Virtual Phonology:
Rule Sandwiching and Multiple Opacity
in North Saami

Patrik A. Bye

University of Tromsø, February 23, 2002
Four people in particular have made an inestimable contribution to this dissertation: Curt Rice, John McCarthy, Ove Lorentz, and Pekka Sammallahti.

As a supervisor and a friend, Curt has been extraordinarily kind and patient, engaged and encouraging exactly when I needed it most. Curt has overseen my development as a phonologist for almost seven years now, and I look forward to many more years of collaboration with him.

In 1998, I had the good fortune of spending a year as a visiting scholar at the University of Massachusetts, Amherst, where I received supervision from John McCarthy. It was a privilege to be able to discuss this material with him, and my debt to him, both intellectual and personal, should be evident from every page. This was an exciting time, when so many of the ideas about dealing with opacity in OT were coming together, and I consider myself immensely fortunate in having been around to witness and take part in the ferment.

Without Ove I would never have made it to Tromsø in the first place. If anyone deserves to be credited with pioneering the generative approach to Saami phonology, it is Ove. When I first arrived in Tromsø in 1991, the first thing I did was to take a couple of courses in Saami phonology from him. Well, the fascination for Saami stuck, and without his tuition and guidance, particularly in those early stages, I simply wouldn’t have known where to begin.

Over the years I’ve built up quite a lengthy correspondence with Pekka Sammallahti. Approaching the phonology of a language like Saami for the first time is daunting indeed, and Pekka’s contribution to my understanding has been nothing short of invaluable. Sammallahti’s own work represents a true feat of analysis of what is, it has to be said, the most intricate phonological system I have ever encountered. Without his pioneering publications and our correspondence on the nuances of Saami phonology, I would have been lost.

There are three other people I’d like to acknowledge with a special commendation.

Florin Oprina has generously stood by my conversion to \LaTeX and Linux. I’ve been an eager student of both, and there have been times where my eagerness has not stopped short of ruthless exploitation of his skills.

I also owe a debt of thanks to Paul de Lacy for his challenging input on many aspects of this work. Paul has made a significant impact on my linguistic thinking over the last couple of years, and been tremendous fun to work with.

In 1999, the department at Tromsø was bestowed the boon of a visit from David Odden, with whom I have had many stimulating discussions on Saami and the nature of grammar in general. His views have served as a contrasting foil for my own, and I’m sure the content of the dissertation is sharpened as a result.

Occasionally you get to take part in an event which profoundly changes the course of one's thinking. The First North American Conference on Phonology in Montréal in April 2000 was an event of this kind, and I'd like to thank Charles Reiss and Mark Hale for organizing it. Many thanks to Suzanne Levi and Ellen Kaise, who first threw down the gauntlet to Sympathy Theory which is picked up in these pages.

I've been lucky to have some excellent teachers over the years. The Department of Linguistic Science at the University of Reading provided a broad and interdisciplinary background in linguistics. In particular, I'd like to thank my first (and only) Norwegian teacher Paul Kerswill, who gave me the idea of going to Tromsø for my year abroad, as well as my first phonology teachers, Erik Fudge, Ron Brasington, Bill Hardcastle, and Linda Shockey, who share culpability for channeling my interests into phonology all those years ago.

I'd like to express my gratitude to all those who have taught me over the years and contributed to my development as a linguist, including Anders Holmberg, Ove Lorenz, David Olden, Curt Rice, and Taraki Taraldsen (Tromsø), José Ignacio Hualde and Harry van der Hulst (Girona), Ken Rice and Ken Hale (Albuquerque), John Harris, Jun'ko Itô, Paul Kiparsky, Gjert Kristoffersen, Armin Mester, and Tomas Riad (Hvalfjarðarströnd, Iceland), John Kingston, John McCarthy, and Lisa Selkirk (UMass).

Soon I will have spent ten years in Tromsø, and during that time I have had the pleasure of seeing the department grow into the vibrant and international centre of linguistic research excellence it is today. People have come and gone, and I'd just like to acknowledge a few of those who have journeyed with me some of the way. Thanks to Wilma ter Haar, Anya Hogoboom, Þorbjörg Hróarsdóttir, Omar Ikleane, Patrycja Jabłońska, Marit Julien, Bodil Kappelen-Schmidt, Svetoslav Marinov, Lucie Medová, Krzysztof Migdałski, Øystein Nilsen, Florin Oprina, Marie Safářová, Øystein Vangsnes, and Mark de Vos.

Many thanks too to my classmates at UMass for such a stimulating and enjoyable year, John Alderete, Jan Bell, Katy Carlson, Nancy Hall, Heli Harrikari, Caroline Jones, Eva Juarros, Ed Keer, Paul de Lacy, Meredith Landman, Ania Lubowicz, Elliot Moreton, Marcin Morzycki, Anita Nowak, Steve Parker, Jen Smith, Mariko Sugahara, Pius Tamani, and Hisao Tokizaki. Thanks also to Kathy Adamczyk, Lynn Ballard, and Ellen Woolford, who saw to it that my sojourn at UMass went so smoothly.

I'd also like to thank our excellent library staff, in particular Bente Stor-
vestre, for never once failing to procure material, no matter how obscure.

For computer assistance of various kinds I’d like to thank Holger Hagan, Geir Tore Voktor, Leigh Murray, and Alexander Berdnikov.

For other kinds of assistance I’d like to thank Dag Andersson, Finn Reidar Johannessen, Eirik Liland, Niels Windfeld Lund, Marit Nilsen, Unni Norum, Else Olsen, Nina Oppedal, Siri Sorland, and Toril Swan.

For all kinds of support of a non-academic nature I’d like to acknowledge my special gratitude to my family and friends, Ashley and Lisbeth Bye, Anna Bye, Paul de Lacy, Tore Nesset and Tanja Kudrijavitseva, Sarah Harding, Florin Oprina, Curt Rice, Hilde Brox, Rolv Jørgensen and Liv Dahle, Heejin Kim, Inger-Johanne Fjeldså, and my son Allan.

This research was funded by Norwegian Research Council (NFR) Grant No. 115785/520
# Contents

1 Introduction ............................................................. 1
   1.1 Opacity .............................................................. 1
   1.2 Virtual Phonology .................................................. 2
   1.3 North Saami .......................................................... 3
   1.4 Morphological Conditioning ....................................... 4
   1.5 Theoretical Alignment ............................................. 7
   1.6 Dissertation Outline ............................................. 8

I Theoretical Preliminaries ............................................ 11

2 Optimality Theory .................................................... 13
   2.1 The Basics ........................................................... 13
      2.1.1 Rules versus constraints .................................... 13
      2.1.2 The Architecture of the OT Grammar ..................... 16
   2.2 H-EVAL ............................................................... 17
      2.2.1 Constraint Interaction and Tableaux ...................... 17
      2.2.2 Factorial typology .......................................... 19
      2.2.3 Parallelism .................................................. 20
   2.3 Universality ....................................................... 20
   2.4 Richness of the Base ............................................. 25
      2.4.1 Allophony, Richness of the Base, and Robust Grammars 26
   2.5 Correspondence Theory ........................................... 28
      2.5.1 The Emergence of the Unmarked ............................ 30
      2.5.2 Positional Faithfulness ..................................... 31
      2.5.3 Lexical exceptionality ...................................... 32
   2.6 Phonological Opacity ............................................. 33
      2.6.1 Sympathy Theory ............................................. 34
      2.6.2 Local Conjunction .......................................... 39

3 Virtual Phonology .................................................... 41
   3.1 The Theory .......................................................... 41
      3.1.1 Architecture .................................................. 41
      3.1.2 Predictions ................................................... 43
   3.2 Multiple Opacity .................................................. 46
      3.2.1 Vowel Harmony ............................................... 47
      3.2.2 Vowel Shortening ............................................ 49
      3.2.3 Long High Vowel Lowering .................................. 49
CONTENTS

3.2.4 The Opaque Interaction of Vowel Harmony and Vowel Shortening ........................................ 50
3.2.5 A Virtual Phonological Account of Opacity in Yawelmani Yokuts ........................................ 53
3.3 Rule Sandwiching ........................................ 58
  3.3.1 Preglottal Vowel Shortening in Yawelmani ........................................ 58
  3.3.2 Rule Sandwiching in Modern Hebrew ........................................ 60
  3.3.3 Contextual Metrical Invisibility in Mohawk ........................................ 64
  3.3.4 Discussion ........................................ 69
3.4 Virtual Markedness ........................................ 70
  3.4.1 Nasalized Vowels in American English ........................................ 70
  3.4.2 Palato-alveolar Affricates in Abersoch Welsh ........................................ 71
  3.4.3 Ga-gyō Variation in Tokyo Japanese ........................................ 73
  3.4.4 Disjoint Metrical Tiers in Huariapano ........................................ 76
  3.4.5 Virtual feet in Cairene Classical Arabic ........................................ 79
  3.4.6 Yidin Deep Phonotactics ........................................ 80
  3.4.7 Vowel Harmony in Finnish ........................................ 88
  3.4.8 Duke of York Effects ........................................ 90
3.5 Learnability issues ........................................ 91
3.6 Residual Problems ........................................ 95

II Quantity Alternation ........................................ 99
4 Grade Alternation ........................................ 101
  4.1 Grade Alternation in Saami ........................................ 101
  4.2 Grade Alternation in other languages ........................................ 104
5 Segment Inventory ........................................ 109
  5.1 Stress assignment ........................................ 109
  5.2 Traditional Terminology ........................................ 110
  5.3 Vowels ........................................ 111
  5.4 Consonants ........................................ 112
    5.4.1 Consonants in word-initial position ........................................ 112
    5.4.2 Consonants in marginal and final position ........................................ 113
    5.4.3 Consonants in foot-medial position ........................................ 114
  5.5 Triliteral clusters ........................................ 120
  5.6 Conclusions ........................................ 122
6 Historical origins of gradation ........................................ 123
  6.1 Historical Fortition ........................................ 124
  6.2 Historical Morphologization ........................................ 127
  6.3 Four degrees of quantity? ........................................ 130
  6.4 Conclusions ........................................ 132
7 A rule-based account ........................................ 135
  7.1 Bisyllabic Stems ........................................ 136
    7.1.1 Consonant Gradation: Lenition or Fortition? ........................................ 137
    7.1.2 Epenthesis ........................................ 139
    7.1.3 Balance I: Primary Lengthening ........................................ 141
CONTENTS

7.1.4 Co da Maximization ........................................ 144
7.1.5 Balance II: Secondary Lengthening ....................... 145
7.1.6 Intervocalic Voicing and Spirantization ................. 149
7.1.7 The phonology of voiceless stops: preaspiration and ho-
morganic stop deletion ........................................ 149
7.1.8 The phonology of prestopped nasals: glottalization and
voicing alternation ............................................. 152
7.1.9 The opaque interaction of Balance and Lactic Lengthening 159
7.1.10 Preyotic shortening and rule sandwiching ............. 161
7.1.11 Other rules .................................................. 162
7.2 Quadrisyllabic stems ............................................ 164
7.3 Trisyllabic stems ................................................ 165
7.4 Morphological conditioning vs. abstract phonology ....... 170
7.5 Conclusions ....................................................... 171

8 Quantity ............................................................ 173
8.1 Segmental quantity ............................................ 174
8.2 Prosodic quantity ................................................ 175
  8.2.1 Hypercharacterization .................................... 175
  8.2.2 Hyperchrony ................................................ 177
8.3 Consonant Quantity and Vowel Length ..................... 181
8.4 Conclusions ....................................................... 184

9 Syllabication ...................................................... 185
9.1 Typology of Syllable Structure ............................... 186
9.2 Syllabication in the Strong Grade ........................... 187
  9.2.1 Epenthesi s .................................................. 187
  9.2.2 Complex onsets and the syllabication of Strong Grade
  triliteral clusters .............................................. 187
9.3 Co da Maximization in the Weak Grade ..................... 188
  9.3.1 Co da maximization in biliteral clusters ................ 188
  9.3.2 Co da maximization in triliteral clusters ............... 191
  9.3.3 Blocking of epenthesi s in Q2 weak clusters .......... 192
9.4 Conclusions ....................................................... 195

10 Segmental processes ............................................ 197
10.1 Homorganic voiced stop deletion ............................ 197
10.2 Preaspiration .................................................. 198
  10.2.1 The underapplication of preaspiration ................. 200
10.3 Prestopping and glottalization .............................. 202
  10.3.1 Prestopping ............................................... 202
  10.3.2 Glottalization and nasal release ....................... 206
  10.3.3 Clusters of liquid+nasal ................................ 208
10.4 Laryngeal Spreading ......................................... 209
10.5 Voicing and Spirantization .................................. 212
11 Consonant Gradation 213
  11.1 Consonant Gradation: Lenition or Fortition? .................. 213
  11.2 Consonant Gradation: An OT Analysis ......................... 214
    11.2.1 Morphologically conditioned grade alternation .......... 219
  11.3 Balance .................................................. 220
    11.3.1 Balance as an opacifier ................................ 220
  11.4 Balance opacified .......................................... 225
    11.4.1 Morphological blocking of Balance ........................ 226
    11.4.2 The underapplication of Balance: Uniqueness counterex-
        emplified .................................................. 227
    11.4.3 Preyotic shortening ..................................... 231
    11.4.4 The proliferation of virtual forms ....................... 233
    11.4.5 The Fear of Overgeneration ............................... 234
  11.5 The Phonology of Trisyllabic Stems ............................ 236
  11.6 Concerns about the Richness of the Base ...................... 240
  11.7 Conclusions ................................................ 242

12 Conclusion 243
Chapter 1

Introduction

This thesis has two interlocked goals. The first is to develop a theory of opacity consistent with the principles and spirit of parallelist Optimality Theory (OT). The second is to provide, for the first time, a generative account of a substantial portion of the phonology of North Saami.

1.1 Opacity, Multiple Opacity, and Rule Sandwiching

A central challenge to any generative theory of phonology is to reconcile the existence of two apparently incommensurable forms of process interaction, conspiracy and opacity. The two great theoretical alternatives within generative phonology, derivational (rule-based) theory and Optimality Theory, are almost defined by the analytical success each have had in dealing with these kinds of process interaction.

While Optimality Theory has had resounding success in dealing with the problem of conspiracies, rule-based theory has been most notably successful in its treatment of opacity.

Nevertheless, rule-based theory has had little to say on the subject of conspiracies, and Optimality Theory has yet to match the success of rule-based approaches to opacity, at least when under the constraint of full parallelism.

