (2000).

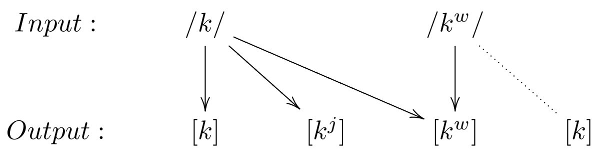
(37) Zuni Labialization (Buckley, 2000)

a. [ʔaːk^{w}u] ‘purple sage’

b. [k^{w}’okʃi] ‘be good’

Because of the labialization process, the contrast between /k/ or /k’/ and /k^{w}/ or /k^{w}’/ is neutralized before round vowels. For the sake of clarity, the underlying-to-surface mappings of the pulmonic velar phonemes is shown graphically in figure 5.1.

Figure 5.1: Realisation of Velars



Note that the non-pulmonics have the same pattern of underlying-to-surface mappings as the pulmonics observed in figure 5.1.

The realisation of /k^{w}/ as [k] before consonants is mentioned in Newman (1996). However, it is not strictly relevant to the issue at hand. As Buckley (2000) does not mention it, neither will I attempt to account for this fact here.

The total absence of [k] and [k’] directly preceding vowels is strong evidence for a constraint explicitly forbidding these sequences. Such a constraint needs to forbid plain velars (both pulmonic and non-pulmonic) whilst allowing the palatalized and labialized variants [k^{j}, k^{w}]. This could be formulated in a number of different ways, depending on which theory of features we adhere to. For example, *[+dorsal, -labial]V would work, but would not be possible in any feature theory which treats [labial] as a privative feature. Similarly, different theories of feature representation could analyse the coarticulation [k^{w}] in any number of different ways. In order to remain agnostic regarding theories of feature representation, this analysis will rely on a simple, if superficial definition of the constraint:

(38) *kV

Assign one violation mark for each sequence of [k] (or [k’]) and a vowel.

A discussion of the naturalness of this constraint is given in section 5.3.3. Implicitly, this constraint doesn’t distinguish between [k] and [k’], it treats them the same. This is entirely reasonable given that, whichever feature theory we choose, [k] and [k’] will share the same passive and active articulators and the same manner of articulation, the only difference between them being pulmonic vs non-pulmonic airflow.

Ranking 38 high will ensure that any prevocalic /k/ or /k'/ present in the input will not make it to the surface. What is required now is a way to cause underlying /k/ and /k'/ to undergo palatalization preceding [i, e, æ] and labialization preceding [u, o]. This can be accomplished with agreement constraints. A general definition for agreement constraints is given in 39:

(39) AGREE(F)
Assign one violation mark for each CV sequence where C and V do not share the same specification for the feature [F].

In keeping with feature-theoretical-agnosticism, the definition in 39 is intended to be broad enough to cover both binary and privative features. For binary features, 39 should be understood as banning $C_{[\alpha F]}V_{[\beta F]}$ where $\alpha \neq \beta$. For privative features it should be understood as banning $C_{[F]}V$ and $CV_{[F]}$.

For clarity, the following tableaux demonstrate the effect of AGREE on either binary or privative features:

(40) AGREE(F) with binary features:

			AGREE(F)
a.		$C_{[+F]}V_{[-F]}$	*!
b.		$C_{[-F]}V_{[+F]}$	*!
c.	☞	$C_{[-F]}V_{[-F]}$	
d.	☞	$C_{[+F]}V_{[+F]}$	

(41) AGREE(F) with privative features:

			AGREE(F)
a.		$C_{[F]}V$	*!
b.		$CV_{[F]}$	*!
c.	☞	$C_{[F]}V_{[F]}$	
d.	☞	CV	

Similarly, AGREE constraints work equally well in a feature geometry approach (e.g. Morén, 2003), where agreement would be regarded as two segments sharing the same feature.

The definition of AGREE given in 39 applies only to CV sequences. This is because the is the only domain of agreement under analysis here. It should be noted that other domains of agreement are entirely possible, e.g. agreement between all vowels in a prosodic word. The type of agreement constraint defined in 39 should be regarded as one member in a 'family' of agreement constraints.[3]

[3]If necessary we could easily distinguish between the different variants of AGREE with the addition of a second argument which defines the domain of agreement. For example, 39 could be AGREE(CV)(F). As there is only one domain of agreement under analysis here, the second argument is largely redundant and thus has been omitted for the sake of readability.

The analysis of Zuni requires two agreement constraints, one for the feature [round] and one for the feature [back]:

(42) AGREE([BACK])
Assign one violation mark for each CV sequence where C and V do not share the same specification for the feature [back].

(43) AGREE([ROUND])
Assign one violation mark for each CV sequence where C and V do not share the same specification for the feature [round].

To produce the palatalization and labialization effects observed in Zuni, these two constraints are ranked below *kV, as shown in 44:

(44) *kV $\gg$ AGREE([BK]), AGREE([RD])

This mini-grammar will successfully derive [kʲi] from /ki/, as in [lakʲ'i] 'today' (shown in 45), and [kʷu] from /ku/, as in [ʔaːkʷu] 'purple sage' (shown in 46).

(45) /ki/ $\longrightarrow$ [kʲi]

/lak'i/		*kV	AGREE([BK])	AGREE([RD])
a.	lak'i	*!	*	
b. ☞	lakʲ'i			
c.	lakʷ'i		*!	*!

(46) /ku/ $\longrightarrow$ [kʷu]

/ʔaːku/		*kV	AGREE([BK])	AGREE([RD])
a.	ʔaːku	*!		*
b. ☞	ʔaːkʷu			
c.	ʔaːkʲu		*!	*!

In both cases, [k] fatally violates *kV. This means it falls to the agreement constraints to determine the most optimal candidate. In 45, the most optimal candidate is [lakʲ'i] because the palatalized [kʲ'] agrees with the following vowel [i] in both backness and rounding. In 46, the most optimal candidate is [ʔakʷu] because [kʷ] agrees with the following vowel [u] in backness and rounding.

A keen-eyed reader might spot that at this point the *kV constraint is actually redundant, the agreement constraints alone would be enough to cause palatalization and labialization. However, the *kV constraint will become indispensible once faithfulness constraints are introduced.

Introducing Faithfulness

So far the grammar only works with a very limited set of inputs and candidates. As it stands, palatalization and labialization will apply to all prevocalic consonants, not just velars. The tableau in 47 shows the grammar falsely predicting *[tʷu] from /tu/:

(47) False prediction: /tu/ ⟶ *[tʷu]

/tu/			*kV	Agree([bk])	Agree([rd])
a.	✓	tu		✱!	✱!
b.	✗	tʷu		*	

Since it is only velars which undergo palatalization and labialization in Zuni, this must be incorrect. The solution is to introduce faithfulness constraints. The following constraint forbids consonants from changing their place features:

(48) Ident$_C$([Place])

Assign one violation mark for every output consonant that differs from its input correspondent in any place feature.

Again, this is a broadly defined constraint designed to be agnostic regarding the representation of features. This constraint could be thought of as representing a bundle of equally ranked constraints, each of which demands faithfulness for a particular consonantal place feature e.g. ≫Ident([lab]), Ident([cor]), Ident([dor]), etc. ≫ . This constraint needs to be ranked above the Agree constraints but below *kV:

(49) *kV ≫ Ident$_C$([Place]) ≫ Agree([bk]), Agree([rd])

The ranking in 49 solves the problem of the over-application of labialization and palatalization:

(50) /tu/ ⟶ [tu]

/tu/			*kV	Ident$_C$(Place)	Agree([bk])	Agree([rd])
a.	☞	tu			*	*
b.		tʷu		*!	*	

Unlike 47, the grammar in 50 rejects the unfaithful form *[tʷu] before the agreement constraints come into play. Therefore [tu] is the optimal candidate, and no labialization is observed.

Crucially, the addition of a faithfulness constraint does not prevent velars from undergoing labialization and palatalization. The ranking in 49 predicts the correct surface forms for [ʔaːkʷu] ‘purple sage’, and [lakʲ’i] ‘today’:

(51) /ki/ ⟶ [k^ji]

/lak'i/			*kV	Ident_C(Place)	Agree([bk])	Agree([rd])
a.		lak'i	*!		*	
b.	☞	lak^j'i		*		
c.		lak^w'i		*	*!	*!

(52) /ku/ ⟶ [k^wu]

/ʔaːku/			*kV	Ident_C(Place)	Agree([bk])	Agree([rd])
a.		ʔaːku	*!			*
b.	☞	$ʔaːk^w$u		*		
c.		$ʔaːk^j$u		*	*!	*!

In both 51 and 52 the optimal candidate (b. in both cases) violates the faithfulness constraint Ident_C(Place). However the violation is not fatal because the only possible candidates which don't violate Ident_C(Place), i.e. candidates which are totally faithful, violate the higher-ranked *kV. As long as /kV/ is present in the input, any candidate which does not violate *kV *must* violate Ident_C(Place)[4], and the optimal candidate will always be determined by the agreement constraints.

Explaining hyper-palatalization

The data in 36 show the vowels [æ] and [a] are in complementary distribution, with [æ] occurring after palatals and [a] occurring elsewhere. There is reason then to suppose that [a] and [æ] are allophones, and that the grammar can derive [k^jæ] from underlying /ka/. This is the process of /ka/ ⟶ [k^jæ] which Buckley cites as unnatural.

The ranking developed so far (given in 49) is already enough to trigger palatalization of the velar and fronting of the vowel when given the input /ka/. This is shown in 53 with the word [k^j'æweʔ] 'water':

(53) /k'a/ ⟶ [k'[-back]]

/k'aweʔ/			*kV	Ident_C(Place)	Agree([bk])	Agree([rd])
a.		[k'a]weʔ	*!		*	
b.		[k^j' a]weʔ		*	*!	
c.	☞	[k^j'æ]weʔ		*		
d.	☞	[k^j'e]weʔ		*		
e.	☞	[k^j'i]weʔ		*		

Note, however, that the grammar cannot distinguish between candidates c., d., and e. ([k^j'æ], [k^j'e] and [k^j'i]). Since [i, e, æ] are all front vowels, all of them satisfy the Agree([bk]) constraint. Given that we know that Zuni phonology prefers /a/→[æ] over/a/→[e] or /a/→[i], this leaves us with the question of how to modify the grammar to capture this fact.

[4]Excluding candidates involving epenthesis or deletion. It is assumed throughout that in Zuni, Max-IO and Dep-IO are ranked higher than any of the constraints examined here, ensuring that epenthesis and deletion are never optimal repair strategies.

The most natural explanation would be that /a/→[e] and /a/→[i] involve a change in backness *and* height, whereas /a/→[æ] only involves a change in backness. The preference for /a/→[æ] then is explained by vowel height faithfulness. This is expressed by the following constraint:

(54) IDENT$_V$([LOW])

Assign one violation mark for every output vowel that differs from its input correspondent in the feature [low].

The constraint defined in 54 is a variant of the faithfulness constraint defined in 48. However, 54 is specific to the feature [low] on vowels.

Looking the tableau in 53, we can see that placing IDENT$_V$([LOW]) anywhere in the ranking would achieve the desired effect. But until we encounter evidence to the contrary, the most parsimonious option would be to rank IDENT$_V$([LOW]) equally with our other faithfulness constraints, under the assumption that all faithfulness constraints are equally ranked in the intial state (see chapter 4).

This gives the ranking shown in 55:

(55) *kV ≫ IDENT$_C$(PLACE), IDENT$_V$([LOW]) ≫ AGREE(BK), AGREE(RD)

This ranking solves the problem observed in 53. The tableau in 56 shows the grammar rejecting [k̇ʲe] and [k̇ʲi] as possible realisations of /ka/:

(56) /k'a/ ⟶ [kʲ'æ]

/k̇awe?/		*kV	ID$_C$(PL)	ID$_V$([LO])	AGREE([BK])	AGREE([RD])
a.	[k'a]weʔ	*!			*	
b.	[kʲ'a]weʔ		*		*!	
c. ☞	[kʲ'æ]weʔ		*			
d.	[kʲ'e]weʔ		*	*!		
e.	[kʲ'i]weʔ		*	*!		

Candidates d. and e. fatally violate the newly introduced IDENT$_V$([LOW]) constraint, leaving c. as the most optimal candidate.

Why not hyper-labialization?

The introduction of unfaithful vowels into the candidate set, and the subsequent faithfulness constraints required to reject them, presents a new problem, namely, the absence of hyper-labialization. The tableau in 57 demonstrates this problem explicitly: the hyper-labialization of /ka/→[kʷɒ] appears to be an equally valid repair strategy as hyper-palatalization.

(57) [kʲ'æ] or *[kʷ'ɒ]?

/k'aweʔ/			*kV	ID$_C$(PL)	ID$_V$([LO])	AGREE([BK])	AGREE([RD])
a.		[k'a]weʔ	*!			*	
b.		[kʲ'a]weʔ		*		*!	
c.	☞	[kʲ'æ]weʔ		*			
d.		[kʷ'a]weʔ		*			*!
e.	☞	[kʷ'ɒ]weʔ		*			

Candidates c. and e. are both optimal, raising the question of why *[kʷ'ɒweʔ] is not attested in Zuni.

The proposed answer to this question is that vowels are always faithful to their input in their specification for the feature [round]. In other words, rounding never spreads *to* vowels from preceding consonants. The is evidenced by the existence of attested forms in which [kʷ] precedes a front, unrounded vowel:

(58) Rounding before front vowels (Buckley, 2000; Newman, 1996)

a.	[ɬakʷiʔkʲæ]	'it was inserted'
b.	[ʔehkʷiʔkʲæ]	'he was first'
c.	[ʔalekʷi]	'to be fried'
d.	[ʔakʷa]	'it is getting ripe'

Note that the [kʷ] in the examples above is an instance of a different phoneme to the [kʷ] produced by labialization (see figure 5.1). However, even if an underlying /kʷ/ is assumed, the grammar thus far cannot successfully produce [kʷi] as an output:

(59) False Prediction: /kʷi/ ⟶ *[kʷu]

/ɬakʷi-ʔka/			*kV	ID$_C$(PL)	ID$_V$([LO])	AGR([BK])	AGR([RD])
a.	✓	ɬa[kʷi]ʔkʲæ				*!	*!
b.		ɬa[ki]ʔkʲæ	*!	*		*	*
c.		ɬa[kʲi]ʔkʲæ		*!			
d.	✗	ɬa[kʷu]ʔkʲæ					

The grammar falsely selects candidate d. *[ɬakʷuʔkʲæ], instead of the correct candidate a. [ɬakʷiʔkʲæ], because the underlying /kʷ/ does not agree with the following vowel [i] in either backness or rounding. However, any change to /kʷ/ is rejected by IDENT$_C$(PLACE). Therefore the grammar resolves the agreement violation by rounding and backing the following vowel /i/→[u].

This is *not* the pattern observed in Zuni however. This solution to both these problems is the introduction of another faithfulness constraint, which prevents rounding processes from apply to vowels:

(60) IDENT$_V$([ROUND])

Assign one violation mark for every output vowel that differs from its input correspondent in the feature [round].

Newman (1996) notes that the only phonological change that applies to vowels in Zuni is a process of shortening. Apparently then, Zuni vowels are very faithful to their inputs in terms of place specification. It would therefore seem reasonable that an analysis of Zuni should be heavily reliant rely on highly ranked faithfulness constraints applying to vowels.

Again, the most parsimonious ranking solution is to rank IDENT$_V$([ROUND]) equally with the other faithfulness constraints, giving the following ranking:

(61) *kV $\gg$ ID$_C$(PLACE), ID$_V$([LOW]), ID$_V$([RD]) $\gg$ AGR([BK]), AGR([RD])

The addition of the new faithfulness constraint allows the grammar to realise /k^{w}i/ faithfully as [k^{w}i]:

(62) /k^{w}i/⟶[k^{w}i]

/ɬak^{w}i-ʔka/		*kV	ID$_C$(PL)	ID$_V$([LO])	ID$_V$([RD])	AGR([BK])	AGR([RD])
a.	☞ ɬa[k^{w}i]ʔk^{j}æ					*	*
b.	ɬa[ki]ʔk^{j}æ	*!	*			*	*
c.	ɬa[k^{j}i]ʔk^{j}æ		*!				
d.	ɬa[k^{w}u]ʔk^{j}æ				*!		

Candidate d. is no longer the optimal candidate, as it violates the newly introduced IDENT$_V$([ROUND]) constraint. This means that although candidate a. still violates the agreement constraints, it is nonetheless the most optimal candidate.