The term conspiracy was first introduced into linguistic theory by Kisselcherth (1970) to refer to a situation in which two or more formally independent processes 'conspire' to produce a specific surface configuration (cf. Haiman 1972). It is now generally agreed that the Conspiracy Problem is in principle solved. Nevertheless, OT has been rather less successful to date in dealing with the Problem of Opacity.

Opacity refers to an interaction between two processes in which one process obscures or opacifies the conditions in which the other process applies.

Opacity comes in two main forms, and these are best illustrated considering a derivation consisting of two interacting processes, \( P \) and \( Q \), where \( P \) is ordered prior to \( Q \). \( Q \) may either obliterate the environment in which \( P \) applies (in which case \( Q \) 'counterbleeds' \( P \)), or \( Q \) may create the environment for the application of \( P \) (in which case \( Q \) 'counterfeeds' \( P \)). In the first case, we say that \( P \) overapplies, since \( P \) applies although the conditions for its application are not met on the
CHAPTER 1. INTRODUCTION

Surface (non-surface-apparent): it looks like $P$ shouldn't have applied, but does so all the same. In the second case, we say that $P$ under applies, since $P$ does not apply even though the conditions for its application are met on the surface: $P$ looks as though it should have applied but does not.

Rule-based theory has been remarkably successful at dealing with interactions of this kind. Indeed, there is nothing anomalous about the phenomenon of opacity on derivational assumptions. It is, rather, expected, and emerges as a mere epiphenomenon of the ordering of rules.

OT's output-orientation and parallelism puts the theory at a distinct disadvantage in dealing with phonological generalizations which are non-surface-true or non-surface-apparent. Indeed, the success of OT in this domain has been comparatively modest seen measured against the successes of rule-based theory. Opacity is still a true challenge to OT's parallelism.

This challenge has made opacity a hot research topic in OT in the last few years. Sympathy Theory (McCarthy 1999) represents a watershed in our understanding of how opacity can be integrated into a fully parallelist OT. Nevertheless, there remain significant problems with Sympathy Theory, both conceptual and empirical, and at least two kinds of three-rule opaque interaction remain beyond its scope. Specifically, Sympathy Theory fails to predict the existence of multiple opacity. Multiple opacity is of two kinds, and, ultimately, neither of them can be handled by Sympathy.

Multiple interacting opacity involves the interaction of at least three rules and the opacification of at least two of these rules, $P > Q > R$, such that $R$ opacifies (counterfeeds or counterbleeds) $Q$, and $Q$ opacifies $P$.

Multiple non-interacting opacity is not been previously addressed in the literature explicitly, but it is apparently instantiated in Saami. Multiple non-interacting opacity involves the opacification of two rules, but these rules do not interact. It is exemplified by a four-rule derivation where $P$ is opacified by some rule $R$, and another rule $Q$ is opacified by some rule $S$, but in which $P$ and $Q$ do not themselves interact and so are not ordered with respect to eachother.

Sympathy Theory also fails to generate a specific three-rule interaction called rule-sandwiching. Rule-sandwiching has the form $P > Q > R$, where, crucially, $P$ and $R$ introduce identical faithfulness violations, and where $P$ bleeds/feeds $Q$, but $R$ counterbleeds/counterfeeds $Q$. McCarthy denies that rule interactions of this type exist, but they turn out to be quite well attested, for example in Yawelmani Yokuts, Modern Hebrew, Mohawk, Yiddish and Saami. Saami in particular is rich in rule interactions of this type. Of course, neither of these forms of interaction are problematic within a derivational approach. This dissertation aims at bringing these phenomena into the purview of a fully parallelist OT.

1.2 Virtual Phonology

An alternative theory of phonological opacity, Virtual Phonology, is developed, which is at once fully compatible with OT's parallelism and addresses the empirical defects just mentioned. The claims of Virtual Phonology are enumerated below:

1. The output is a binary array, or pair of representations consisting of a unique surface form ($S$) and a unique virtual form ($V$). The $S$-form encodes overt structure, while the $V$-form encodes covert structure.
2. The role of the $\mathcal{V}$-form is to encode, transparently, those phonological generalizations which may be non-surface-true or non-surface-apparent. The $\mathcal{V}$-form thus covers the same functions as the sympathetic ($\bigstar$) candidate in Sympathy Theory.

3. Unlike Sympathy Theory, which admits of the selection of multiple $\bigstar$-candidates, the $\mathcal{V}$-form is unique, at least as a null hypothesis. While the $\bigstar$-candidate may encode at most a single recuperatum, multiple opacity in Virtual Phonology involves multiple recuperata all distributed throughout the unique $\mathcal{V}$-form.

4. The $\mathcal{V}$-form is subject to constraints on $\mathcal{V}$-Faith. These play the role of the selector ($\bigstar$) constraints of Sympathy Theory. Unlike selector constraints, which, for the purposes of sympathetic selection, always pattern as undominated, $\mathcal{V}$-Faith is freely rerankable.

5. The $\mathcal{S}$-form is subject to constraints on $\mathcal{S}$-Faith. In this regard, Virtual Phonology does not differ in essence from the original version of Sympathy Theory which employed constraints on intercandidate faithfulness.

From points 3 and 4 flow two predictions which crucially serve to distinguish Virtual Phonology from Sympathy Theory.

From 3, Virtual Phonology predicts the possibility of multiple non-interacting opacity, which I argue is attested in West Finnmark Saami. Uniqueness is ultimately rejected, in §11.4.2, although it is convenient enough to serve as a point of departure for discussing the next point.

From 4, Virtual Phonology predicts the existence of rule sandwiching. Because $\mathcal{V}$-Faith may be ranked anywhere relative to markedness, finer control over what goes into the virtual form is the result.

### 1.3 The Alternation of Grade and Quantity in North Saami

The body of the thesis is dedicated to a thorough elucidation of the grade and quantity alternation system of the West Finnmark dialect of North Saami,\(^2\) a Finno-Ugric language spoken in Northern Norway, Sweden, Finland, and the Kola Peninsula in Russia.

North Saami is in fact relatively well-documented, supported by a rich research tradition with roots as far back as the nineteenth century (Setälä 1896, 1912a,b; Wiklund 1896, 1913, 1915, 1919; Nielsen 1902, 1905, 1913, 1926; Lagercrantz 1927, 1929; Sammallahti 1971, 1977, 1984, 1988a,b; Korhonen 1988a; Magga 1984).

Nevertheless, data from Saami has played a less than negligible role in the development of generative theory despite the inherent interest and importance of the phenomena which adorn its phonological system.

---

1. The term *recuperatum* refers to any element of phonological representation which is absent on the surface but present opaquely. *Virtual property* is synonymous.

2. Saami was formerly known as 'Lappish', although recent scholarship has rendered the latter term increasingly obsolete.
The reasons for this lack of coverage are probably complex. What material there is has remained relatively inaccessible to phonologists working from within generative theory, much of it being written in lesser-known languages such as Finnish and Norwegian. Another reason may be the daunting complexity of both the phonological phenomena themselves, as well as the Finno-Ugric system of transcription used in much important descriptive work. What confronts the novice is frequently a welter of unphonemicized and unsystematized data. A particular problem is the transcription of quantity, of which many dialects of Saami have three phonologically contrastive degrees. In the transcriptions with which one often has to work, quantity is frequently overdetermined. The narrow phonetic transcription of many descriptions distinguishes up to five degrees of quantity. Phonetic duration is a complex function of the interaction between segmental quantity contrasts and higher-level prosodic factors, and, despite the accuracy to which Finno-Ugric transcriptions aspire, there is a deeply frustrating lack of sensitivity to the distinction between the two.\footnote{Sammallahti (1998b: 193) provides an accessible overview of the FU transcription system. Posti and Itkonen (1973) is a collection of proposals aimed at simplifying the system in various ways.}

Sammallahti’s work (Sammallahti 1971, 1977, 1984, 1998a) is a major leap forward in the phonological analysis of North Saami, because it gives us, probably for the rst time, an attempt at a \textit{phonemicization} of North Saami, laying bare the underlying system of quantity contrast.

The phonology of quantity and quantity alternation is strikingly complex, and provides an exciting testing ground for phonological theories of opacity. The alternation of quantity constitutes the main empirical focus of the dissertation.

Several dialects of Saami boast a typologically marked three-way length distinction in consonants: short, long, and ‘overlong’. These ‘quantities’ interact in a process of Grade Alternation, or Consonant Gradation. Examples of the alternation are given in (1). Geminates alternate with short consonants, and ‘overlong’ geminates alternate with plain geminates preceding a syllable closed by a suffix.

(1) \textit{Grade Alternation in North Saami}

\begin{tabular}{l l l l}
nammaa & ‘name’ & namaas & ‘in a name’
\end{tabular}
\begin{tabular}{l l l l}
tsammaa & ‘kiss’ & tammaas & ‘in a kiss’
\end{tabular}

1.4 Morphological conditioning and language-specific restrictions on the input

Opacity is intimately connected with two other kinds of anomaly, and ultimately OT will have to deal with both of them. They are (i) morphological conditioning, and (ii) language-specific restrictions on the input. Neither of these have received particular attention in the literature on OT. The rst problem, that of morphological conditioning, is probably the one which most immediately cries out for attention.

Morphologized alternations result when the original trigger or blocker of the process underlying some alternation is lost to historical erosion. For example,
1.4. MORPHOLOGICAL CONDITIONING

historically, certain fortition-blocking suffixes were apocopated, but the stem alternation they induced remained. For example, Proto-Saami $\text{*}\text{kuollee}$～$\text{*}\text{kuolee}$-n ‘fish: NOM, SG～GEN, SG’ becomes in Modern North Saami $\text{kuollii}$～$\text{kuollii}$. The standard way to deal with this kind of phenomenon has been to assume that the trigger/blocker of the process is some kind of morphological entity. In this case, that entity is plausibly the morphosyntactic environment: GENITIVE SINGULAR.

This dissertation breaks with the view that alternations such as these are morphologically conditioned, and proposes to view them, instead, in abstract phonological terms, i.e. as phonologically conditioned by some abstract phonological element that does not itself appear in the surface form because it undergoes absolute neutralization, in this case to $\emptyset$.

The idea is essentially to bring precisely the same apparatus used in the analysis of phonological opacity to bear on morphological conditioning. There is, of course, no strictly formal let or hindrance to this move. It is obvious that morphological conditioning can be reconstrued as no more than a form of phonological opacity, albeit of a special kind. The question is whether it should.

It is argued that morphologization not only can be, but should be accounted for using the tools independently motivated for dealing with phonological opacity. The pay-off is increased restrictiveness, since the need to posit constraints with simultaneous access to both phonological and morphosyntactic category information is obviated.

The move does raise its problems, though, and these are not to be taken lightly. Whereas, in cases of phonological opacity, the relevant phonological trigger/blocker can be recovered from elsewhere in the paradigm (i.e. it surfaces in another form with which the opaque form alternates), the abstract phonological trigger/blocker of a ‘morphologized’ alternation cannot be recovered in this way.

So, while reducing morphological conditioning to phonological opacity may be worthwhile in terms of the overall restrictiveness it brings, we have to make sure that this restrictiveness is not purchased at too high a cost: psychological reality is bound to catch up with us sooner or later. If our formal models purport to describe a psychologically real grammar, then it has to be the case that actual speakers treat morphologically conditioning on a par with plain old phonological opacity. Speakers have to be able, in a sense, to ‘reconstruct’ the relevant trigger/blocker as a phonological one, and, moreover, imbue it with the right phonological properties.

At first glance, this claim would appear to involve imputing miraculous powers of representation to the learner — after all, learners do not have access to the history of their language. So, how can a speaker perform the necessary feat of reconstruction, if the relevant phonological trigger/blocker is exceptionlessly absent in the stimulus?

An answer is developed to this question which draws on recent work, in particular by Hayes (1996, 1999a), on the learning of OT constraints. Hayes proposes a radical break with the innatist tradition that has informed work in generative phonology since its inception, and proposes instead that constraints are induced from phonetic experience.

Hayes’ empiricism is put to novel use here in dealing with morphological conditioning. Although Hayes does not develop the idea, his work paves the road to a major philosophical distinction between constraints and phonological generalizations that is entirely obscured on a rule-based approach to grammar.
In the rule-based approach, phonological generalizations amount, essentially, to the rules that describe them. The rules are, in a sense, the very 'stuff' of the phonological generalizations they describe. In contrast, Hayes' empiricism adumbrates a sundering of the phonological generalization, seen as a description over a set of data, and the constraints whose interaction generate the observed patterns. Put differently, constraints are not themselves the stuff of phonological generalizations. Whereas rules may be said to describe the data, it would be inappropriate to say that the interaction of the constraints describe it in the same way. It would be more accurate to say that phonological generalizations in the data emerge or are generated by constraint interaction. The relationship between constraints and generalizations in the data is thus more tenuous than that between generalizations and rules. Constraints are in a sense 'inward-looking', and represent not so much knowledge of a specific language, but functional knowledge of how the articulatory and perceptual systems work. It is precisely this knowledge which, it is argued here, is brought to bear on the task of reconstructing abstract phonological triggers in cases of morphologization. Language-users are not only in possession of essentialist knowledge about the terms of the phonological system: they also have access to a fund of functional knowledge about these terms. Tacit knowledge combines knowledge of what something is as well as what it does.

Let us return to the issue of language-specific restrictions on the input. Much emphasis is placed throughout on the OT heuristic of the Richness of the Base. Nevertheless, there seem to be examples in the literature of 'rogue' inputs which are very difficult to deal with, given the available apparatus. The problem is invariably exacerbated in the context of some phonological alternation. In these cases, it can appear tempting to impose arbitrary language-specific restrictions on what the grammar may accept as an input. By Richness of the Base, however, no grammar may be said to be complete unless it can account for all of the possible inputs to the grammar. Rogue inputs are essentially those which diagnose a grammar's incompleteness, by returning ungrammatical expressions when fed through the evaluation mechanism. What is the appropriate way of dealing with the problem?

Richness of the Base is essentially OT's negative heuristic, in the sense of Lakatos (1970), in that it directs the mode of inquiry away from the possibility of language-specific restrictions on the input to the output constraints and their interaction. Richness of the Base is apparently 'counterexampled' by rogue inputs. However, given the status of Richness of the Base as a heuristic assumption as opposed to a falsifiable hypothesis, it is inappropriate to argue for the existence of language-specific restrictions on the basis of 'counterexamples'. The gaps must be explained, not stipulated.

Some of these cases can be addressed reasonably simply within Virtual Phonology. Indeed, for static phonological generalizations, the problem of rogue inputs is in principle solved. Dynamic phonological generalizations (i.e. those involving alternations), on the other hand, are more problematic, and limitations of space preclude the development of a fully worked out theory of how do deal with these here.