Returning now to the issue of why *[k^{w}ɒ] is not attested in Zuni, we now have an explanation of this fact. The realization of /ka/ as *[k^w'ɒ] would mean changing the [round] specification on the vowel /a/, and therefore is a violation of IDENT$_V$([ROUND]). The tableau in 63 demonstrates this with [k^j'æweʔ] 'water':

(63) [k^j'æ] not *[k^w'ɒ]

/k'aweʔ/		*kV	ID$_C$(PL)	ID$_V$([LO])	ID$_V$([RD])	AGR([BK])	AGR([RD])
a.	[k'a]weʔ	*!				*	
b.	☞ [k^j'æ]weʔ		*				
c.	[k^w'ɒ]weʔ		*		*!		

The grammar now rejects *[kʷ'ɒweʔ] in favour of [kʲæ'weʔ]. This is because /k'a/→[kʷ'ɒ] violates $\textsc{Ident}_V$([ROUND]). Thus, there is no hyper-labialization in Zuni.

Interestingly, [kʷɒ] is a possible output of the grammar, but only if derived from the input /kʷɒ/. Since there is no attested [kʷɒ] in zuni, this analysis predicts that are no lexical items in Zuni which are stored with a /kʷɒ/ in the underlying form, i.e. it is simply a lexical gap owing to diachronic sound changes. It would be interesting to test this hypothesis on native speakers of Zuni. Unfortunately, doing so is far beyond the scope of this dissertation.

5.3.2 [ka] in Loanwords

Buckley notes that in fact [ka] *is* is attested in Zuni, but only in loanwords:

(64) Loanwords with [ka] (Buckley, 2000)

a.	[melika]	'white man'
b.	[kaːpo]	'cowboy'
c.	[katʃuːtʃanpoʔjanne]	'railroad (man's) cap'

A naive and hasty reaction to this fact might be to reject entirely the analysis in section 5.3.1 on the grounds that [ka] is attested in the language and therefore not ungrammatical. However, before we jump to such extremes it is worth considering the significance of the 'loanword' status of the data in 64.

When loaning foreign words, speakers frequently apply their native phonotactics to the imported words. This results in the loanword having a significantly different pronunciation from that of the donor language, e.g. Hawaian [kalikimaki] from English 'Christmas' (Kang, 2011). This is not always the case however, sometimes loanwords are seemingly able to violate the phonology and/or phonotactics of the recipient language. Consider, for example, the lack of native English words with word-initial [ʒ]. This is plausibly a fact of the phonotactics of the language. Nonetheless, many speakers of English will use word-initial [ʒ] with recently imported French words, such as 'jus' [ʒu:]. These non-native pronunciations are often used consciously, as a way of emphasizing the foreign or exotic nature of the loanword.

Taking these facts into account, it can reasonably be argued that such loanwords belong to a different sociolinguistic register. Different registers often obey different phonological rules. In effect, speakers momentarily switch grammars every time they switch registers. Thus the differing phonology of loanwords with exotic pronunciations can be modeled with different grammars. Naturally, we should expect there to be limits on just how far loanword grammars can deviate from the speaker's standard grammar. Therefore this hypothesis is more plausible if we can show that the phonological/phonotactic violations incurred by loanwords are producible by a grammar which is minimally different from the standard grammar of the recipient language. For example, in OT terms, we could posit that an ideal loanword grammar should be no more than a single constraint promotion/demotion from the standard grammar.

The question then, is this the case for loanwords in Zuni? Can we derive a grammar capable of /ka/→[ka], that is only a single step from the grammar which gives /ka/→[kʲæ]? The answer is yes. We need only move *kV down a single step to below the faithfulness constraints (shown in 65).

(65) Deriving the loanword grammar:

$$\begin{tikzcd} {*}\mathrm{kV} \arrow[drr] & \gg \textsc{Faithfulness} & \gg \emptyset & \gg \textsc{Agree} \\ \emptyset & \gg \textsc{Faithfulness} & \gg {*}\mathrm{kV} & \gg \textsc{Agree} \end{tikzcd}$$

This new ranking will derive the correct output for loanwords containing [ka] i.e. without hyper-palatalization.This is shown in 66 with the word [melika] 'white man'.

(66) Loanword grammar: /melika/ ⟶ [melika]

/melika/	$\mathrm{ID_C}$(Pl)	$\mathrm{ID_V}$([Lo])	$\mathrm{ID_V}$([Rd])	*kV	Agr([Bk])	Agr([Rd])
a. ☞ [melika]				*	*	
b. [melikʲæ]	*!					

Candidate b., which has the hyper-palatalized [kʲæ], would normally be the optimal candidate. However, because *kV is demoted, the violation of $\mathrm{ID_C}$(Place) is now fatal. Therefore candidate a. is the optimal candidate.

5.3.3 How Natural is *kV?

On the face of it, *kV might not seem to be particularly natural. Certainly, [k] is not a rare segment, 97.12% of the languages in the UPSID database have [k] `http://web.phonetik.uni-frankfurt.de/upsid.html`. Similarly, there is little reason to suppose that [k] is a particularly difficult segment to articulate. This makes ease of articulation an unlikely motivation for a *kV constraint.

For this reason, the most plausible phonetic explanation for this constraint lies not in articulatory difficulty but in perceptual salience. So how perceptually salient is prevocalic [k]?(Hume, Johnson, Seo, & Tserdanelis, 1999) studied of the relative perceptual salience of [p, t, k] before [i, a, u] for native speakers of Seoul Korean and American English. Dorsals are commonly argued to be more salient than coronals or labials, a claim confirmed by Hume et al.. This explains why, in voiceless stop clusters in Korean, a coronal or labial before a dorsal will assimilate to the dorsal place of articulation, but dorsals never assimilate to coronals or labials (*ibid.*). Note however that this refers to dorsal stops before other stops, whereas *kV refers specifically to prevocalic environments.

The findings of Hume et al. suggest that the salience of [k] before vowels is somewhat more complex. While [p, t] exhibit a similar degree of salience in front of all the vowels tested [i, a, u], the salience of [k] varied depending on which vowel it preceded. That is, the salience of prevocalic [k] depends heavily on the quality of the following vowel, and to a much greater degree than is the case for [p] and [t].

It is quite plausible that the inconsistent perceptibility of [k] in prevocalic positions could lead to the grammar treating kV as a marked sequence. Certainly, we should expect consistency to be a major factor in any communicative system. This *salience instability*, as we might call it, would be neatly resolved by phonological processes of assimilation, which *kV helps motivate in the case of Zuni.

It should be noted that Hume et al. are cautious of reliance on perceptual salience as the sole motivator of phonological markedness:

(67) [The results suggest] caution in appealing to auditory-perceptual data as the sole motivation for processes such as place assimilation and markedness, in general. This is underscored by the observation that patterns of place assimilation can differ cross-linguistically. For example, place assimilation in English and Sri Lankan Portuguese Creole differ from the pattern noted in [...] Korean: in English, only coronals typically undergo assimilation, while in Sri Lankan Portuguese Creole, of the three places of articulation, only labials and dorsals appear to be targets of assimilation. (Hume et al., 1999, p. 2071)

This does not necessarily contradict any of the assumptions in this analysis. The analysis given here supposes only that *kV is motivated by salience instability. It is entirely compatible with the claim that other types of phonological markedness, e.g. those that force labials and coronals to assimilate, are motivated by different phonetic factors.

5.3.4 How Natural is Agree?

Place agreement by itself is not particularly difficult to reduce to phonetic factors. By sharing the same feature specification, adjacent segments become more similar in their place of articulation, which means less movement and effort for the active articulators.

What is perhaps more interesting is the claim that the domain of agreement can be restricted to CV sequences[5] given that articulatory ease is no more of a factor in a CV sequence than in any sequence of segments. There is however a perceptual argument that can be made for CV sequences. Firstly, place of articulation errors are more perceptible in onset positions than in codas (Cole, Jakimik, & Cooper, 1978). This fact alone would suggest that the grammar should be more concerned with place agreement in CV sequences than, say, VC sequences.

Additionally, allowing place agreement to CV sequences, and not across syllable boundaries, has the benefit of making the syllable structure more explicit in the sound stream. When place agreement is restricted to syllable internal sequences, similarity in place of articulation is evidence that two adjacent segments should be parsed in the same syllable, and dissimilar places of articulation become evidence of a syllable break. Parsing a stream of segments into syllables is by no means a trivial task, yet it is certainly an essential part of language comprehension (Dorffner, Kwasny, & Port, 1987) and by extension language learning.

Speculatively, we should expect a phonetically grounded phonology to incorporate precisely these kinds of perceptual aids.

5.3.5 Local Summary

The pattern of /ka/→[$\mathrm{k^j}$æ] in Zuni can be derived purely from the interaction of agreement constraints, faithfulness constrains, and the markedness constraint *kV, which is plausibly a product of the perceptual instability of [k] in prevocalic contexts. The apparently unnatural pattern can be derived by constraints which are phonetically grounded. Thus, the Zuni data is not evidence against phonetic grounding.

Unlike the Kashaya and Odawa analyses, the Zuni analysis does not rely on local conjunction. Arguably, this is a positive result. While the hyperpalatalization in Zuni is not entirely phonetically sensible, it is considerably less outlandish than the unnatural processes in Kashaya and Odawa. This would seem to support the hypothesis that markedness constraints are phonetically grounded, and local conjunctions are the primary source of phonological craziness.

Finally, the Zuni analysis predicts that the absence of [$\mathrm{k^w}$ɒ] is because of a lexical gap, rather than because the sequence [$\mathrm{k^w}$ɒ] is ungrammatical. In principle, this is a testable hypothesis. It would be interesting to see how Zuni speakers react to nonsense forms containing [$\mathrm{k^w}$ɒ], and whether they judged them to be more or less unacceptable than forms containing, for example, [$\mathrm{k^w}$æ].

[5] Or alternatively, Onset-Nucleus sequences. As Zuni lacks complex onsets at the surface level, the two notions are extensionally equivalent for the Zuni analysis.

5.4 Eastern Ojibwa/Odawa

Buckley (2000) cites the palatalization processes in Eastern Ojibwa as unnatural. Buckley argues that the palatalization was historically a phonetically natural process, but that subsequent changes in the language have rendered the the resulting alternations phonetically opaque.

The analysis given below demonstrates how the seemingly opaque rules in Eastern Ojibwa can result from the interaction of grounded or 'natural' constraints. The analysis does however rely on a degree of abstractness in the underlying representations. Arguably, this deviates from the standard OT assumption of Lexicon Optimization (Prince & Smolensky, 1993). Nonetheless, there is evidence that some degree of abstractness in underlying representations is both possible and necessary in OT (Krämer, 2003, 2006). Crucially, this abstractness is entirely in keeping with the view of phonology advocated in this thesis, i.e. an autonomous, symbolic computer which is grounded in phonetic factors. This section builds on the analysis of Odawa in Piggott (1980).

5.4.1 Analysis

The Eastern Ojibwa dialect of Odawa exhibits some unusual patterns of coronal palatalization. Firstly, there exists a fairly natural alternation between [n] and [ʃ], with [ʃ] occurring before [i] and [n] occurring elsewhere:

(68) [n]-Palatalization (Piggott, 1980)

a.	[kimiːnaː]	'you give him'
b.	[kimiːʃimi]	'you give us'
c.	[kinaːnaː]	'you fetch him'
d.	[kinaːʃimi]	'you fetch us'

So far this appears to be a reasonably typical pattern of palatalization. However, there are exceptions, as certain stems appear to never undergo palatalization. The data in 69 shows the same suffix which produced palatalization in 68, failing to produce palatalization when attached to the stem /weːpin/ 'leave behind':

(69) Non-Palatalizing Stems (Piggott, 1980)

a.	[kweːpinaː]	'you leave him behind'
b.	[kweːpinimi]	'you leave us behind'

Furthermore, there appear to be affixes where [i] fails to cause the palatalization of [n], even when attached to stems which otherwise undergo palatalization. The data in 70 show the same verb stems as in 68, failing to undergo palatalization before the suffix /-in/:

(70) Non-Palatalizing Affixes (Piggott, 1980)

a.	[kimiːnaː]	'you give him'
b.	[kimiːnin]	'I give you'
c.	[kinaːnaː]	'you fetch him'
d.	[kinaːnin]	'I fetch you'

Perhaps most perplexing, are those affixes which don't cause palatalization of [n] but which do cause palatalization of [t], such as the passive suffix [ikaːso]:

(71) Variably Palatalizing Affixes (Piggott, 1980)

a.	[miːnikaːso]	'he is given'
b.	[weːppitaw]	'hit someone'
c.	[weːppiʧikaːso]	'be hit'

The solution to these apparently unnatural patterns will assume some degree of abstractness in the underlying representations. Following Piggott, I will assume that those instances of [n] which alternate with [ʃ], are in fact allophones of an underlying /l/. Those instances of [n] which do *not* alternate with [ʃ] are derived from underlying /n/. Also following Piggott, I assume that those instances of [i] which fail to cause palatalization of [n] (from /l/), are instances of underlying /e/. Only those instances of [i] that cause palatalization of [n](from /l/) are derived from underlying /i/. Piggott (1980) provides some extended argumentation for these assumptions within a rule-based framework, however as the rest of this section will demonstrate, these assumptions work equally well in an OT analysis.

Motivating Palatalization

Assuming that both [n] and [ʃ] are allophones of /l/, the first step is build a grammar that derives [n] and [ʃ] from /l/. Doing this requires three constraints, the first of which is a markedness constraint banning laterals:

(72) *[lateral]

Assign one violation mark for each instance of a lateral in the output.

Ranking this constraint high prevents /l/ from ever surfacing as a lateral. Next, we need constraints to help choose between [n] and [ʃ], depending on the environment:

(73) Ident([sonorant])

Assign one violation mark for every output segment that differs from its input correspondent in its specification for the feature [sonorant].

(74) PalatalizeCoronal

Assign one violation mark for each instance of an unpalatalized coronal before [i].[6]

The palatalization constraint defined above is the same as the constraint given in 25c for the Kashaya analysis. Like the constraint given in 25c, the definition is deliberately superficial but hopefully uncontroversial.

The Ident([sonorant]) constraint punishes a change in the specification for the feature [sonorant]. Given that [n] and [l] both share the same specification for this feature, this constraint explains why [n] is the first choice for the realisation of /l/.

These three constraints are ranked as shown in 75:

(75) *[lateral] ≫ PalatalizeCoronal ≫ Ident([sonorant])

This ranking will produce the correct pattern of palatalization for underlying /l/, producing [ʃ] before [i] as in [kimiːʃimi] 'you give us' and [n] elsewhere as in [kimiːnaː] 'you give him'. This is demonstrated by the tableaux below:

(76) /li/ ⟶ [ʃi]

		/ki-miːl-i-mi/	*[lateral]	PalCor	Ident([son])
a.		kimiː[li]mi	*!	*	
b.	☞	kimiː[ʃi]mi			*
c.		kimiː[ni]mi		*!	
d.		kimiː[ʎi]mi	*!		

[6]This constraint is a deliberate oversimplification. See section 5.2.3.

(77) /la/ ⟶ [na]

/ki-miːl-aː/			*[lateral]	PalCor	Ident([son])
a.		kimiː[laː]	*!		
b.		kimiː[ʃaː]			*!
c.	☞	kimiː[naː]			
d.		kimiː[ʎi]mi	*!		

In both cases, candidates which realise /l/ as [l] or [ʎ] are rejected because they fatally violate *[lateral]. In 76, b. is the optimal candidate because the sequence *[ni] violates PalatalizeCoronal. In 77 however, neither b. nor c. violate PalatalizeCoronal meaning that it falls to Ident([son]) to determine the optimal candidate. As [n] shares the same specification for the feature [sonorant] as /l/, candidate c. does not violate Ident([son]) and thus is the optimal candidate.

This analysis doesn't explain why palatalization results in [ʃ] and not, for example, [ɲ]. Since [ɲ] is entirely lacking from the inventory of Odawa, we can simply assume the existence high ranking constraints banning [ɲ]. However, making such constraints explicit in the tableaux would only serve to take up space and hinder readability.

/n/ vs /l/

As already noted, those stems in which [n] does *not* palatalize to [ʃ] are treated as having an underlying /n/. To stop /n/ from being realised as [ʃ], a faithfulness constraint is required.

(78) Max([nasal])

Assign one violation mark for each instance of the feature [nasal] in the input which does not have a corresponding instance of the feature [nasal] in the output.

This constraint is violated if /n/ is realised as anything other than a nasal. Implicitly 78 treats [nasal] as a privative feature. If using binary features, the constraint required to achieve the same effect is Ident([+nas]).

If Max([nasal]) is to be effective at preventing palatalization then it needs to be ranked higher than PalatalizeCoronal, as shown below:

(79) *[lateral] ≫ Max([nasal]) ≫ PalatalizeCoronal ≫ Ident([sonorant])

This ranking will not result in the palatalization of underlying /n/. This is demonstrated in the tableau below with the word [kiwe:pinimi] 'you leave us behind'.