Whenever constraints on the input loom, opacity is implicated. In American English, for instance, it is opacity in the surface distribution of vowel nasalization that raises the spectre of language-specific restrictions in the first place. The surface distribution of vowel nasalization involves the interaction of two
phonological generalizations: (i) vowels are nasalized preceding a tautosyllabic nasal (e.g. /seitn/ → [sē̃t]), and (ii) nasals delete preceding a tautosyllabic voiceless stop (e.g. /seint/ → [sē̃]).

In the surface form of saint, these two generalizations interact opaquely, since the nasal deletes but the vowel nasalizes regardless, in the absence of a surface trigger. But if nasalized vowels can appear on the surface without a following nasal, as in [sē̃t], how do we capture the fact that *sē̃ isn’t a possible word in this variety of English? At this point, it seems the only way to obtain this gap is by assuming a restriction on the input, specific to American English, that input vowels cannot bear the feature [+nasal], in violation of the Richness of the Base. The dilemma is resolved by invoking the distinction made in Virtual Phonology between the surface form and the virtual form. The virtual form can be invoked to function as a repository for non-surface generalizations of this kind. Specifically, it is argued that apparent restrictions on the input are explained by invoking the activity of a set of virtual markedness constraints, whose jurisdiction is restricted to virtual forms.

The counter-instances furnished by languages such as Welsh (§3.4.2) and North Saami (§11.6) are more serious, because they involve dynamic phonological generalizations. It is hoped that the tools of Virtual Phonology will eventually be brought to bear on such cases as well.

1.5 Theoretical alignment of the dissertation

This proposal is made against the backdrop of a growing disenchchantment with OT. The problems with Sympathy Theory have led to calls from some quarters for a return to rule-based theory. This is inappropriate for at least two reasons. First, a return to derivationalism would be to neglect OT’s significant theoretical gains in dealing with conspiracies. Second, it assumes, I think erroneously, that OT’s heuristic power has exhausted itself. Still, the last two or three years have seen the development of what may eventually amount to a significant critique of OT. Some of this work is explicitly anti-OT and advocates, in essence, a return to rule-based theory, in which opacity was dealt with unproblematically, and restrictions on the input were no issue. Examples of this work include Idsardi (1997), Kaise (2000), Levi (2000), and Odden (2000). All of these writers have been concerned to show that OT’s treatment of opacity (specifically, Sympathy Theory) brings pathology in its wake.

Another line of phonological inquiry, represented by Hale and Reiss (1998, 2000), Dolbey and Hanssón (1999a,b), and foreshadowed in the work of Anderson (1981) and Lass (1984), takes issue with prevailing approaches to incorporating phonetic naturalness into synchronic descriptions. OT’s universalism has foregrounded the functional basis of constraints, and, while not strictly an entailment of OT, it is certainly the case that OT has in practice engendered a pronounced functionalist bias. Analysts working within the OT framework thus by and large incorporate considerations of phonetic naturalness into the design of the synchronic grammar itself. Grammars are standardly described in terms of constraints which refer directly and in detail to preferred states of affairs on various articulatory and acoustic dimensions, a practice memorably dubbed ‘substance abuse’ by Hale and Reiss.

This substantive approach is most obviously and explicitly represented in the
work of the UCLA school (sometimes referred to as 'West Coast Functionalism'),
in particular Flemming (1995), Hayes (1996, 1999a), Kain (1995), Kirchner

For researchers such as Hale & Reiss, and Dolbey & Hansson, though, the
synchronic grammar is precisely the wrong place to incorporate such consider-
ations of naturalness. They argue, instead, that the synchronic grammar is
functionally blind, and that naturalness in synchronic grammars is to be at-
tributed to factors operating in diachronic change and language acquisition.
Their approach is accordingly informed by an interest in dysfunctional gram-
matical patterns.

Dysfunctional patterns, of which morphological conditioning is a prime ex-
ample, are indeed an anomaly for OT, just like opacity and apparent language-
specific restrictions on the input. However, it is argued that they may be ac-
counted for without doing violence to the principle that OT constraints are thor-
oughgoingly functionally motivated, provided that inputs are permitted to be
sufficiently abstract. I indicated earlier that functionalism suggests an answer to
the problem this poses to learnability and the psychological plausibility of gram-
mars incorporating abstractness: dysfunctional patterns may be reconstrued in
functional terms assuming that learners have access to functional knowledge.

This dissertation thus addresses three kinds of phenomenon which are anom-
alous given key OT heuristics. Output-orientation and parallelism are initially
incompatible with the existence of opacity. *Richness of the Base* is incompatible
with the existence of language-specific restrictions on the input. Morphological
conditioning, and dysfunctional phonology in general, is incompatible with OT's
universalism and functionalist bias.

This dissertation takes all of these heuristic principles seriously, and sets out
to show that novel insights into the architecture of the grammar can be obtained
by focusing precisely on the counter-instances to these principles.

While not entirely a complete solution to these problems, if this serves to
inspire further work on these issues, then this dissertation will have achieved its
aim.

1.6 Dissertation Outline

The remainder of this dissertation consists of two main parts. The first part
consists of two chapters.

Chapter 2 is an introduction to the basic tenets and analytical techniques of
Optimality Theory. It also covers Sympathy Theory, the best known extension
of the theory for dealing with opacity.

Chapter 3 introduces the proposal, Virtual Phonology. This chapter includes
an extended critique of McCarthy's own Sympathy-theoretic analysis of Yawel-
mani Yokuts, and shows how Sympathy Theory breaks down when it comes to
dealing with multiple opacity. The remainder of the chapter adduces evidence
in favour of Virtual Phonology from the phenomenon of rule-sandwiching in
various natural languages.

The second part is dedicated to an analysis of the system of grade and
quantity alternation in North Saami. It is structured as follows.

Chapter 4 gives a brief overview of the phenomenon, and places Saami Grade
Alternation in its context amongst other Uralic languages with similar phenom-
Chapter 5 presents the rhythmic structure of North Saami, and ties this in with the segment inventories of the language. North Saami is a language in which metrical position impinges deeply on the possibilities for segment contrast, and metrical structure is also key in understanding the alternation of grade and quantity.

Chapter 6 examines the historical origins of Consonant Gradation. We assume and develop the basic position espoused by Sammallahti (1998a,b) that Consonant Gradation is, historically speaking, fortition. We develop this point as synchronically true, as well.

Chapter 7 is a rule-based treatment of Grade Alternation and the processes, segmental and quantitative, with which it interacts. This chapter is included essentially because of its descriptive utility as a foil to Chapter 11, which address the same interactions from the standpoint of an Optimality-theoretic framework.

Chapter 8 considers the representation of quantity in North Saami.

Chapter 9 deals with syllabification from a constraint-based perspective.

Chapter 10 examines segmental processes with which Consonant Gradation interacts from an Optimality-theoretic point of view.

Chapter 11 is a constraints-based and Virtual-phonological account of Grade Alternation in North Saami.

Chapter 12 concludes with a summary of the major results of the dissertation and suggests directions for future research and the resolution of outstanding issues.
Part I

Theoretical Preliminaries
Chapter 2

Optimality Theory

The theoretical framework for this dissertation is Optimality Theory (OT) as developed by Prince and Smolensky (1993), McCarthy and Prince (1993a), and many others.\(^1\)

In addition to serving a tutorial function for readers with little exposure to OT, this chapter also aims to locate the current dissertation with respect to current theoretical concerns. The layout of the chapter is as follows.

§2.1 presents the basic architecture and tenets of OT, and discusses the notion of ‘conspiracy’, which provides so much of the impetus for conducting phonological research within an OT framework in the first place.

§2.2 introduces constraint interaction, and equips the reader with the necessary background for interpreting OT tableaux.

§2.3 addresses the OT tenet of universality and functional grounding of constraints, and proposes that only functionally grounded constraints be admitted to the grammar. The ramifications of this principle for dealing with morphologically conditioned phonological alternations are discussed.

§2.4 examines the principle of the Richness of the Base, the idea that there are no language-specific restrictions on the input.

§2.5 turns to Correspondence Theory and Faithfulness.

Finally, §2.6 addresses the treatment of opacity within OT, and reviews Sympathy Theory (McCarthy 1999).

2.1 The Basics

2.1.1 Rules versus constraints

In OT, the grammar is essentially a mechanism for resolving conflicts between competing constraints. The grammar consists in an ordered set (hierarchy) of violable, but universal constraints on output representations. OT differs substantially in this regard from rule-based phonological theory (a.k.a. derivation-alism, or operationalism), in which outputs are derived in serial fashion by an ordered sequence of rewrite rules, the vast majority of which are context-driven. A standard phonological rule describes a generalization in terms of a structural

\(^1\) Much of this material is available on-line on the Rutgers Optimality Archive at http://ruccs.rutgers.edu/roa.html
change in some structural description. For example, rule (2) applies to change input B to output D whenever the string ABC is encountered.

(2) $B \rightarrow D/A_C$

A notational variant of (2) is the structural change shown in (3) below.

(3) $ABC \rightarrow ADC$

Elucidating the principles of interaction between rules has formed much of the basis of phonological research during the last twenty or thirty years. Nevertheless, one form of rule interaction which has remained outside the purview of the rule-based theory is the conspiracy. It has long been known that rules may 'conspire' to achieve some target output form Kisseberth (1970).\(^2\)

A subtype of the Conspiracy Problem is the Duplication Problem (Kenstowicz and Kisseberth 1977: 136ff.). Rule-based theory recognizes a dichotomy between Morpheme Structure Conditions (MSC's) and phonological rules. MSC's apply directly to the lexicon, and have isolated morphemes as their domain of application. Phonological rules, on the other hand, are triggered by concatenation of morphemes. The consequence of maintaining an analytical division between MSC's on the one hand and phonological rules on the other is that certain phonological generalizations remain unexpressed.

For example, the Yawelmani vowel system /i a o u/ is characterized by a MSC which supplies the redundant feature [+back] for any vowel which is lexically specified as [+round]. However, this lexical redundancy is also duplicated by the grammar.

A phonological rule of Vowel Harmony converts an underlying suffixal /i/ to /u/ following a root whose vowel is /u/. The vowel harmony rule involves the spreading of [+round] from the root vowel onto the vowel of the suffix. However, the output of Vowel Harmony is not the high front rounded vowel [u], as might be expected: the actual pronunciation is [u]. By maintaining this division between MSC's and repairs, we are forced to state the redundancy [+round]→[+back] as a condition not only on lexical items but on the output of phonological rules as well, duplicating information already contained in the MSC. This misses the key generalization that all rounded vowels are [+back] in Yawelmani, whether they are underlyingly [+round] or derived as such by Vowel Harmony.

Kenstowicz and Kisseberth (1977) make a similar point for Chamorro. Indeed, examples of this kind can be multiplied, and they all serve to point to one conclusion: that the constraints in force on 'underlying' representations are the same as those operating on the outputs of phonological rules. The division between MSC's and phonological rules is unable to capture this insight.

The broader Conspiracy Problem (Kenstowicz and Kisseberth 1977: 142ff.) is the observation, originally due to Kisseberth (1970), that different rules may "conspire" to yield outputs of a certain shape. In Yawelmani, for example, several rules conspire to avoid consonant clusters in one way or another. The rules themselves are however formally diverse. One morphophonemic rule of epenthesis applies to break up clusters of three consonants, while another rule deletes an initial laryngeal consonant (? or h) from certain suffixes when they follow a verb stem ending in a consonant cluster. Both of these rules invoke...
a constraint militating against triliteral clusters on the surface to trigger some process when a CCC cluster would otherwise result. The configuration is also implicated in the blocking of a third process: syncope is blocked from applying just in case it would generate a three-consonant cluster, and so is restricted to applying in a "double-sided" open syllable, i.e. in the environment VC.CV. The same constraint may thus be active in triggering some processes while blocking the application of others.

The 1980's saw the development of rule-based approaches to phonology which placed an ever greater emphasis on constraints on output form. In the theory of Constraints and Repair Strategies (Paradis 1988), for example, a rule can be blocked from applying if it yields an output violating some output constraint. In OT, though, the break with rules is complete, and the descriptive burden shifts entirely to the constraints themselves.

Constraints express 'desired' states of affairs along a wide variety of dimensions, such as perception, articulation, learnability, economy, and so on. Constraints are often inherently conflicting, and as a consequence rampant violated in output forms. Certain constraints therefore have to be accorded priority over other constraints. In OT, priority is formalized by organizing the constraints on a language-particular basis into a strict hierarchy of dominance. Because of their defeasibility, OT constraints capture generalizations which are not necessarily surface-true. Surface-trueness is sacrificed in favour of maintaining maximal generality of constraint formulation.

Most implementations of OT distinguish between two broad sets of constraints: markedness (or structural well-formedness) constraints on the one hand, and faithfulness constraints on the other. Markedness constraints penalize particular phonological configurations or structures. Their general form is *X ('Do not have X'), where X is some aspect of structure. Markedness constraints are frequently referred to generically as *Struc — 'structure is disallowed'.

Left to their own devices, markedness constraints would conspire to ensure every input surfaces as a maximally reduced [∅] (or simply ∅ (since all aspects of structure are penalized by some markedness constraint. The force of the markedness constraints is held in check in natural languages by an opposing set of faithfulness constraints which enforce the preservation of input structure. Faithfulness constraints are of the form Faith-X ('Be faithful [i.e. preserve] input X').

The SPE-style rule (Chomsky and Halle 1968) essentially amalgamates the dimensions of markedness and faithfulness into a single rule, the markedness statement corresponding to the structural description, the faithfulness aspect corresponding to the structural change (or repair). In OT, however, the repair is undetermined by the constraint, thereby achieving a high degree of generality. For example, in the case of the rewrite rule given in (2), OT would simply posit a constraint *ABC ('Do not have ABC'). This formulation underdeter-
mines the specific repair (B→D) instantiated in (2), since satisfaction of this constraint is consistent with all of the following repairs: B→D, A→E, C→F. The choice of repair is determined by the ranking of the other constraints. The point can be further illustrated by a ‘cross-linguistic conspiracy’.