(80) /ni/ $\longrightarrow$ [ni]

/ki-we:pin-i-mi/			*[lateral]	MAX([NAS])	PALCOR	IDENT([SON])
a.		kiwe:pi[li]mi	*!		*	
b.		kiwe:pi[ʃi]		*!		*
c.	☞	kiwe:pi[ni]mi			*	

Even though candidate c. violates PALATALIZECORONAL, it is nonetheless the most optimal candidate because candidate b. fatally violates MAX([NASAL]), thus /n/ cannot be palatalized to [ʃ].

Underlying /e/

Having provided a solution to the issue of non-palatalizing stems, we may turn to the issue of non-palatalizing affixes. As noted, instances of [i] which don't trigger palatalization are assumed to be derived from underlying /e/. Of course, to stop /e/ surfacing as [e] a markedness constraint is required:

(81) *e

Assign one violation mark for each instance of [e] present in the output.[7]

Since markedness constraints like PALATALIZECORONAL only apply to what is present in the output, not the input, simply analyzing some instances of [i] as underlying /e/ is not enough to block palatalization.

In order for /e/ to block palatalization, the analysis will rely on a local conjunction of two faithfulness constraints:

(82) IDENT([SONORANT])$\&_\sigma$IDENT([HIGH])

Assign one violation mark for each syllable which contains a violation of both IDENT([SON]) and IDENT([HI])

Note that this is a local conjunction in the syllable domain, a particularly powerful form of local conjunction. See section 1.5.3 for a disucssion of the typological implications of this sort of conjunction.

The first of the constraints in the local conjunction is defined in 73. The second, IDENT([HIGH]), is defined below:

(83) IDENT([HIGH])

Assign one violation mark for each output segment that differs from its input correspondent in its specification for the feature [high].

[7]This constraint is in fact a shorthand for the combined effect of multiple, stringently related constraints. This is discussed at length in section 5.4.2

The local conjunction must be ranked higher than PALATALIZECORONAL if it is to block palatalization. This gives the final ranking (including *e) shown below:

(84) *[lateral], *e ≫ IDENT([SONORANT])$\&_\sigma$IDENT([HIGH]), MAX([NASAL]) ≫ PALATALIZECORONAL ≫ IDENT([SONORANT])

This ranking will not produce palatalization in cases where adjacent /l/ and /e/ are present in the input. This is demstrated in the tableau below with [kimiːnin] 'I give you':

(85) /le/ ⟶ [ni]

	/ki-miːl-el/	*[lateral]	*e	ID([SON])$\&_\sigma$ID([HI])	MAX([NAS])	PALATALIZE	IDENT([SON])	IDENT([HI])
a.	kimiː[lel]	*!*	*					
b.	kimiː[nen]		*!					
c. ☞	kimiː[nin]					*		*
d.	kimiː[ʃin]			*!			*	*

Candidate a. and b. are ruled out because both [l] and [e] fatally violate high-ranked markedness constraints. Therefore /e/ has to raise to [i], violating IDENT([HI]). On its own this violation is not fatal because IDENT([HI]) is ranked low. However, if the underlying /l/ is also realised as a non-sonorant [ʃ] then the higher-ranked local conjunction is violated, ruling out the palatalization in candidate d.

Variably Palatalizing Affixes

The one outstanding issue is the case of affixes which appear to cause the palatalization of [t] but not [n] (i.e. /l/), such as the passive suffix [ikaːso]. In fact, the ranking given in 84 already produces the correct pattern in this case. Since the passive marker does not cause palatalization of /l/, we can take the underlying form to be /-ekaːso/ (underlying /e/). As we would expect, when attached to a stem which ends in /l/, no palatalization occurs:

(86) /le/ ⟶ [ni]

/miːl-ekaːso/		*[lateral]	*e	ID([son])&$_\sigma$ID([hi])	Max([nas])	Palatalize	Ident([son])	Ident([hi])
a.	miː[le]kaːso	*!*	*					
b.	miː[ne]kaːso		*!					
c. ☞	miː[ni]kaːso					*		*
d.	miː[ʃi]kaːso			*!			*	*

However, when /-ekaːso/ is attached to a stem which ends in /t/, palatalization *does* occur. This is demonstrated below with [weːppiʧikaːso] 'be hit':

(87) /te/ ⟶ [ʧi]

/weːppit-ekaːso/		*[lateral]	*e	ID([son])&$_\sigma$ID([hi])	Max([nas])	Palatalize	Ident([son])	Ident([hi])
a.	weːppi[te]kaːso		*!					
b.	weːppi[ti]kaːso					*!		*
c. ☞	weːppi[ʧi]kaːso							*

Because palatalizing /t/ to [ʧ] does not involve a change in sonority, neither Ident([son]) nor the local conjunction is violated. Thus candidate c. is the most optimal candidate.

5.4.2 How Natural is *e?

The markedness of vowels is a somewhat different issue from the markedness of consonants or consonant clusters. Speculatively, we should expect articulatory ease to play a much smaller role in vowel markedness. For example, the difference in articulatory ease between [i] and [e] is plausibly much smaller than [t] and [ʧ], given that vowels do not require the same precise coordination that stops do. Similarly, because all vowels involve very little closure in the oral cavity, the difference in air pressure across the glottis is likely much smaller between different vowels than between different types of consonants. While ease of articulation might explain the cross-linguistic tendency for vowels to reduce to [ə] and not [ʉ], for example, we shouldn't expect ease of articulation to tell us much about Odawa's preference for [i] over [e].

Instead, the markedness of a vowel in any given language should be understood in relation to the entire vowel system used by that language. Vowel systems tend to spread to fill the entire acoustic space, in order to maximize the perceptual contrast between the different vowels (Liljencrants & Lindblom, 1972). Examples of vowel systems which maximize perceptual contrast would be [i, a, u], [i, e, a, o, u] and [i, ɪ, e, a, o, ʊ, u]. Whereas a vowel system such as [i, ɪ, ʊ, u] would *not* be maximally distinctive, because it lacks low vowels and there fore does not 'fill out' the perceptual space.

As such, vowel markedness can be understood as the failure of a vowel to conform to a perceptually distinctive vowel system. But how does one capture such a definition using OT constraints? Set theory can help us to accomplish this. First, each of the maximally distinctive vowel systems given above is defined as a set:

$$A = \{\text{i, a, u}\} \tag{5.1}$$

$$B = \{\text{i, e, a, o, u}\} \tag{5.2}$$

$$C = \{\text{i, ɪ, e, a, o, ʊ, u}\} \tag{5.3}$$

$$\tag{5.4}$$

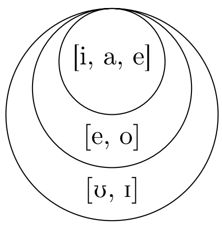
Since each set is a maximally distinctive vowel system, the 'phonetically natural' tendency for languages should be to use a vowel system which corresponds to one of these sets. Vowel markedness can be defined as *not* being a member of a set of maximally distinctive vowels. So, if a language has B as its vowel system, any member of B would be unmarked, while any element which is not in B would be marked.

So far this only delineates marked from unmarked within a given language. This does not yet give us a universal definition of vowel markedness. Since any vowel belongs to *some* set of maximally distinctive vowels, any vowel is unmarked in *some* vowel system.

A universal definition of vowel markedness requires an understanding of the relationships between the different vowel systems. Notice that the examples of maximally distinctive vowel systems above exhibit a sub-/superset relationship:

$$A \subset B \subset C \tag{5.5}$$

Or:



Therefore any element (i.e vowel) which is a member of A, must also be a member of B and C:

$$\forall x \in A : x \in B \tag{5.6}$$
$$\forall x \in B : x \in C \tag{5.7}$$
$$\therefore \forall x \in A : x \in B \cap C \tag{5.8}$$

Bearing in mind our definition of markedness, we can say that if a language contains vowels that are members of B, any vowel which is a member of A must be unmarked in that language, since any vowel which is a member of B *must* also be a member of A.

Consequently, by comparing all the possible maximally distinctive vowel systems (A, B and C in this case). We can see that some elements are more frequently unmarked than others. Specifically, the elements in A are always unmarked, because the elements in A are always members of every other set. The elements in B-A are sometimes unmarked, since they are members of B and C but not A. While the members in C-B are rarely unmarked since they are only members of C and not B or A. This gives a universal, cross-linguistic definition for vowel markedness: The fewer maximally distinctive vowel systems a vowel belongs to, the more marked that vowel is.

Moreover, this makes it is possible to construct a scale of vowel markedness, since the relative compliment of any set will always be more marked than the set itself:

$$A < B - A < C - B \tag{5.9}$$
$$\{\text{i, a, u}\} < \{\text{e, o}\} < \{\text{ɪ, ʊ}\} \tag{5.10}$$

What this hierarchy expresses is that he further to the left a vowel is, the less marked that vowel is.

Universal Hierarchies in OT

Now we are left with the question of how to express this hierarchy using OT constraints. Since the hierarchy is universal, it is not enough to simply posit constraints like *{e, o} and *{ʊ, ɪ}, because in OT constraints can be freely ranked in any order. If *{e, o} were ranked higher than *{ʊ, ɪ}, for example, then the the hierarchy would no longer be represented in the grammar.

One solution to the problem of expressing universal hierarchies using freely rankable constraints is the use of stringently related set constraints (de Lacy, 2006). Using this approach, the scale of vowel markedness can be expressed using the following constraints:

(88) a. *{ɪ, ʊ}
Assign one violation mark for each instance of [ɪ] or [ʊ] present in the output.

b. *{ ɪ, e, o, ʊ,}

Assign one violation mark for each instance of [ɪ], [e], [o] or [ʊ] present in the output.

c. *{ i, ɪ, e, a, o, ʊ, u}

Assign one violation mark for each instance of [i] [ɪ], [e], [a], [o], [ʊ] or [u] present in the output.

These constraints will always produce the same output, irrespective of their relative ranking. For example, in 89 we can see that [ɪ] and [ʊ] violate *all* the constraints, thus the relative ranking makes no difference to the outcome:

(89) Stringently related constraints:

	*{ɪ, ʊ}	*{ ɪ, e, o, ʊ,}	*{ i, ɪ, e, a, o, ʊ, u}
[i]			*
[a]			*
[u]			*
[e]		*	*
[o]		*	*
[ɪ]	*	*	*
[ʊ]	*	*	*

In the absence of any faithfulness constraints, the stringently related constraints will cause all vowels to be realised as either [i],[a] or [u]. In order capture the grammar of Odawa, where underlying /e/ is realised as [i], we need only include a single faithfulness constraint:

(90) IDENT([BACK])

Assign one violation mark for each output segment that differs from its input correspondent in its specification for the feature [back].

The addition of this constraint ensures that /e/ will be realised as [i]:

(91) /e/ ⟶ [i]

/e/	*{ɪ, ʊ}	*{ ɪ, e, o, ʊ,}	*{ i, ɪ, e, a, o, ʊ, u}	ID([BK])
☞ [i]			*	
[a]			*	*!
[u]			*	*!
[e]		*!	*	
[o]		*!	*	*
[ɪ]	*!	*	*	
[ʊ]	*!	*	*	*

To bring us back to the topic at hand: What does this tell us about the naturalness of the constraint *e? The constraint *e used in the Odawa analysis should be understood as a simplified shorthand for the entire set of constraints used in 91. More explicitly, wherever *e occurs in the analysis, it can be substituted with all the constraints in 91. As the combined constraints in 91 have the same effect as the constraint *e, the predicted output will remain unchanged. It is nothing more than a difference in notation.

Since the effect of explicitly writing out all constraints in 91 would be a lot of very large and unreadable tableaux, I have chosen not to do so. Instead I trust the reader to keep in mind that *e is a shorthand when reading the analysis. See 103 for an example using all the constraints in full.

Imperfect Vowel Systems in Natural Language

The set-theoretic account of vowel markedness given here is highly idealised. The example vowel systems used are very simple and ignore a vast number of different vowels that language employ in practice. Of course, attempting to include every possible set of 'perfect' vowels would have made the analysis unreadable. But there is no obvious reason why the basic principle would not hold when expanded to include larger numbers of possible vowel systems.

A more serious concern however, is that real languages are rarely so neat and tidy that their vowel are perfectly distributed to ensure maximum perceptual distinctiveness. While there are certainly languages which employ a neat 3 or 5 vowel system, there are many which do not. In fact Odawa is one such language. While Odawa apparently dislikes [e] it has no problem with [o]. This might seem to be a problem, given that [e] and [o] both occupy the same rung in the vowel markedness hierarchy. If [e] and [o] are equally marked, why should Odawa allow one and not the other?

It is worth bearing in mind that the stringently related constraints are only intended to explain the effect of perceptual distinctiveness on the grammar. This thesis assumes that phonetics influences phonology through ontogeny, but it does not claim that phonology should be entirely reducible to phonetics. Phonological grammars can vary greatly and are capable of being as phonetically 'unnatural' as the computational system will allow. In OT, the idiosyncrasies of individual languages are explained through the ranking of constraints. Therefore the fact that Odawa allows [o] and not [e] should be explained as the effect of constraints other than just the stringently related markedness constraints.

For example, we could posit the existence of constraints which causes underlying /o/ to be more faithful than underlying /e/:

(92) IDENT([ROUND])

Assign one violation mark for each output segment that differs from its input correspondent in its specification for the feature [round].

(93) *[ROUND]&IDENT([HIGH])

Assign one violation mark for each output segent which differs from its input correspondent in its specification for the feature IDENT([HI]) and which bears the feature [round]

These constraints ensure that underlying /o/ will be realised as [o], without effecting the realisation of /e/ as [i]:

(94) /o/ ⟶[o]

/o/	IDENT([ROUND])	*[ROUND]&IDENT([HIGH])	*{ɪ, ʊ}	*{ ɪ, e, o, ʊ,}	*{ i, ɪ, e, a, o, ʊ, u}	IDENT([BACK])
[i]	*!				*	*
[a]	*!				*	
[u]		*!			*	
[e]	*!			*	*	*
☞ [o]				*	*	
[ɪ]	*!		*	*	*	*
[ʊ]		*!	*	*	*	

The local conjunction punishes round segments which are not faithful to their height specification, ensuring that /o/ cannot be realised as the less marked [u]. The constraint ID([ROUND]) ensures that the local conjunction cannot be 'circumvented' by realising /o/ as an unrounded vowel.

The local conjunction has no effect on /e/ being realised as [i], since neither /e/ nor [i] are rounded vowels:

(95) /e/ ⟶[i]

/e/	IDENT([ROUND])	*[ROUND]&IDENT([HIGH])	*{ɪ, ʊ}	*{ ɪ, e, o, ʊ,}	*{ i, ɪ, e, a, o, ʊ, u}	IDENT([BACK])
☞ [i]					*	
[a]					*	*!
[u]	*!				*	*
[e]				*!	*	
[o]	*!			*	*	*
[ɪ]			*!	*	*	
[ʊ]	*!		*	*	*	*

5.4.3 How Natural is *[lateral]?

Laterals are somewhat typologically rare, which implies that they may have a marked status. Laterals of one form or another are only present in 84.04% of the languages in the UPSID database (`http://web.phonetik.uni-frankfurt.de/upsid.html`). While one could reasonably argue that laterals are inherently more articulatorily 'difficult' than other consonants, given the unusual use of the tongue required to produce lateral airflow, this may not be the best explanation for why Odawa disprefers laterals.

Instead, I propose that the phonetic origins of *[lateral] can be explained in terms of segmental sonority prominence (Prince & Smolensky, 1993). Independent evidence for this proposal comes from the fact that Odawa lacks not only laterals but also rhotics. Therefore the lack of laterals (and rhotics) can be understood as a ban on highly-sonorant segments in low prominence positions like onsets and codas. This can be shown to fall out from from the alignment of two universal hierarchies, the segmental prominence hierarchy and the syllabic prominence hierarchy. Prominence Alignment is essentially a phonetic or 'natural' concern, as it is based on the notion of maximizing the perceptual difference between peaks and margins, based on sonority.