Both English, Norwegian, and Swedish prohibit the cluster -sm word-finally, yet have different ways of negotiating a word-final -sm in the input. Given an input such as /sarkasm/, Norwegian epenthesizes word-finally to give sarkasm, while English epenthesizes medially, giving sarkasm. Swedish demonstrates a third option and devoces the final sonorant, giving sarkasm. What rule-based approaches fail to capture is that all three strategies share the same functional unity — the avoidance of the same marked reverse-sonority cluster "sm]. The choice of repair devolves onto other constraints. Swedish prefers not to epenthesize. English and Norwegian go down the road of epenthesi, but the outcome in each case is different. In Norwegian, the preservation of the underlying adjacency relations in /-sm-/ is purchased at the cost of allowing a consonant cluster on the surface. The English strategy avoids the consonant cluster, but sacrifices the preservation of the underlying adjacency of the segments in the string /-sm-/.

2.1.2 The Architecture of the OT Grammar

The architecture of OT consists of three major components (adapted from McCarthy and Prince 1994: pg.4):

(4) Components of the OT Grammar

- **CON.** The structured set of constraints out of which language particular grammars are constructed.

- **GEN.** A function defining, for each possible input $\text{In}_k$, the range of candidate structural descriptions (parses) available to $\text{In}_k$.

- **$H$-Eval.** A function that comparatively evaluates sets of forms with respect to a given constraint hierarchy $H$, which is a ranking of CON.

The input to the grammar itself is universal and not subject to any language-specific restrictions. This principle is known as Richness of the Base, which we shall return to examine in considerable detail in §2.4.

**GEN** is defined as a function which defines the range of possible structural descriptions for each input. By “freely exercising the basic structural resources of the representational theory”, the generation function **Gen** takes some input (In$_k$) and generates an infinite set of candidate output structural descriptions (Prince and Smolensky 1993: pg.4). Thus, Gen can add syllable nodes, epenthesize, delete segments and change features. Gen can in principle supply a candidate with syllabic, moraic, and other prosodic nodes, ranging from no structure at all to indefinite amounts of superfluous structure. There are no limits on the operations Gen can perform, beyond the parameters set by the representational theory (for example generate structures where moras dominate feet). This feature goes under the name of freedom of analysis. The candidate forms thus generated are submitted to a second subfunction, $H$-Eval. (for ‘harmonic evaluation’), which evaluates the candidates for relative harmony against a set of
ranked, *violable* constraints. Evaluation produces a *harmonic ordering* on the set of candidate forms. The unique optimal candidate, the actual output, is the one which best satisfies the ranked *hierarchy* of constraints.

\[ \text{Structure of Optimality-theoretic grammar} \]
\begin{align*}
    \text{a. } & \text{Gen} (\text{In}_k) & \rightarrow & \text{Cand}_k = (\text{Out}_1, \text{R}_1), (\text{Out}_2, \text{R}_2), \ldots \\
    \text{b. } & \mathcal{H}\text{-Eval(In}_k, \text{Cand}_k) & \rightarrow & \text{Out}_{\text{real}}
\end{align*}

Below I summarize the basic tenets of OT, drawing on McCarthy and Prince (1994) and Prince and Smolensky (1993).

\[ \text{Principles of Optimality Theory} \]
1. **Universality**
   Universal Grammar provides a set $\text{Con}$ of constraints that are universal and universally present in all grammars.

2. **Violability**
   Constraints are violable; but violation is minimal.

3. **Ranking**
   Constraints of $\text{Con}$ are ranked on a language particular basis; the notion of minimal violation is defined in terms of this ranking. A grammar is a ranking of the constraint set.

4. **Inclusiveness**
   The constraint hierarchy evaluates a set of candidate analyses that are admitted by very general considerations of structural well-formedness.

5. **Richness of the Base**
   There are no restrictions on the input.

6. **Parallelism**
   Best-satisfaction of the constraint hierarchy is computed over the whole hierarchy and the whole candidate set. There is no serial derivation.

We have already discussed the notion that OT constraints are *violable*. *Inclusiveness* refers to GEN’s freedom of analysis. *Richness of the Base* and *Parallelism* will be discussed in greater detail in sections §2.4 and §2.2.3 respectively.

### 2.2 $\mathcal{H}$-EVAL

#### 2.2.1 Constraint Interaction and Tableaux

Let us consider how constraints interact in $\mathcal{H}$-EVAL. Constraints are ranked in a language-particular hierarchy, and optimality is computed in parallel fashion, against the entire hierarchy.

Consider two constraints $\mathcal{A}$ and $\mathcal{B}$, and two candidates $\text{cand}_1$ and $\text{cand}_2$. Let us suppose that $\text{cand}_1$ violates $\mathcal{A}$ and satisfies $\mathcal{B}$, and $\text{cand}_2$ violates $\mathcal{B}$ and satisfies $\mathcal{A}$. The information is summarized in the tableau in (7). Violations are shown with an asterisk (*) in the relevant cell — candidates are literally...
‘marked’ with respect to the constraint they violate. Constraint satisfaction has no special designation. Normally this will be shown simply by leaving the relevant cell blank, although, where crucial a check (✓) may be used.

(7)

<table>
<thead>
<tr>
<th></th>
<th>A : B</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>cand₁</td>
</tr>
<tr>
<td>b.</td>
<td>cand₂</td>
</tr>
</tbody>
</table>

In cases of conflict, the ranking between A and B can only be established by determining which of the candidates cand₁ and cand₂ is the actual surface form. Suppose that cand₁ is optimal, and cand₂ suboptimal. Given this state of affairs, we have a ranking argument for the relative priority of A and B: B crucially dominates A. The tableau in (8) incorporates this new information on optimality and constraint domination.

(8)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>✽</td>
<td>cand₁</td>
</tr>
<tr>
<td>b.</td>
<td>cand₂</td>
<td>*</td>
</tr>
</tbody>
</table>

Crucial domination is shown graphically by left-right precedence in the top row of the tableau and a single solid vertical line dividing the relevant columns.

There are three additional conventions which are made use of in constructing tableaux which are by now standard practice. The optimal candidate is indicated with a pointing hand (✽). The exclamation mark (!) marks the point at which a violation is fatal, i.e. where the candidate is eliminated from the contest. Finally, shading is a typographical aid used to indicate that how the candidates fare on certain portions of the constraint hierarchy is irrelevant—usually because there is already a clear winner. Thus, in (8), it doesn’t matter whether cand₁ violates A or not, since cand₂ has already scored a fatal mark on a higher ranked constraint (B). A related point goes under the slogan of the strictness of strict domination. In the above tableau, the number of violations on the lower ranked constraint A is irrelevant. No number of violations on A by cand₁ would be sufficient to boost cand₂’s relative harmony over cand₁. There are no trade-offs between constraints: domination is absolute.

In many cases, constraints simply do not conflict, and in this case precedence is irrelevant. Where conflict is absent, or the dominance relation between two conflicting constraints has not been established due to lack of evidence, the lack of any dominance relationship is shown by a dotted line. For example, the following violation profile in (9) does not provide any ranking argument for ordering A and B.

(9)

<table>
<thead>
<tr>
<th></th>
<th>A : B</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>cand₁</td>
</tr>
<tr>
<td>b.</td>
<td>cand₂</td>
</tr>
</tbody>
</table>
Both candidates fare equally well (or badly) on both constraints. No conclusions about the relative ordering between them can be drawn, and the resolution of the conflict has to be adjudicated by other constraints.

Another frequently encountered situation which can serve no basis for establishing constraint rankings is where the set of marks incurred by one candidate is a proper subset of the marks incurred by some other candidate. An example is shown in (10) below.

(10)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>cand₁</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>cand₂</td>
<td>*</td>
</tr>
</tbody>
</table>

In (10), *cand₁* has proper subset of *cand₂*’s marks. Thus *cand₁* cannot fail to win on any ranking of *A* and *B*. In this situation, we say that *cand₁* harmonically bounds *cand₂*.

### 2.2.2 Factorial typology

The goal of generative grammar is to enumerate the set of possible grammars. Although some rankings are assumed to be universally given (see Prince and Smolensky 1993: 127–172), the rankings between the vast majority of constraints is free to vary on a language particular basis and has to be learnt.

In OT, the set of possible languages is thus generated by factorially permuting all the constraints in \( \text{CON} \) (modulo the universal rankings) to output a typology of possible systems. In a system of 3 constraints \{A, B, C\}, the number of possible total orderings on the set is \( 3! = 6 \). (11) enumerates the set of possible permutations (total orderings) on this set. ‘\( \gg \)’ designates the relation ‘dominates’.

(11) **Total orderings on ‘constraint set’ \{A,B,C\}**

- A\( \gg \)B\( \gg \)C
- A\( \gg \)C\( \gg \)B
- B\( \gg \)A\( \gg \)C
- B\( \gg \)C\( \gg \)A
- C\( \gg \)A\( \gg \)B
- C\( \gg \)B\( \gg \)A

A central part of Optimality-theoretic is testing the typologies generated by factorially permuting the constraints against the range of attested grammar types. Nevertheless, this kind of testing has its limitations. This is most obvious when we consider that OT predicts a number of unviable grammar types. For example, a grammar in which all markedness constraints dominated all faithfulness constraints would be communicatively unviable (although something precisely like this has been argued to represent the initial state of the language faculty in the pre-linguistic child—Tesar and Smolensky 1993, 1996 and Hayes 1999a). Conversely, the domination of all markedness by all faithfulness would
yield a language which exploited every human phonological capacity: it would make full use of the attested range of prosodies and segmental contrasts. Even if such a system didn’t place excessive burdens on memory, it would use far more contrasts than is needed in order to communicate efficiently. The set of languages which are in principle capable of taking up residence in a human brain is thus vastly larger than the set of languages we would expect to find attested on grounds of communicative efficiency and learnability.

Given Universality, the OT linguist is under special pressure to ensure that the constraints he proposes are genuinely principled and not merely ad hoc solutions to language particular problems. This is also crucial, given the nature of OT compared with parametric theories. A binary parametric system with $n$ parameters yields $2^n$ possible grammars, while an OT system consisting of $n$ constraints yields $n!$ totally ranked hierarchies. At $n = 30$, the number of possible parameter-based grammars is a mere $1,073,741,824$. The number of OT grammars, by contrast, is an astronomical $29,472,540,000,000,000,000,000,000,000$. Without a truly principled means of reigning in the number of constraints, the typological enterprise risks being rendered vacuous. We will return to this question in §2.3.

2.2.3 Parallelism

Most OT work has adopted as its null hypothesis that optimization proceeds in parallel, without serial ordering of rules or strata. Candidates are generated and evaluated simultaneously against the entire constraint ranking. Derivations in OT are thus one-step mappings from an input to a set of fully-formed output representations. All-out parallelism is adopted here.

Parallelism is not a necessary property of OT. Several researchers have proposed that strata of constraint hierarchies may be ordered in a way which recapitulates the rule-ordering of rule-based serialism. See McCarthy and Prince (1993a), Black (1993), Kenstowicz (1995), and Kiparsky (1997, 1998) for examples of proposals of this type.

2.3 Universality and the functional basis of constraints

What exactly makes for a principled constraint? This question is in OT an especially pressing one in view of the vast space of possible grammars generated when OT’s combinatorial possibilities are utilized to the full. The standard answer has been that a constraint is principled if it is supported by a significant body of typological evidence, and is proven to work in many different languages.

Hayes (1996) argues for a different emphasis in evaluating the legitimacy of constraints. A principled constraint, argues Hayes, is one which can be justified on functional grounds. (See also Steriade 1997; Boersma 1998; Kirchner 1998; Haspelmath 1999 for recent discussion, as well as Deacon 1997 for a similar non-OT view.) Hayes proposes, contrary to prevailing assumptions (Prince and Smolensky 1993; Tesar and Smolensky 1993, 1996), that constraints are not innately given. Innatism of course significantly decreases the learning burden, since the learner only has to acquire the ranking of the constraints. However, the burden placed on evolution by innatism is unacceptably great. Hayes makes
2.3. UNIVERSALITY

the point like this. Many of the constraints attested in natural languages ban rare segment types, in which case there would have been no selective pressure to evolve innate representation of them. More generally, constraints which drive common processes don’t have to be innate if they can be reliably learned from the environment. It makes poor evolutionary sense to develop rich innate specifications when the environment is such that the necessary connections can be acquired quickly and at little cost (Hinton and Nowlan 1987). Hayes takes issue with the innatism of the standard view and adopts what I will call the *enactionist* view of constraint acquisition. The essence of Hayes’ proposal is this: constraints are crystallized by the learner during the process of language acquisition by phonetic experience. This mechanism responsible is a *selectionist* one he terms *inductive grounding*.

Inductive grounding functions similarly to the OT grammar, in the sense it involves the generation of pools of candidate constraints out of the basic resources of the representational theory, and the selection of the optimal (phonetically most efficient) constraints. The constraints that the learner arrives at through this process reflect a trade-off between the conflicting demands of phonetic accuracy and formal simplicity. Hayes also argues that the data on language acquisition by the child supports the view that children do create constraints by inductive grounding.

From the phonetic experience acquired from operating his own articulatory and perceptual apparatus, the learner constructs *phonetic maps*. The phonetic map may be envisioned as a landscape which describes the phonetic difficulty associated with articulation and perception at certain points in the landscape. The degree to which phonological constraints can faithfully capture this landscape is offset by the need for constraints to be formally simple. Real life phonological constraints reflect a trade-off between these two conflicting demands. *Grounded constraints* are those which optimize these conflicting demands best. They are selected from the pool of candidates on the basis of how they fare in terms of phonetic efficiency. Constraints are effective to the extent they ban hard configurations and allow easy configurations.

It is clear that children’s phonologies do a number of things there may be no evidence for in the stimulus. They don’t imitate adult forms so much as follow their own phonetic bent. Hayes, citing a study by Smith (1973) of his daughter Amahl, describes her rendering of input stops. Voicing distinctions in stops were neutralized to voiceless unaspirated lenis word-initially ([ʰebu] ‘table’), voiceless finally ([aːt] ‘hard’), and voiced medially ([w@ːgin] ‘working’). The processes reflect the demands of constraints universal to all human beings.

Let us sum up this discussion at this point and pose the question which the Optimality-theoretic perspective so loudly begs. Constraints require the truth of certain phonological generalizations. But are these phonological generalizations abstracted directly from the data with which the learner is confronted itself, or are *all* constraints (phonological generalizations) induced by phonetic experience? Put differently, are there any linguistic generalizations which are encoded into grammatical mechanisms other than the language particular ranking of the universal constraints? Hayes’ position with respect to this question

---

5The term *enactionism* in the sense I intend here, see Foley (1997). The term refers to any theory which holds that the acquisition of knowledge is mediated by physical interaction with the external world.