Like the discussion of *e, this section will involve unpacking *[lateral] into several stringently related constraints. Following Uffmann (2007), I assume the following hierarchy of segmental prominence:

(96) Vowels $>$ r $>$ laterals $>$ nasals $>$ obstruents $>$ laryngeals

This hierarchy states that vowels are the most prominent, and laryngeals the least prominent. This hierarchy is then aligned with another hierarchy, the syllabic prominence hierarchy:

(97) Peak $>$ Margin

This hierarchy states that peaks (i.e. nuclei) are more prominent than margins (i.e. onsets and codas). The two hierarchies are aligned such that the more prominent segments are preferred in more prominent positions. Uffmann (2007) derives the two following markedness scales:

(98) *Margin/V $\gg$ *Margin/r $\gg$ Margin/lat $\gg$ *Margin/nas $\gg$ *Margin/obs $\gg$ *Margin/lar

(99) *Peak/lar $\gg$ *Peak/obs $\gg$ *Peak/nasal $\gg$ *Peak/lat $\gg$ *Peak/r $\gg$ *Peak/V

Note however that these two scales face the same problem as the scale of vowel markedness in section 5.4.2. If these constraints are simply included in a grammar as-is then they can easily be reranked such that they no longer reflect the universal nature of the hierarchies. Once again, we can turn to stringently related constraints to solve the problem (de Lacy, 2006). Rewriting the margin markedness scale using stringently related constraints gives the following constraints:

(100) *Margin/V ≫ *Margin/{V, r} ≫ Margin/{V, r, lat} ≫ *Margin/{V, r, lat, nas} ≫ *Margin/{V, r, lat, nas, obs} ≫ *Margin/{V, r, lat, nas, obs, lar}

When arranged into a tableau, we can see how these constraints work. Like the vowel constraints in section 5.4.2, the relative ranking of these constraints makes no difference to the outcome:

(101) Stringently related constraints:

	*Margin/V	*Margin/{V, r}	*Margin/{V, r, lat}	*Margin/{V, r, lat, nas}	*Margin/{V, r, lat, nas, obs}	*Margin/{V, r, lat, nas, obs, lar}
[ʔ]						*
[t]					*	*
[n]				*	*	*
[l]			*	*	*	*
[r]		*	*	*	*	*
V	*	*	*	*	*	*

Variation between languages, in terms of which segments are allowed in which syllabic positions, depends on the interaction of the markedness scale with other constraints. In the case of Odawa, the first three constrains in the scale are ranked high, while the others are ranked low. The tableau below is the same as in 86, but with *[lateral] unpacked into the segmental markedness scale. Note that only the first three constraints from the markedness scale take the place of *[lateral], while the others are ranked so low as to have no effect:

(102) /le/ ⟶ [ni]

/miːl-ekaːso/		*Margin/V	*Margin/{V, r}	*Margin/{V, r, lat}	*e	ID([SON])&$_\sigma$ID([HI])	MAX([NAS])	PALATALIZE	IDENT([SON])	IDENT([HI])	*Margin/{V, r, lat, nas}	*Margin/{V, r, lat, nas, obs}	*Margin/{V, r, lat, nas, obs, lar}
a.	miː[le]kaːso			*!	*						*	*	*
b.	miː[ne]kaːso				*!						*	*	*
c. ☞	miː[ni]kaːso							*		*	*	*	*
d.	miː[ʃi]kaːso					*!			*	*		*	*

Cross-linguistic variation can be modeled simply by ranking individual constraints from the markedness scale lower or higher. For example, for a language which allowed laterals in margins but not rhotics, the constraint *Margin/{V, r, lat} would be moved down past the faithfulness constraints, alongside *Margin/{V, r, lat, nas}. However there is no possible way to rerank these constraints such that the grammar violates segmental sonority prominence, which is assumed to be a universal property of phonology.

As with *e, throughout this analysis the segmental markedness scale constraints are collapsed into *[lateral] for the sake of readability. See 103 for an example using all the constraints in full.

5.4.4 Local Summary

The complex patterns of palatalization in Odawa can be derived by a complex interaction of faithfulness constraints, the PALATALIZECORONAL constraint, and a number of stringently related markedness constraints which capture universal phonetic facts about sonority and the perceptibility of the vowel space.

This analysis follows from Piggott (1980)'s assumptions about the underlying representations in Odawa. Those instances of [n] which alternate with [ʃ] are realisations of underlying /l/, while those instances of [i] which fail to cause palatalization are realisations of underlying /e/. The unusual patterns in Odawa follow from the local conjunction IDENT([SON])&$_\sigma$IDENT([HI]), which forbids changes in height and sonority within the same syllable.

The analysis demonstrates how the apparently unnatural alternations in Odawa, can be produced by a phonetically grounded phonology. Therefore the Odawa data is not evidence against phonetic grounding.

(103) /le/ ⟶ [ni]

/miːl-ekaːso/		*Margin/V	*Margin/{V, r}	*Margin/{V, r, lat}	IDENT([ROUND])	*[ROUND]&IDENT([HIGH])	*{ɪ, ʊ}	*{ ɪ, e, o, ʊ,}	*{ i, ɪ, e, a, o, ʊ, u}	IDENT([BACK])	ID([SON])&$_\sigma$ID([HI])	MAX([NAS])	PALATALIZECORONAL	IDENT([SON])	IDENT([HI])	*Margin/{V, r, lat, nas}	*Margin/{V, r, lat, nas, obs}	*Margin/{V, r, lat, nas, obs, lar}
a.	miː[le]kaːso			*!				*	*							*	*	*
b.	miː[ne]kaːso							*!	*							*	*	*
c. ☞	miː[ni]kaːso								*				*		*	*	*	*
d.	miː[na]kaːso								*	*!					*	*	*	*
e.	miː[nu]kaːso				*!			*	*	*					*	*	*	*
f.	miː[ʃi]kaːso								*		*!			*	*		*	*
g.	miː[di]kaːso								*				*	*!	*	*	*	*

5.5 Conclusions

The analyses presented in this chapter show how unnatural phonological processes can emerge in a phonetically grounded phonology. Thus the existence of such patterns cannot be considered evidence against phonetic grounding of the sort advocated here. In fact, the analyses demonstrate why the existence of an unnatural pattern can never 'disprove' phonetic grounding, in principle. To prove that an unnatural process is reducible to grounded constraints, one need only provide a single successful analysis. But to prove the opposite, that a given process is irreducible to grounded constraints, one would need to prove that no such analysis exists. This is strictly impossible, by virtue of the Induction Problem.

This presents grounded phonology with a clear model for conducting phonological analyses. Irreducibility to grounded constraints should be regarded as a null hypothesis, since it is strictly unprovable. Thus the goal of the phonologist is to disprove the null hypothesis by constructing an analysis which reduces a given phenomenon to the interaction of grounded constraints.

The Kashaya and Odawa analyses relied on the use of local conjunctions to produce the unnatural processes. The combination of grounded constraints with local conjunction makes for a powerful yet conceptually coherent phonology. The strict limitations imposed by the grounding criterion prevent the phonologist from simply inventing constraints to fit the data. Instead, under the approach advocated here, the phonologist must attempt to reduce the data to the interactions of a relatively small number of constraints which reflect a common principle. Conversely, if one allows any and all sorts of constraints, then phonological patterns are not reduced to anything, thus nothing is explained. The use of local conjunctions provides the grammar with the power to account for the unnatural processes observed in languages, whilst maintaining the criterion of phonetic grounding.

The Odawa analysis also relied on underlying segments which never surface unchanged. Tentatively, this might indicate that a phonetically grounded approach to phonology should rely more on abstract representations. Conceptually, we might think of markedness constraints as representing the limits imposed on the grammar by phonetic factors, while representations are limited by more general cognitive capacities (long-term memory, efficient computation, etc.) which are not themselves subject to any phonetic factors.

The Zuni analysis makes a potentially interesting prediction, namely that the absence of the sequence k^wɒ is because of a gap in the lexicon, rather than because it is strictly ungrammatical. This could be tested on native speakers of Zuni, to see whether their intuitions or reactions to k^wɒ are different from other unattested sequences in the language.

Chapter 6

Final Conclusions

This dissertation has argued for a modally-dependent, or grounded approach to phonology. This argument is based on the view of phonology as an evolved, biological system. Explaining such a system entails answering Tinbergen's Four Questions: function, mechanism, ontogeny and phylogeny. A grounded approach to phonology presents us with congruent potential answers to the first three.

By positing that the function of phonology is maximizing articulatory and perceptual ease, the Mechanistic Question (the synchronic study of grammar) becomes an issue of understanding how the phonology accomplishes this function. Under this view, the most interesting and revealing data is precisely when the phonology fails to do what we would expect. Much like the giraffe's recurrent laryngeal nerve (chapter 3), unnatural phonological processes are an insight into the nature and implementation of this biological system. The is the approach pursued in chapter 5, and it makes for a potentially productive research strategy. It provides the phonologist with a clear goal when performing phonological analyses, i.e. the reduction of unnatural processes to the interaction of grounded constraints. We should expect this approach to be enlightening as to the nature of phonology. The analysis in chapter 5 indicate that, within an OT framework, local conjunctions and abstract underlying representations may be an essential part of the mechanism of phonology.

Crucially, the strongest support for a modally-dependent phonology comes from confronting the Ontogenetic Question, that is, the question of how a working phonology develops from a fertilised egg. The neuroscientific findings discussed in chapter 2 indicate that cortical tissue develops under the influence of the external organs. This would imply that the most likely ontogenetic account of phonology is one in which the phonology develops under the influence of the articulatory and perceptual organs, entailing a modally-dependent phonology. The evidence for this comes from studies into the visual cortices of cats, studies into cross-modal plasticity in the blind and deaf, as well as data from lesion studies. Moreover, this is simply a more parsimonious account of how so few genes can build such complex cortical structures. The modally-dependent account removes the need for any kind of innate specification of cortical tissue, significantly reducing the burden on the genome.

While substance-free phonology, as it is presented in Hale and Reiss (2008), is arguably a more elegant answer to the Mechanistic Question, it fails entirely to provide answers to the Functional and Ontogenetic Questions. I do not suppose however, that anything in this dissertation is a death blow to substance-free phonology. What I hope has is evident, is that there are clear avenues for rebuttal by anyone advocating substance-free phonology. These include positing a function for phonology, but most importantly, providing a plausible ontogenetic account consistent with the neuroscientific literature.

By confronting the Ontogenetic Question, phonologists open the door to all manner of potentially interesting collaborations with neuroscientists. The Ontogenetic Question is not only relevant to current theories of phonology, it also presents an exciting future for the field as a whole.

References

Anderson, S. R. (1981). Why phonology isn't "natural". *Linguistic Inquiry*, *12*(4), 493-539.

Archangeli, D., Mohanan, K., & Pulleyblank, D. (2009). The emergence of optimality theory. In L. Uyechi & L. H. Wee (Eds.), *Reality exploration and discovery: Pattern interaction in language & life.* C S L I Publications/Center for the Study of Language & Information.

Archangeli, D., & Pulleyblank, D. (1994). *Grounded phonology.* The MIT Press.

Archangeli, D., & Pulleyblank, D. (2012). Emergent phonology: Evidence from English. In *Issues in english linguistics.*

Bates, E. (1999). Plasticity, localization and language development. In S. Broman & J. Fletcher (Eds.), *The changing nervous system: Neurobehavioral consequences of early brain disorders.* Oxford University Press.

Bedny, M., Pascual-Leone, A., Dravida, S., & Saxe, R. (2011). A sensitive period for language in the visual cortex: Distinct patterns of plasticity in congenitally versus late blind adults. *Brain and Language*, *122*(3), 162 - 170. Available from `http://www.sciencedirect.com/science/article/pii/S0093934X11001799`

Bermudez-Otero, R., & Börjars, K. (2006). Markedness in phonology and in syntax: the problem of grounding. *Lingua*, *116(5).*

Boeckx, C., & Uriagereka, J. (2007). Minimalism. In G. Ramchand & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces.* Oxford University Press.

Buckley, E. (1994). *Theoretical aspects of Kashaya phonology and morphology.* CSLI Publications, California.

Buckley, E. (2000). On the naturalness of natural rules. *USCB Working Papers in Linguistics*, *9*.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Massachusetts institute of technology MIT Press.

Chomsky, N. (1995). *The minimalist program.* The MIT Press.

Chomsky, N. (2009). *Cartesian linguistics: A chapter in the history of rationalist thought.* Cambridge University Press. (Original work published 1966)

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America*, *64*(1), 44-56.

Dawkins, R. (2009). *The greatest show on earth: The evidence for evolution.* Free Press.

de Lacy, P. (2006). *Markedness: Reduction and preservation in phonology.* Cambridge University Press.

Dorffner, G., Kwasny, S., & Port, R. (1987). *Parsing phonetic segments into syllables* (Tech. Rep.). Computer Science Department Indiana University. Available from http://www.cs.indiana.edu/pub/techreports/TR217.pdf

Edelman, G. M. (1993). Neural darwinism: Selection and reentrant signaling in higher brain function. *Neuron*, *10*(2), 115 - 125. Available from http://www.sciencedirect.com/science/article/pii/089662739390304A

Fitch, W. T. (2010). *The evolution of language.* Cambridge, UK: Cambridge University Press.

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

Fodor, J. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, *28*(2), 97-115.

Fodor, J. (1975). *The language of thought.* HARVARD University Press.

Fodor, J. (2008). *LOT 2.* Oxford University Press.

Freberg, L. (2010). *Discovering biological psychology* (2nd ed.). Wadsworth.

Gisiger, T., & Boukadoum, M. (2011). Mechanisms gating the flow of information in the cortex: What they might look like and what their uses may be. *Frontiers in computational neuroscience*, *5*(1).

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

Hayes, B. (1999). Phonetically driven phonology: The role of optimality theory and inductive grounding. *Functionalism and Formalism in Linguistics*, *1*, 243-285.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393-402.

Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology*, *26*(6), 994-1002. Available from http://jn.physiology.org/content/26/6/994.short

Hume, E., Johnson, K., Seo, M., & Tserdanelis, G. (1999). A cross-linguistic study of stop place perception. *Proceedings of the International Congress of Phonetic Sciences*.

Joanisse, M. F. (1999). Exploring syllable structure in connectionist networks. In *In proceedings from the 14th international congress of phonetic sciences, (icphs-99* (p. 731-734).

Kager, R. (1999). *Optimality theory.* Cambridge University Press.

Kang, Y. (2011). Loanword phonology. In M. van Oostendrop, C. Ewen, E. Hume, & K. Rice (Eds.), *Companion to phonology.* Wiley-Blackwell.

Krämer, M. (2003). Intervocalic s-voicing, geminates and the Richness of the Base in Vene Italian. *Rivista di Grammatica Generativa*, *28*, 71-85.

Krämer, M. (2006). Optimal underlying representations. In L. Bateman & C. Ussery (Eds.), *Proceedings of the thirty-fifth annual meeting of the north east linguistic society* (p. 351-365). GLSA Publications.

Lieberman, P. (2007). The evolution of human speech: Its anatomical and neural bases. *Current Anthropology*, *48*(1).

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), pp. 839-862. Available from `http://www.jstor.org/stable/411991`

MacSweeney, M., Woll, B., Campbell, R., McGuire, P. K., David, A. S., Williams, S. C. R., et al. (2002). Neural systems underlying british sign language and audiovisual english processing in native users. *Brain*, *125*(7), 1583-1593. Available from `http://brain.oxfordjournals.org/content/125/7/1583.abstract`

McCarthy, J. J. (2002). *Comparative markedness* (Tech. Rep.). University of Massachusetts, Amherst. Available from `http://roa.rutgers.edu/article/view/499`

Morén, B. (2003). The parallel structures model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory*, 194-270.

Newman, S. (1996). Sketch of the Zuni language. In I. Goddard (Ed.), *Handbook of North American Indians* (Vol. 17). Washington DC: Smithsonian Institution.

O'Connor, T., & Wong, H. Y. (2012). Emergent properties. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2012 ed.). `http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/`.

Oswalt, R. L. (1961). *A kashaya grammar (southwestern pomo).* University of California, Berkeley.

Padgett, J. (2002). *Constraint conjunction versus grounded constraint subhierarchies in optimality theory* (Tech. Rep.). University of California, Santa Cruz. Available from `http://people.ucsc.edu/ padgett/locker/conjunct.pdf`

Piggott, G. L. (1980). *Aspects of Odawa morphophonemics.* Garland Publishing.

Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones.* Routledge.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar* (Tech. Rep.). Rutgers University and the University of Colorado at Boulder. Available from `http://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF`

Smolensky, P., & Legendre, G. (2006a). *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 2). Cambridge, Massachusetts: The MIT Press.

Smolensky, P., & Legendre, G. (2006b). *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 1). Cambridge, Massachusetts: The MIT Press.

Tesar, B., & Smolensky, P. (2000). *Learnability in optimality theory.* Mit Press.

Tinbergen, N. (1963). On aims and methods in ethology. *Zeitschrift für Tierpsychologie*, *20*, 410-433.

Uffmann, C. (2007). Intrusive [r] and optimal epenthetic consonants. *Language Sciences*, *29*, 451476.

Walker, W. (1972). Toward the sound pattern of zuni. *International Journal of American Linguistics*, *38*(4), 240-259. Available from http://www.jstor.org/stable/1264302

Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, *26*(6), 1003-1017. Available from http://jn.physiology.org/content/26/6/1003.short

Wiesel, T. N., & Hubel, D. H. (1965). Extent of recovery from the effects of visual deprivation in kittens. *Journal of Neurophysiology*, *28*(6), 1060-1072. Available from http://jn.physiology.org/content/28/6/1060.short