6The term is a misnomer, since what Hayes proposes is a *selectionist* algorithm for learning constraints and not, strictly, an inductive one.
is clear: he at least allows for the possibility that the learner may posit ungrounded constraints in order to be able to capture generalizations about their language which lack a functional basis. This point is considerably developed in Hayes (1999b), where Hayes proposes that certain facts of the phonology of the Australian language Yidp must be handled in terms of such ungrounded constraints (see also Bolognesi 1998 and Kilre 1998 for a similar point of view).

For Hayes, acquisition is a mixture of experience-based and data-based learning. This is quite a major concession to declarative ways of thinking, however: if language learners have in principle construct constraints with brute-force descriptive power, it begs the question why shouldn’t they capture all of their phonological knowledge in these terms? Why not go in for a grammar which is all-out declarative? In contrast to grounded constraints, ungrounded constraints would be learned on the basis of gaps in the input data itself. Such constraints lack the character of OT constraints: since they are abstracted directly from the data, they are language-specific and inviolable. To incorporate them into the OT grammar would essentially subvert the typological aims of the enterprise, do violence to OT’s universalism, thereby considerably reducing OT’s appeal. In short, why aren’t all constraints simply declarative ones, as opposed to violable ones?

Perhaps we can rather strengthen Hayes position and push everything through relying exclusively on the kind of learning based on phonetic experience alone. A stronger version of the enactionist hypothesis would rule out the possibility of ungrounded constraints altogether, only admitting the set of inductively grounded functional constraints to the grammar. This is a radical stance, since it entails that the learner construe all phonological generalizations and alternations as functionally motivated, including those which either lack an overt functional motivation or which are patently dysfunctional. All dysfunctional grammatical patterns must emerge on reranking the grounded constraints. I will call this position the strong enactionist hypothesis.

(12) **Strong Enactionist Hypothesis**

The grammar admits only grounded constraints.

The claim is admittedly extreme. It nevertheless has to be made explicitly, since it is effectively implied by OT’s insistence on the universality of the constraint set. If OT’s typological claims are not to be rendered vacuous, then we must not admit language-specific constraints to the grammar, i.e. any constraint which serves merely to capture some pattern in the stimulus which is not motivated by functional concerns. Hayes himself has explicitly proposed the existence of ‘Anticorrespondence’ constraints in Yidp (Hayes 1999b). His proposal entails that grammars simply have to have recourse to some brute-force descriptive power in order to capture dysfunctional patterns. In §3.4.6 we will examine the relevant Yidp data and find that resort to brute-force descriptive constraints is not warranted in this case.

Countless researchers have emphasized that not all phonological structure is phonetically sensible. Such quirks arise as a result of a conspiracy of historical circumstances, such as borrowing, analogy, or rule telescoping. Although the diachronic chain of events which lead to some synchronic state of affairs may resolve itself into a sequence of entirely natural stages, the result may often be unnatural. What is central for us is how such accidents of history are to be handled in the grammar.
But does the existence of such dysfunctionality mitigate the postulation of functionally arbitrary constraints, which permit the triggering of phonological alternations by morphosyntactic features, and even arbitrary morphological features. If constraints must be functionally grounded, it makes no sense to allow morphological information such as this to serve as raw material for phonological constraints. For us, the most important candidates for ungrounded constraints are those which posit specific morphological or syntactic elements as triggers for phonological alternations (e.g. the conditioning of lenition in the environment [+feminine singular] in Welsh — Pyatt 1997). (12) precludes the learning of constraints from distributional regularities in the stimulus. (12) also precludes any mixing of levels. Thus, the triggering of a phonological alternation by a morphosyntactic feature cannot be made the subject of a grounded constraint. This follows from the autonomy of phonology and syntax: since by hypothesis there is no interaction between the two modules, there can be no functional basis for the alternation.

Strong enactmentism also entails therefore that learners must posit a covert functional phonological basis for any phonological alternation with only an overt morphological or syntactic trigger. I will term this recoverability through function. Put differently, abstract phonological content must be recoverable to some degree from the alternations it induces, not merely the alternations (if any) in which it participates. For example, the observation that word-initial consonants are limited in some particular context, or set of contexts, may be taken as reliable evidence that the trigger of the operation is preceded by some (abstract) phonological element, one of whose typical effects is lenition of a following consonant. There may be several candidates for the identity of the abstract trigger, but the least marked choice is probably a vowel. In the absence of an overt vowel in the relevant positions on the surface, the learner is by hypothesis able to 'reconstruct' this abstract 'ghost' vowel as part of the speaker's 'intentions'. Let us dub this the Ghost Hypothesis.

\[(13) \text{Ghost Hypothesis}\]

All 'morphological conditioning' is reducible to the triggering and blocking effects of covert ('ghost') phonological elements.

Morphologization is really part of a larger pattern of absolute neutralization of some phonological contrast. In Welsh and Saami, this neutralization is to zero.

Abstract analyses have been variously faulted on grounds of psychological implausibility, excessive power, and that they dress up diachronic facts as syn-
chronic truth. Lass (1984: 211f.) is a good overview of the criticisms. I will briefly review his most cogent points here, and attempt to counter them.

Abstract analyses are frequently justified on the grounds that they allow the linguist to capture the relevant generalizations in terms of exclusively natural rules. Thus, if phonetic motivation is highly valued, and arbitrariness in rule formulation is not, then this counts in favour of the abstract analysis. Lass argues that this concern with naturalness is misplaced on the grounds that a speaker’s knowledge of the phonology of his language is, as it were, pre-phonetic: the claim that coarticulatory and perceptual effects can have any relevance in a purely mental sphere is not coherent, according to Lass. Yet, if Hayes is correct in assuming that constraints, the very stuff of which the grammar is constituted, are crystallized in the process of actualizing one’s phonetic capabilities, then knowledge of phonology entails a functional knowledge which is at a fundamental level ‘embodied’ and ceases to be exclusively mental. Naturalness is no longer misplaced, and Lass’ objection loses much of its force.

Lass marshals a further objection to abstractness on the grounds of the supposed psychological implausibility of abstract analyses. Ultimately, any analysis worth its salt has to model the knowledge actually internalized by the speaker as grammatical competence. When practising linguists bring considerations of elegance and restrictiveness to the task of analyzing dysfunctional patterns, phonological solutions are preferred to those involving some measure of morphological arbitrariness. This is a purely methodological criterion, though, argues Lass, and it is unconscionable to impute it to the language learner as a psychological one, because it involves making “extraordinary claims about the representational powers of naïve speakers” (Lass 1984: 213). If knowledge of the phonology also involves the functional knowledge of how segments behave in context, the extraordinaryness of this claim pales. For example, the recoverability-by-function hypothesis suggests that a speaker of Welsh can reconstruct a vowel in the relevant position to trigger lenition by virtue of a functional knowledge, acquired through phonetic experience, of how vowels behave in context.

A third objection is that abstract analyses involve projecting the history of the language onto the competence of the speaker. Again, this objection can be countered. Abstract synchronic analyses and diachrony may coincide in many cases, but this is not a ground for suspicion. Acquiring a morphologically conditioned alternation as opaquely phonological does not require that the learner be able to divine the history of the language: abstract phonological entities are recovered through their function and behaviour in context.

Lass’ final objection, and in his opinion the most damning, is that the abstract phonological solution involves the merely diacritic use of phonological features. Lass contends that, in abstract phonological analyses, “phonological specifications [...] are being used as mere notational variants of purely morphological specifications” (Lass 1984: 214). Again, the objection seems to spring from Lass’ structuralist bias, that the terms of phonological analysis do not refer in any way to physical realities. Distinctive features merely serve to pick out sets of segments, but are in themselves untainted by phonetic content. Thus, the sense in which abstract phonological specifications are notational variants of morphological ones is that they serve to pick out the same sets.

The view I am developing here thus contrasts with the more typical mentalist view, that phonological content must be present in the stimulus to be recoverable, at least as an alternation. Phonological opacity, therefore, is not
2.4. RICHNESS OF THE BASE

generally considered problematic for the mentalist perspective, since whatever
is opacified is recoverable from some alternation. In morphological conditioning,
on the other hand, this is not the case. There are other reasons for rejecting
the mentalist view. The strong enactionist hypothesis is a powerful constraining
force on grammars and gives much tighter typologies. Morphological condition-
ing is not qualitatively different from common-or-garden phonological opacity.
The resolution of the issue as to which grammars are used by speakers turns on
whether learners actually are able to posit abstract underlying representations
on the basis of observed phonological alternations. This ascribes impressive ca-
pabilities to the learner to recover phonological information which is absent in
the stimulus. Yet, grammatically, what we appear to have is a cline of opacity,
ranging from full transparency (the relevant element participating in some
phonological generalization is exceptionlessly present on the surface), through
opacity (the relevant element participating in some phonological generalization
is sometimes absent on the surface), to morphologization (the relevant element
participating in some phonological generalization is always absent on the sur-
face). If the strong enactionist hypothesis is correct, then morphologization is
not different in kind to opacity, and we are at liberty to opt for what is formally
the most restrictive approach.

2.4 Richness of the Base

Despite its fundamental importance, Richness of the Base is surely the most
misunderstood of all OT’s tenets. This is significantly due to the scarcity of
accessible material on the subject. As a consequence, many of OT’s practitioners
still seem to work from within a rule-based serialist mind-set.

Adopting an idealization which might be dubbed richness of the base\(^9\), we
have abstracted away from considerations of gaps or systematic omissions
in the lexicon, and assumed, for the purposes of deducing the possible
outputs of the grammar, that all inputs are possible.

(Prince and Smolensky 1993: 191)

Paul Smolensky sums up the relation between Richness of the Base and
OT’s output-orientation in the following way:

The source of all systematic cross-linguistic variation is constraint reranking.
In particular, the set of inputs to the grammars of all languages is the
same. The grammatical inventories of a language are the outputs which
emerge from the grammar when it is fed the universal set of all possible
inputs.

(Smolensky 1996b: 3)

Richness of the Base is a logico-methodological assumption about the set
of inputs to the OT grammar. It specifies that there are no language-specific
restrictions on the input beyond whatever the representational theory admits
as a valid expression. Some of these restrictions are obvious. For instance,
valid phonological expressions have to build on phonological primitives such
as features, moras, syllables, stems, affixes, edges, and so on, and not things
like functional heads, kinship terms, or neutrinos. These primitives have to

\(^9\)The base refers to the input set itself, richness to the all inclusive nature of its members.
be combined in a licit fashion. Thus, while structures in which feet dominate syllables are valid inputs, structures in which syllables dominate feet are not.\textsuperscript{10}

Beyond these broad limits, anything goes, and the grammar has to be able to cope with anything the input might throw at it. In this sense, \textit{Richness of the Base} has a real-world analogue. Real world grammars \textit{do} in practice take novel inputs and successfully manage to generate some kind of output on the basis of them. Children in particular are constantly confronted with novel inputs, which their grammar has to deal with. Children’s productions typically diverge markedly from the inputs they aspire to render, reflecting the effect of the child’s grammar (Gnanadesikan 1995; Smolensky 1996a,b,c). In adult language as well, novel inputs may come in the form of loanwords, which are fed through the grammar and thus become nativized.

One of the rationales for adopting the \textit{Richness of the Base} tenet is economy. In OT, the burden of explanation is on the output-oriented constraints and their interaction. All language variation is reducible to constraint interaction in the \textit{Eval} component of the grammar.\textsuperscript{11} In order that this claim be made as strongly as possible, we have to assume that there are no constraints on either the generator (\textit{Gen}) or the set of inputs. \textit{Richness of the Base} is thus partly an attempt to stave off the threat of duplication: constraints on inputs would risk duplicating information which is independently expressed in the ranked hierarchy of output-oriented constraints.

2.4.1 \textbf{Allophony, Richness of the Base, and Robust Grammars}

Let’s now examine one concrete application of the \textit{Richness of the Base}. Since the input set to the grammar is universal, and the set of language-specific outputs is a proper subset of the universal input set, it follows that input contrasts are massively neutralized in the output. One good example of this is the allophonic process, where two or more allophones participate in a neutralization which is essentially bidirectional. Allophones are classically in complementary distribution with each other. The impossibility of lexical contrast in OT is derived by ranking the relevant faithfulness constraint(s) low. For example, in North Saami, the labiodental fricative [v] and the bilabial approximant [B] are in complementary distribution with each other. The [v] allophone is restricted to occurring in onset position (either as a singleton, or as ‘part’ of a geminate). Examples illustrating the distributional difference are shown in (14).

\begin{verbatim}
(14) a. [v]uulos down
    tu[vv]aa dove
    a[vv]e quite
    saa[vv]e large tub

b. o[B]tteh previous
    kaa[B]t[tu]j it was found
\end{verbatim}

\textsuperscript{10}This is standardly understood as a \textit{restriction} on \textit{Gen}.

\textsuperscript{11}The other side of this coin is that there are \textit{no} universal properties which are the result of constraint interaction in \textit{Eval}.
In structuralist terms, both [v] and [β] are contextually determined allophones of the phoneme /β/ (or /v/). In the OT grammar, the distribution of each allophone is governed by constraints on structural harmony, the ‘basic’ allophone occurring in the more general context. It isn’t entirely clear which of the two contexts is functioning as the more specific in the example at hand, but let’s assume that the ‘basic’ allophone is [β]. (The notion of ‘basic’ allophone is exclusively a matter of whether the allophone is contextually-determined or not.) That is, we can write a rule which maps /β/ → [v]/σ, with the [β] as the ‘elsewhere’ realization. In OT terms, the unfaithful mapping of /β/ to [v] in onset position is driven by a positional markedness constraint *[σβ], which bans [β] in syllable-initial position. Suppose, for concreteness, that the unfaithful mapping in question is a violation of Ident[son]. Because of Richness of the Base, there can be no sense in which either allophone is privileged over the other as an input. Both have to be considered. The inputs to the allophonic alternation are non-unique, and the grammar has to ensure that the potential underlying contrast is neutralized irrespective of which input is chosen. Taking /buvulos/ as input, the onset is unfaithfully mapped to [v] due to the exigencies of highly-ranked *[σβ].

(15)

<table>
<thead>
<tr>
<th>[βuvulos]</th>
<th>*[σβ]</th>
<th>Ident[son]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. vuvulos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. βuvulos</td>
<td>![ ]</td>
<td></td>
</tr>
</tbody>
</table>

Setting /vuvulos/ as the input, as in (16) below, simply means that the optimal candidate fails to incur a mark on Ident[son]. Instead, the mark is incurred by suboptimal (b), which is eliminated by highly-ranked *[σβ] anyway.

(16)

<table>
<thead>
<tr>
<th>/vuvulos/</th>
<th>*[σβ]</th>
<th>Ident[son]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. vuvulos</td>
<td>![ ]</td>
<td></td>
</tr>
<tr>
<td>b. βuvulos</td>
<td>![ ]</td>
<td></td>
</tr>
</tbody>
</table>

Regardless of the choice of input (/v/ or /β/), then, the victory of the desired winner is ensured. However, now we have to ensure that the same result follows in what we have designated as the general context (positions other than onset). By Richness of the Base, the grammar has to be able to return the correct output form on both possible inputs. In this case, there must be a high-ranking constraint which ensures an underlying /v/ is mapped to [β] in the elsewhere case. This can be modelled by allowing the general markedness constraint *v outrank Ident[son]. The constraint *v must itself be outranked by the more specific markedness constraint, *[σβ], otherwise /v/ and /β/ would map to surface [β] in all contexts.

(17)

<table>
<thead>
<tr>
<th>/otette/</th>
<th>*[σβ]</th>
<th>*v</th>
<th>Ident[son]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. otette</td>
<td>![ ]</td>
<td>![ ]</td>
<td></td>
</tr>
<tr>
<td>b. e otette</td>
<td>![ ]</td>
<td>![ ]</td>
<td></td>
</tr>
</tbody>
</table>
Again, the optimization of the correct winner is assured, irrespective of our assumptions about the input. I will designate grammars with this property as \textit{robust}. The opposite of \textit{robust} is \textit{fragile}. Fragile grammars are essentially incomplete because they either omit certain crucial constraints or tacitly assume the activity of constraints on the input. We shall return to language-specific restrictions on the input at several points throughout the dissertation, in particular sections §3.4 and §11.6.

### 2.5 Correspondence Theory and Faithfulness

In earlier versions of OT (sometimes called the Parse/Fill theory), a principle of \textit{Containment} was assumed. Containment specified that everything in the input was also present in the output — nothing was literally deleted. \textsc{Gen} was thus restricted to feature-filling and structure-adding operations. Developing an idea originally proposed by Itô (1986, 1989), deletion and feature change was implemented formally as \textit{underparsing}, i.e. the 'deleted' segment or feature simply failed to get incorporated into higher-level prosodic structure. In the phonetic component, such prosodically unlicensed material was simply left uninterpreted.

Correspondence Theory, as developed by McCarthy and Prince (1995), McCarthy (1995), and Urbanczyk (1996) marks a significant break with this way of thinking. Originally developed as a theory of how reduplicates acquire phonological content, Correspondence developed into a general theory of the relations between phonological strings.

The formal definition of Correspondence is given in (19).

\begin{equation}
\text{Correspondence (McCarthy and Prince 1995: 262)}
\end{equation}

Given two strings \( S_1 \) and \( S_2 \), Correspondence is a function \( \mathcal{R} \) from the elements of \( S_1 \) to those of \( S_2 \). Elements \( \alpha \in S_1 \) and \( \beta \in S_2 \) are referred to as correspondents of one another when \( \alpha \mathcal{R} \beta \).

Correspondents are indicated using subscript indices. For example, in the input-output mapping \(/\text{wi}z\text{a}z\text{a}f\text{a} + s_{5}/ \rightarrow /\text{wi}z\text{a}z\text{a}v\text{a}v\text{a}z\), the correspondence between input /\text{f}/ and output \( v \) is indicated by the identity of the subscript index.

An inventory of the most important kinds of correspondence-based faithfulness constraints is found in McCarthy and Prince (1995: 370ff.), which is given in full in (20) to (28) below. All the constraints refer to pairs of representations \( (S_1, S_2) \). The strings which stand in correspondence may vary. In addition to Input–Output correspondence, correspondence may hold between a Reduplicant and its Base (McCarthy and Prince 1993a, 1995) and a Base and a Derived form (Beau 1998). Each of these mappings is regulated by distinct, but formally identical, families of faithfulness constraints.
2.5. CORRESPONDENCE THEORY

(20) **Max** — “No deletion”\(^{12}\)
Every element of \(S_1\) has a correspondent in \(S_2\).
Penalizes unfaithful mappings of the form: \(/\alpha/ \to \emptyset\).

(21) **Dep** — “No epenthesis”\(^{13}\)
Every element of \(S_2\) has a correspondent in \(S_1\).
Penalizes unfaithful mappings of the form: \(/\emptyset/ \to \alpha\).

(22) **Ident**[\(F\)]
Corresponding segments must have identical values for the feature \(F\).
Penalizes unfaithful mappings of the form: \([\alpha F] \to \neg[\neg \alpha F]\).\(^{14}\)

(23) **I-Contiguity**
The portion of \(S_1\) standing in correspondence forms a contiguous substring.
Penalizes unfaithful mappings of the form: \(/\alpha\beta/ \to \alpha\gamma\beta\).

(24) **O-Contiguity**
The portion of \(S_2\) standing in correspondence forms a contiguous substring.
Penalizes unfaithful mappings of the form: \(/\alpha\beta\gamma/ \to \alpha\gamma\).

(25) **\{Right, Left\}-Anchor(\(S_1, S_2\)**
Any element at the designated periphery of \(S_1\) has a correspondent at the designated periphery of \(S_2\).
Let \(\text{Edge}(X, \{L, R\}) = \text{the element standing at the Edge = L, R of X.}\)
\textbf{Right-Anchor.} If \(x = \text{Edge}(S_1, R)\) and \(y = \text{Edge}(S_2, R)\) then \(x \Re y\).
\textbf{Left-Anchor.} If \(x = \text{Edge}(S_1, L)\) and \(y = \text{Edge}(S_2, L)\) then \(x \Re y\).

(26) **Linearity** — “No metathesis”
\(S_1\) must be consistent with the precedence structure of \(S_2\) and vice versa.
Let \(x, y \in S_1\) and \(x', y' \in S_2\). If \(x \Re x'\) and \(y \Re y'\), then \(x < y\) iff \(x' < y'\).
Penalizes unfaithful mappings of the form: \(/\alpha\beta/ \to \beta\alpha\).

(27) **Uniformity** — “No coalescence”
No element of \(S_2\) has multiple correspondents in \(S_1\).
For \(x, y \in S_1\) and \(z \in S_2\), if \(x \Re z\) and \(y \Re z\), then \(x = y\).
Penalizes unfaithful mappings of the form: \(/\alpha_1\beta_2/ \to \gamma_{1,2}\).

(28) **Integrity** — “No breaking”
No element of \(S_1\) has multiple correspondents in \(S_2\).

\(^{12}\)Etymologically, Max is a truncated form of maximality. The term derives its use from the constraint requiring the copying of base material into the reduplicant to be maximal. Other more transparent terms have been proposed for this constraint, but Max has gained the widest currency.

\(^{13}\)The etymology is base-dependency, from McCarthy and Prince (1994). The constraint required the material of the reduplicant to be ‘dependent’ on that of the root, i.e. not add anything that wasn’t strictly base material.

\(^{14}\)Note that deletion of a segment does not entail violation of Ident[F]. Since a deleted segment has no output correspondent, Ident[F] is vacuously satisfied.
For \( x \in S_1 \) and \( w, z \in S_2 \), if \( x \not\Rightarrow w \) and \( x \not\Rightarrow z \), then \( w = z \).

Penalizes unfaithful mappings of the form: \( /\alpha_1/ \rightarrow /\beta_1\gamma_1/ \).

Bye (1996), Benua (1998), and Beckman (1998) extend the use of Correspondence to regulate autosegmental relations between tiers of the representation. Autosegmental constraints include the following.

(29) **Totality**
Every element of \( T_1 \) has an association to an element in \( T_2 \).

(30) **Unique** (Benua 1998; Beckman 1998)
No element of \( T_2 \) has multiple correspondents in \( T_1 \).
For \( x, y \in T_1 \) and \( z \in T_2 \), if \( x \not\Rightarrow z \) and \( y \not\Rightarrow z \), then \( x = y \).

### 2.5.1 The Emergence of the Unmarked

The resolution of Faithfulness constraints into formally identical constraints holding of different pairs of related forms makes a crucial prediction. The Emergence of the Unmarked (TETU) has played a central role in the development of OT, and a wide variety of grammatical patterns bear its imprint. Below we will examine its role in describing the distribution of prosodically-driven reduction and neutralization, and how it clarifies our understanding of lexical exceptionality. First, though, let us take a look at its role in reduplication, for which it was originally developed. An instructive example is the emergence of unmarked syllable structure in the Sanskrit perfect reduplicant. In Sanskrit, the perfect is formed by copying root material onto a CV prefix template. Examples are from Steriade (1988), cited in Kenstowicz (1994: 653). Reduplicates are shown underlined.

(31) **Perfective reduplication in Sanskrit**

<table>
<thead>
<tr>
<th>Base</th>
<th>Reduplicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>pat</td>
<td>pa-pat-a</td>
</tr>
<tr>
<td>prath</td>
<td>pr-prath-a</td>
</tr>
<tr>
<td>kṣād</td>
<td>k-kṣād-a</td>
</tr>
<tr>
<td>nna:</td>
<td>mn-mnā-u</td>
</tr>
</tbody>
</table>

As the examples show, the copy is incomplete in the case of roots with initial complex onsets. What emerges in the reduplicant, though, is an unmarked syllable structure type: a complex onset in the base doesn’t emerge as such in the reduplicant. Phenomena such as these are captured by positing separate Input-Output (IO) and Base-Reduplicant (BR) correspondence constraints, and letting them interact with the markedness constraints separately. In the case at hand, the relevant markedness constraints are \(*\text{ComplexOnset} \ (\ ^*\text{COns}) \) and \( \text{NoCoda} \).

(32) \(^*\text{COns}\)
Complex onsets are disallowed.

(33) \(\text{NoCoda}\)
Codos are disallowed.
2.5. Correspondence Theory

Now, complex onsets and codas are permitted in the base, as evidenced by forms such as \( k\ddot{a}d \). This shows that the constraint enforcing the faithful parsing of input consonants in the base, \( \text{IO-MAX-C} \), must dominate \(*\text{CxONS}\) and \( \text{NoCoda}\). However, complex onsets and codas fail to surface in the reduplicant. Reduplicant identity is thus mediated by a separate faithfulness constraint, \( \text{BR-MAX-C}\). The emergence of the unmarked in the reduplicant shows that \(*\text{CxONS}\) and \( \text{NoCoda}\) must dominate \( \text{BR-MAX-C} \), as in (34). The reduplicant is shown underlined.

\[
| /\text{RED}+k\ddot{a}d/ | \text{IO-MAX-C} | \text{xONS} | \text{Coda} | \text{BR-MAX-C} |
|----------------|----------|--------|--------|
| a. \*
\kappa-a-k\ddot{a}d-a
| * | * | |
| b. \k\ddot{a}d-k\ddot{a}d-a
| * | * | *
| c. \k\ddot{a}d-k\ddot{a}d-a
| * | | * *
| d. k\ddot{a}d-k\ddot{a}d-a
| * | * | *
| e. \k\ddot{a}d-k\ddot{a}d-a
| * | * | *
| f. \k\ddot{a}d-\ddot{a}d-a
| * | * | *

The general schema for the emergence of the unmarked is \( \text{IO-Faith} \gg \text{Markedness} \gg \text{BR-Faith} \).

2.5.2 Positional Faithfulness

It has long been recognized that the environments for reduction and neutralization are positionally circumscribed. Neutralizing processes are usually found in positions with small perceptual load, for example unstressed or unfooted syllables, non-initial syllables, and so on. Conversely, prominent positions, such as stressed syllables or word-initial syllables are primary sites for the preservation of contrast and enhancement of perceptual cues.\(^{15}\) See de Lucy (1998a) for an account of some of these enhancement (positional markedness) constraints. The theory of positional faithfulness, or prosodic faithfulness (Alderete 1995; Casali 1996; Beckman 1997, 1998) capitalizes on the emergence of the unmarked by introducing faithfulness constraints whose jurisdiction is restricted to specific prosodic positions. These include stressed and foot-internal position (Alderete 1995), root-initial (Beckman 1997, 1998), morpheme-initial (Casali 1996). Reduction and neutralization are driven by context-free markedness constraints. The possibilities afforded by reranking the constraints are as in (35). Here, \( \text{PF} = \text{Positional Faithfulness Constraint}; \, \text{F} = \text{Faithfulness}; \, \text{M} = \text{Markedness}. \) (35b) represents the emergence of the unmarked.

\(^{15}\)Preservation of contrast and enhancement sometimes conflict. An example is the word-initial neutralization of the voicing contrast to \([-\text{voiced}]\) in obstruent stops in Bavarian German. The neutralization does, however, enhance the perceptual prominence of the stop burst, presumably facilitating more reliable identification of place features.
CHAPTER 2. OPTIMALITY THEORY

(35) Ranking possibilities with positional faithfulness

a. \( \text{PF} \gg \text{F} \gg \text{M} \)
   No neutralization anywhere.

b. \( \text{PF} \gg \text{M} \gg \text{F} \)
   Faithful preservation in prominent position;
   neutralization in non-prominent positions.

c. \( \text{M} \gg \text{PF} \gg \text{F} \)
   Across-the-board neutralization.

2.5.3 Lexical exceptionality

Pater (1995) lays out a theory for dealing with lexical exceptionality in OT
which capitalizes on the emergence of the unmarked.\(^{16}\) He observes that certain
English words arbitrarily fail to conform to the canonical stress pattern. Thus,
the canonical pattern, represented by \textit{information} (< \textit{infórm}) is repressed in an
otherwise phonologically equivalent word such as \textit{condénsation} (< \textit{condénse}),
which preserves the primary stress of the base as a secondary stress in the
derived form. The same primary stress in \textit{infórm} fails to surface in the corre-
spending derived form. Pater hypothesizes that certain sets of words are subject
to lexically-specific faithfulness constraints whose jurisdiction is restricted to
that particular class of lexical items. Ranking the lexically-specific faithfulness
constraint above the relevant markedness constraints generates non-canonical
phonological patterns. For example, it is a fact about the canonical phonology
of English, that words in English do not contrastively nasalize vowels. This is
captured in terms of the ranking \( *\text{V} \text{nás} \gg \text{IDENT}[\text{nás}] \).

(36) \( *\text{V} \text{nás} \)
   Nasal vowels are disallowed.

However, the set of French loanwords is exempt from this restriction (at least
in higher registers). Thus [f\(\text{Éö}b\)l] is a grammatical rendering of the French
expression \textit{enfant terrible} in RP.\(^{17}\) We may diacritically mark the set of French
loanwords as subject to a lexically-specific faithfulness constraint \( \gamma\text{-IDENT}[\text{nás}] \).
(37) and (38) illustrate the differential treatment of canonical English words
and French loanwords with respect to the preservation of nasalization.

(37) Contrastive vowel nasalization eliminated in canonical phonology of RP

<table>
<thead>
<tr>
<th></th>
<th>/b(\text{ê})/</th>
<th>( \gamma\text{-IDENT}[\text{nás}] )</th>
<th>( *\text{V} \text{nás} )</th>
<th>IDENT[\text{nás}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>( \text{ê} )</td>
<td>b(\text{ê})</td>
<td>( *\text{V} \text{nás} )</td>
<td>( \text{IDENT[\text{nás}]} )</td>
</tr>
<tr>
<td>b.</td>
<td>b(\text{ê})</td>
<td></td>
<td>( *! )</td>
<td>( \text{IDENT[\text{nás}]} )</td>
</tr>
</tbody>
</table>

(38) French loanwords preserve contrastive vowel nasalization in RP

\(^{16}\)Fukazawa (1997) and Fukazawa et al. (1998) have independently made an identical
proposal.

\(^{17}\)Loanwords are rarely borrowed into the target language completely unmodified. While
the vowel nasalization and stress of the French inputs escape modification, other features of
the phonology of the source language are not preserved. Given the French rendering [\(\text{ê} \text{tré} \text{bl}]\), it is apparent that the vowel qualities and rhotic unroundness do not survive the ravages
of the English grammar. Note too that English epentheses to eliminate the reverse sonority
cluster.
The canonical pattern is thus an instantiation of the emergence of the un-marked.

Loanwords exemplify the situation where a particular set of items may be exempt from processes which operate canonically in the language. The converse of this is the minor rule (Kennstowicz and Kisseberth 1977), where some set of lexical items is diacritically marked as being subject to a rule which does not operate for the majority of lexical items. For example, in Cornish (Brown 1993: 14), a few words with initial /d/ are subject to nasalization following the definite article an 'the'. The vast majority of d-initial words do not undergo the rule. Thus, do:n-an do:n 'the son-in-law' instantiates the canonical pattern, while do:r-an no:r 'the world' exemplifies the minor application of nasalization. For the majority of lexical items, Ident[nas] dominates Nasalization, but items such as do:r may be understood as being subject to a lexically-specific version of the markedness constraint, δ-Nasalization, which dominates Ident[nas].

### 2.6 Phonological Opacity

In its classical form, the OT architecture admits of only two levels of representation, an input and an output. This minimal architecture has long been a problem given the ultimate necessity of having to incorporate some kind of explanation for phonological opacity within OT. According to Kiparsky’s classic definition (Kiparsky 1973: 79),

(39) A phonological rule \(\mathcal{P}\) of the form \(A \rightarrow B/C\_D\) is **opaque** if there are surface structures with any of the following characteristics:

a. instances of \(A\) in the environment \(C\_D\)

b. instances of \(B\) derived by \(\mathcal{P}\) that occur in environments other than \(C\_D\).

Given rule \(\mathcal{P}\), instances of \(A\) in the environment \(C\_D\) count as underapplications of \(\mathcal{P}\): although the structural description for \(\mathcal{P}\) is satisfied, the rule fails to apply. Instances of \(B\) derived by \(\mathcal{P}\) that occur in environments other than \(C\_D\) represent overapplication of \(\mathcal{P}\): the rule applies even when the surface form lacks the relevant environment.

Rule-based serialism dealt with the problem of opacity by allowing intermediate representations, whose function was to supply environments which triggered or blocked the rule. Thus, underapplication is ‘explained’ by the absence of the environment at the crucial point in the derivation at which \(\mathcal{P}\) applies, while overapplication is accounted for by supposing the presence of the triggering environment at the relevant point in the derivation, after which the environment is destroyed.

Several OT solutions to the problem of opacity have been proposed, and they differ according to where they locate the conditioning environment of the opaque process. We will focus on Sympathy Theory (McCarthy 1999), which
CHAPTER 2. OPTIMALITY THEORY

is by far the best publicized of any of these approaches, in addition to being the most similar in spirit to Virtual Phonology, which is proposed in the next chapter. In Sympathy Theory, the conditioning environment is found in another non-optimal output candidate, which influences the actual surface form 'sympathetically'.

Some variants of OT, notably Kiparsky’s Stratal OT (Kiparsky 1997, 1998) import the notion of intermediate form directly: grammars consist of a series of strata, each one a total ranking on Con, such that the output of one evaluation is fed into a later stratum. In Stratal OT, the conditioning environment of an opaque process is located in a deeper stratum.

Another early OT attempt at dealing with opacity is McCarthy (1996) (drawing on earlier work by Lakoff 1993 and Goldsmith 1993) which recognizes multi-level constraints. This approach proposes that some markedness constraints may scan the input as well as the output to find the relevant environment. The proposal necessitates the parametrization of C such that in some languages it scans the output material alone, while in others it may scan the input.

Turbidity Theory (Goldrick 2000; Goldrick and Smolensky 1999) grows out of early OT work on the principle of Containment. Unlike Sympathy and Stratal OT, Turbidity remains committed to the strict two-level approach, which entails that “all unfaithful mappings must be surface motivated” (Goldrick 2000). This is achieved at the cost of a considerable increase in representational complexity, though, since Turbidity allows for the presence of covert material in the surface representation which can affect the pronunciation.

2.6.1 Sympathy Theory

Sympathy (McCarthy 1997b, 1998b, 1999) is an extension of Optimality Theory which deals with cases of phonological opacity. We introduce the concept by way of the by now classic example of opacity from Tiberian Hebrew (drawing on Prince 1975). In this language, epentheses into final clusters stands in a counterbleeding relationship to the deletion of a glottal stop in coda position. Both Epenthesis and Glottal Stop Deletion are independently motivated and occur transparently in Hebrew (40a,b). (40c) illustrates the interaction of the two.

(40) a. Epenthesis into final clusters
/melk/ → [melɛx] ‘king’

b. ?-Deletion in codas
/qara?/ → [qarr̥] ‘he called’

c. Epenthesis followed by ?-Deletion
/def?/ → [deʃe?] → [deʃe] ‘tender grass’

Epenthesis is driven by a constraint against tautosyllabic consonant clusters, *ComplexCoda.

(41) *ComplexCoda
Complex codas are disallowed.
2.6. PHONOLOGICAL OPACITY

Deletion of /?/ reflects the activity of \textit{CodaCond}, which bans glottal stops in final position.

(42) \textit{CodaCond}

? is disallowed in final position.

In the example (40c), epenthesis is triggered by the presence of a final cluster /-/?/ in the input. Subsequently, /?/ is deleted by a later rule, rendering the epenthesis opaque: the triggering environment is absent in the output.

Parallel OT is a two-level theory, recognizing only an input and an output level, so the part played by intermediate forms in the rule-based serial approach has to be modeled in some other way. Faithfulness constraints demand that the output preserve features of the input, while markedness constraints are purely output-oriented. Violations of faithfulness are therefore only mitigated to the degree they increase structural harmony (the \textit{harmonic ascent theorem} of Moreton 1996). Given this, opaque cases like (40c) present a serious problem for OT.

In order for the desired winner deSe to be optimized, it has to fare better than all other candidates. Under the classical OT model, however, there will, in cases of opacity, always be some candidate, the transparent candidate, which harmonically bounds the opaque optimal candidate over the set of markedness and faithfulness constraints. This is shown in (43). The expression \(*!\) designates the fatal point at which the desired optimal candidate is incorrectly eliminated from the contest. A frownie (\(\circ\)) marks the undesired but actual winner, the transparent candidate (43a).

(43) \textit{Transparent beats opaque desired winner}

<table>
<thead>
<tr>
<th>/def/?/</th>
<th>CodaCond (\vdash *\text{Complex} )</th>
<th>Max (\vdash \text{Dep} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\circ) def</td>
<td>:</td>
<td>* :</td>
</tr>
<tr>
<td>b. (##) deje</td>
<td>:</td>
<td>* : (*!)</td>
</tr>
<tr>
<td>c. def?</td>
<td>*!</td>
<td>:</td>
</tr>
<tr>
<td>d. deje?</td>
<td>*!</td>
<td>:</td>
</tr>
</tbody>
</table>

As well as faring equally well on markedness, the transparent candidate (43a) is also more faithful than the desired winner, candidate (43b). (43a) has a proper subset of (43b)'s marks, i.e. (43a) harmonically bounds (43b). Sympathy Theory remedies the failure of Classical OT to deal with opacity by making novel use of two independently motivated mechanisms in the OT model. The first of these is the large and inclusive pool of candidates generated by \textit{Gen} (§2.1.2). The second is Correspondence Theory (McCarthy and Prince 1995), reviewed in §2.5. Rule-based theory posits an intermediate level at which the rule is properly conditioned. Now, because the OT grammar generates an inclusive pool of candidates, a subset of those candidates will be characterized by transparent applications of the operation which is opaque in the optimal form. One of these candidates, which corresponds to the intermediate form of rule-based serialist accounts, is accorded formal status in Sympathy Theory, and termed the \textit{sympathetic candidate}.

As well as faring equally well on markedness, the transparent candidate (43a) is also more faithful than the desired winner, candidate (43b). (43a) has a proper subset of (43b)'s marks, i.e. (43a) harmonically bounds (43b). Sympathy Theory remedies the failure of Classical OT to deal with opacity by making novel use of two independently motivated mechanisms in the OT model. The first of these is the large and inclusive pool of candidates generated by \textit{Gen} (§2.1.2). The second is Correspondence Theory (McCarthy and Prince 1995), reviewed in §2.5. Rule-based theory posits an intermediate level at which the rule is properly conditioned. Now, because the OT grammar generates an inclusive pool of candidates, a subset of those candidates will be characterized by transparent applications of the operation which is opaque in the optimal form. One of these candidates, which corresponds to the intermediate form of rule-based serialist accounts, is accorded formal status in Sympathy Theory, and termed the \textit{sympathetic candidate}.
In the Hebrew case, this candidate is $defe$, since it transparently combines epenthesis with its trigger. The sympathetic candidate is so termed because it is capable of ‘sympathetically’ influencing the shape of the output. Sympathetic influence is achieved by constraints sensitive to the correspondence relations holding between candidates: Sympathy requires that there be a correspondence relation between the candidates from the same input. In the Hebrew example, the sympathy candidate $*depe$ is supplied with a correspondence relation to the entire candidate set. (Following standard practice, the sympathy candidate will be designated with a floriform dingbat, $*$. ) Outputs are thus subject to a constraint which demands that the output resemble the sympathetic candidate.

McCarthy has developed two theories of how this notion of ‘resemblance’ is to be encoded formally. In the original version of Sympathy Theory (McCarthy 1997a), resemblance is enforced by a set of Faithfulness constraints predicated on the correspondence relations holding between the sympathy candidate and the other candidates in the pool. This may be termed the Inter-candidate Faithfulness Theory. In the later version, resemblance is enforced by a constraint variously known as $*Sym$ or $*Cum$ which requires that the output share violations of Faithfulness with the sympathy candidate. This variant of Sympathy will be termed Cumulativity Theory. In the most recent published source on the theory (McCarthy 1999) compares and contrasts both variants of the theory.

In a way that will be made precise shortly, the transparent candidate (43a) must fare worse on sympathy than the opaque candidate (43b) because (43a) resembles $desep$ less than (43b). We shall return to the formalization of this notion of resemblance in a moment. Let us see how the sympathetic candidate itself is selected. The sympathetic candidate is the most harmonic candidate which satisfies some designated constraint, termed the selector, and marked with a star ($*$). The most harmonic candidate which satisfies the selector constraint $*f$ is designated $NF$.

McCarthy makes the stipulation that only faithfulness constraints can be selectors. The reason is that if markedness constraints are chosen as selectors, we predict the existence of Duke-of-York effects. These are derivations of the form $/A/ \rightarrow B \rightarrow A$, which, according to McCarthy, are unattested in natural language.\(^{18}\)

For Tiberian Hebrew, the selector must be $Max$. For the purposes of selecting the sympathetic form, the selector acts as if it is top-ranked, although the selector may itself occupy any place in the hierarchy. This is a crucial difference between Sympathy Theory and Virtual Phonology (see in particular §3.3). (44) below shows the full story.

\begin{equation}
\text{(44) Opaque candidate bests transparent candidate with Sympathy}
\end{equation}

(44d) is selected as the $\bullet$-candidate. Of all the candidates which satisfy the selector, $\ast$MAX, (44d) is most harmonic. Its only competitor (44c) adds a violation of highly-ranked $\ast$COMPLEX. The contest between transparent (44a) and opaque (44b) is decided by how the candidates fare on sympathy. (44b) wins owing to its greater resemblance to the sympathetic candidate, in spite of faring worse on faithfulness than (44b).

As mentioned, McCarthy (1999) discusses two significantly different ways of assessing resemblance to the sympathy candidate. One way is to assess resemblance directly by using ‘faithfulness’ constraints on the mapping from the $\bullet$-candidate to the output. These $\bullet$O-constraints can make reference to any feature of the $\bullet$-candidate and demand that the output reflect it. For example, in (44), the constraint $\bullet$O-MAX-V can be substituted for Sympathy. The violations will be exactly the same. The only difference is that resemblance to the sympathy candidate in this case is evaluated according to a specific dimension of similarity to the sympathy candidate: whether or not the output preserves the vowels present in the sympathy candidate.

Citing an argument due to Paul Kiparsky, McCarthy (1999) argues that this way of assessing resemblance overgenerates: Duke-of-York derivations are still possible. He therefore proposes to assess resemblance to the sympathy candidate indirectly, on the basis of the extent to which the output accumulates the unfaithful mappings of the sympathy candidate. The thrust of the idea is that the output contains a superset of the $\bullet$-candidate’s unfaithful mappings and this is captured in terms of a single constraint, $\bullet$Sym, which demands that the output is cumulative with respect to (i.e. accumulates the unfaithful mappings of) the sympathetic candidate.

On the cumulative approach, candidates are compared indirectly, in terms of their unfaithful input-output mappings. If the candidate to be evaluated, E-Cand, has a superset of the unfaithful mappings of the sympathy candidate $\bullet$-Cand, then E-Cand is cumulative with respect to $\bullet$-Cand. That is, E-Cand accumulates at least all of $\bullet$-Cand’s unfaithful mappings and has the option of acquiring additional ones.

Sympathy is evaluated by pairwise comparison of the set of unfaithful mappings relative to the input projected by each candidate E-Cand to be evaluated for sympathy with the unfaithful mappings of the sympathy ($\bullet$) candidate. Letting $\text{In}_{\text{UCand}}$ stand for the set of unfaithful mappings from the input In to the output candidate Cand, then for any pair of output candidates, $\langle \text{Cand}_1, \text{Cand}_2 \rangle$, then there are four logically possible relations that may obtain between $\text{In}_{\text{UCand}_1}$ and $\text{In}_{\text{UCand}_2}$ (McCarthy 1998b: 22):
CHAPTER 2. OPTIMALITY THEORY

Possible cumulativity and comparability relations between sets of unfaithful mappings

\[ \text{In} \cup \text{Cand}_1 = \text{In} \cup \text{Cand}_2 \]
In this case, Cand\(_1\) = Cand\(_2\), modulo properties that do not involve faithfulness (i.e. syllabification).

\[ \text{In} \cup \text{Cand}_1 \subseteq \text{In} \cup \text{Cand}_2 \]
Cand\(_1\) and Cand\(_2\) are distinct but comparable. Cand\(_2\) has a superset of the unfaithful mappings of Cand\(_1\). Cand\(_2\) is therefore cumulative with respect to Cand\(_1\).

\[ \text{In} \cup \text{Cand}_1 \supseteq \text{In} \cup \text{Cand}_2 \]
Cand\(_1\) and Cand\(_2\) are comparable. Cand\(_1\) has a subset of the unfaithful mappings of Cand\(_2\). Cand\(_1\) is therefore cumulative with respect to Cand\(_2\).

\[ \text{In} \cup \text{Cand}_1 \not\subset \text{In} \cup \text{Cand}_2 \]
Cand\(_1\) and Cand\(_2\) are non-comparable. There is no relation of cumulativity between them.

The crucial case is the one where one of the sets of unfaithful mappings is drawn from the \(\diamondsuit\)-candidate. \(\diamondsuit\)SYM is formulated in (46).

\(\diamondsuit\)SYM

Given a sympathetic candidate \(\diamondsuit\)-Cand, to evaluate E-Cand, both derived from the input In:
If \(\text{In} \cup \text{E-Cand} \subseteq \text{In} \cup \diamondsuit\)-Cand, then E-Cand’s performance on \(\diamondsuit\)SYM is proportional to the cardinality of the set \(\text{In} \cup \text{E-Cand} - \text{In} \cup \diamondsuit\)-Cand. If E-Cand and \(\diamondsuit\)-Cand are non-comparable in their unfaithful mappings, then E-Cand’s performance on \(\diamondsuit\)SYM is worse than that of any candidate that is comparable.

Two kinds of relation between \(\text{In} \cup \text{E-Cand}\) and \(\text{In} \cup \diamondsuit\)-Cand receive special treatment on \(\diamondsuit\)SYM. Where the relation between \(\text{In} \cup \text{E-Cand}\) and \(\text{In} \cup \diamondsuit\)-Cand is non-comparable, E-Cand scores an infinite number of violations. Where the relation is comparable, but non-cumulative (i.e. where E-Cand has at least one mark not shared by \(\diamondsuit\)-Cand), E-Cand also scores an infinite number of marks. Only in the case of a cumulative relationship between \(\text{In} \cup \text{E-Cand}\) and \(\text{In} \cup \diamondsuit\)-Cand is a finite number of marks assessed against E-Cand. The rationale for assessing marks in this way is to ensure that candidates which stand in a non-cumulative or non-comparable relation to \(\diamondsuit\)-Cand always fare worse on \(\diamondsuit\)SYM than any candidate that stands in a cumulative relation to \(\diamondsuit\)-Cand.

Using an alternative formalization suggested by Alan Prince, McCarthy (1999) proposes to resolve \(\diamondsuit\)SYM into two separate constraints. \(\diamondsuit\)CUMULATIVITY (\(\diamondsuit\)CUM) evaluates the extent to which the candidate for evaluation accumulates the unfaithful mappings of \(\diamondsuit\)-Cand, while \(\diamondsuit\)DIFF assesses marks for each additional violation incurred by E-Cand relative to the \(\diamondsuit\)-Cand.

\(\diamondsuit\)CUMULATIVITY

Given an IO Faithfulness constraint \(\star\)F which selects a sympathetic candidate \(\diamondsuit\)-Cand\(_F\), to evaluate a candidate E-Cand:

\(\diamondsuit\)DIFF
a. $\diamond \text{Cumul}_F$
E-Cand is cumulative with respect to $\diamond \text{-Cand}_F$. That is, $\diamond \text{-Cand}_F$ has a subset of E-Cand’s IO faithfulness violations.

b. $\diamond \text{Diff}_F$
Every IO faithfulness violation incurred by E-Cand is also incurred by $\diamond \text{-Cand}_F$.

c. Fixed Universal Ranking
$\diamond \text{Cumul}_F \gg \diamond \text{Diff}_F$

### 2.6.2 Local Conjunction

Another approach which has been useful in dealing with certain kinds of opacity is local conjunction (Smolensky 1993, 1995, 1997; Itô and Mester 1998, 1999; Kirchner 1996; Cowlhurst and Hewitt 1997; Łubowicz 1998).

Under local conjunction, two constraints are conjoined as a single composite constraint. This conjoined constraint is violated if and only if both of its conjuncts are violated within some local domain $\delta$.

The schematic tableau in (48) illustrates the violation profile for four candidates representing the logically possible combinations of violations on two constraints, $C_1$ and $C_2$.

(48)

<table>
<thead>
<tr>
<th>$C_1$</th>
<th>$C_2$</th>
<th>${C_1 \land C_2} _\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cand1</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>cand2</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>cand3</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>cand4</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

The conjunction is violated if and only if both of its components are violated. The notion of local domain is crucial to determining a violation: violation is always relative to some local domain (segment, syllable, morpheme, etc.) so that conjoined constraints do not combine violations of randomly distributed marks. Finally, a conjoined constraint is separately ranked from its conjuncts and is assumed universally to dominate them, i.e. $\{C_1 \land C_2\}_\delta \gg C_1, C_2$.

Locally conjoined constraints ban undesired configurations which might be dubbed the worst-of-the-worst (Prince and Smolensky 1993). For example, many languages permit both codas and labial stops. However, a subset of these languages do not permit labial codas. The combination of violating both \textit{NoCoda} and \textit{*Labial} within the same segment violates the conjoined constraint $\{*\text{NoCoda} \& *\text{Lab}\}_\text{seq}$. Despite its uses, local conjunction is a powerful tool, and it has proven difficult to constrain it.

A special case of local conjunction is local \textit{self}-conjunction, where the two components are drawn from one and the same constraint, to give a ‘power’ constraint $\{C_1 \land C_2\}_\delta$ or $C^2$. Alderete (1997) and Itô and Mester (1998) use local self-conjunction to reformulate the OCP (Obligatory Contour Principle), which has informed much work on auto-segmental phonology (e.g. Leben 1973; Goldsmith 1976; McCarthy 1986). The idea is that multiple violations of basic constraints are disallowed, but single violations are tolerated. Local conjunction
has also been argued for on the basis of the existence of chain shifts (see Kirchner 1996 and §11.2), and Lubowicz (1998) utilizes local conjunction in order to get derived environment effects in OT.
Chapter 3

Virtual Phonology

This chapter lays out a proposal, Virtual Phonology, for dealing with opacity in a fully parallel OT grammar. Although the theory owes much of its conceptual apparatus to Sympathy Theory, there are crucial architectural and empirical differences between the two.

The first section, §3.1 lays out the basic architecture and empirical predictions of Virtual Phonology. Considerable emphasis is laid on the predictive differences between Virtual Phonology and Sympathy Theory: the two theories make divergent predictions concerning the possibility of two kinds of rule-interaction, to wit, multiple opacity and rule-sandwiching. These are addressed in considerable detail in the subsequent sections.

§3.2 addresses the problem of multiple opacity and includes an extended critique of McCarthy’s analysis of multiple opacity in the phonology of the Californian language Yawelmani Yokuts. An analysis in the terms of Virtual Phonology is proposed.

§3.3 examines rule-sandwiching and argues for the authenticity of the phenomenon. The phenomenon is beyond the purview of Sympathy Theory. Several examples are documented and analyzed using the tools supplied by Virtual Phonology.

§3.4 examines cases in which the virtual form must satisfy some virtual markedness constraint, and also addresses the problem of apparent language-particular restrictions on the input.

§3.5 addresses issues of learnability which are posed by the model. Specifically, it is argued that the absence of non-trivial Duke-of-York effects arises out of constraints on the learnability of such systems, not from the formal properties of the grammar itself.

§3.6 concludes the presentation of the theory by exploring some residual problems of overgeneration.

3.1 The Theory

3.1.1 Architecture

The architectural properties of Virtual Phonology are enumerated below. Some of these are a first approximation only, and will be revised later.
1. **Dyadism**
   The output is a pair or binary array of representations consisting of a unique **surface form** ($S$) and a unique **virtual form** ($V$). The $S$-form encodes overt structure, while the $V$-form encodes covert structure.\(^1\)

2. **Covertness**
   The role of the $V$-form is to encode, in transparent fashion, phonological generalizations which may be non-surface-true or non-surface-apparent. The $V$-form thus covers the same functions as the sympathetic ($*$) candidate in Sympathy Theory.

3. **Uniqueness**
   Unlike Sympathy Theory, which admits of the selection of multiple $*$-candidates, the $V$-form is unique. While the $*$-candidate may encode at most a single recuperatum\(^2\), the $V$-form may in principle encode more.

4. **Split IO-Faithfulness**
   IO-Faithfulness constraints are split into $IV$- and $IS$-Faith constraints. The $V$-form is subject to constraints on $IV$-Faith. These play the role of the selector ($*$) constraints of Sympathy Theory, singling out some recuperatum for preservation in the $V$-form. The $S$-form is subject to constraints on $IS$-Faith.

5. **Defeasibility**
   Unlike selector constraints, which, for the purposes of sympathetic selection, always pattern as undominated, $IV$-Faith is in principle violable and can be ranked anywhere in the hierarchy.

6. **VS-Faithfulness**
   The $S$-form is subject to constraints on VS-Faith. In this regard, Virtual Phonology does not differ in essence from the original version of Sympathy Theory which employed constraints on intercandidate faithfulness.

7. **Virtual Markedness**
   Virtual Phonology postulates the existence of a set of virtual markedness constraints, whose jurisdiction is restricted to the $V$-form. These fulfill the same role as markedness selector constraints in some versions of Sympathy Theory.

---

\(^1\)Terms of virtual-phonological discourse will be abbreviated using Fraktur type. The input will also be designated in this way by $I$ for reasons of typographic harmony.

\(^2\)The term recuperatum refers to any element of phonological representation which is absent on the surface but present opaquely. Virtual property is synonymous. Three examples will suffice here to elucidate the meaning of the term.

In the Tiberian Hebrew example of final glottal stop deletion, the recuperatum is $?$. Recuperatum may also refer to a feature which maps unfaithfully on the surface. For example, in Yawelmani, long vowels which are underlyingly [+high] and condition Rounding Harmony, are lowered on the surface. They nevertheless still condition Rounding Harmony as if [+high]. In this case, the recuperatum is the value [+high]. Finally, the recuperatum may also refer to a zero element. This is necessarily the case where epenthized elements pattern as if absent. For example, Hagstrom (1997) discusses the metrical invisibility of epenthetic $e$ in Mohawk. Stress in Mohawk is usually penultimate, but where $e$ is inserted to break up a biconsonantal cluster, stress is assigned as if the vowel was not there.
3.1. THE THEORY

Unlike in Sympathy Theory, the virtual form is not a failed candidate, but is generated alongside the actual surface form as part of the output. At this point, Virtual Phonology is the same as the theory proposed by (Jun 1999) in its essential architectural claims. The presence or absence of a sympathetic candidate in the tableau is contingent on some faithfulness constraint being designated as a selector. In Virtual Phonology on the other hand, the virtual form is an obligatory feature of all outputs. The default is identity between virtual and surface form, i.e. for transparency to be preferred. We may assume that the initial state of the language faculty, all \textsc{IS}-\textsc{Faith} constraints outrank all markedness constraints. \textsc{IS}- and \textsc{IV}-\textsc{Faith}, on the other hand may be taken to start off low. Virtual Phonology thus obviates the need to posit intercandidate correspondence.

A brief note on the transcriptional practices used here: where a form is specifically intended as virtual, it will be enclosed in vertices, e.g. $\langle\text{deSe}\rangle$. The corresponding surface form will be written optionally in square brackets, e.g. $[\text{deSe}]$. Virtual markedness constraints will be enclosed in vertices, e.g. $[*X]$, the version of $*X$ whose effect is exclusive to the virtual form. McCarthy (1999: 348) mentions two disadvantages of Jun's theory, which, of course, carry over directly to Virtual Phonology.

\begin{quote}
Jun's theory also has some distinct disadvantages: since every faithfulness constraint comes in two versions, it allows several faithfulness constraints to act in concert as the selector [...]; and situations of multiple opacity, like Yokuts [...], cannot be analysed, unless the hierarchy evaluates ordered $n$-tuples for some arbitrary value of $n > 2$.
\end{quote}

(McCarthy 1999: 348)

It will be shown that the first prediction, the possibility of several faithfulness constraints acting in concert as the selector, is borne out by evidence from Saami (§11.5). McCarthy's second point, that "situations of multiple opacity [...] cannot be analysed, unless the hierarchy evaluates ordered $n$-tuples for some arbitrary value of $n > 2" will be addressed in §11.4.2.

3.1.2 Predictions

There are two particular architectural features of Virtual Phonology which lay the basis for empirical predictions which diverge from those of Sympathy in interesting and significant ways. These are uniqueness and defeasibility. Let us briefly enumerate the predictions of each of these principles in turn. Uniqueness predicts the following:

- Multiple non-interacting opaque interactions are possible.
- Multiple interacting opaque interactions are impossible.

Uniqueness only gets things half right. Although it allows for multiple non-interacting opacities, which are attested in Saami, it doesn't allow for multiple interacting opacities, which is clearly wrong.

---

3 Jun calls his proposal \textit{Generalized Sympathy}. Somewhat confusingly, he refers to the virtual and surface form as the \textit{base} and \textit{derived} form respectively. Jun's theory thus also predicts the existence of rule sandwiching effects, although it would appear he is unaware of this prediction of the theory.
It is easy to see how the second prediction follows from the *Uniqueness* criterion. In cases of multiple interacting opacity, there have to be two opaque (intermediate) forms: one form $\Phi_1$, whose function is transparently to encode some phonological generalization $\gamma_1$ not holding of the surface form $\Sigma$, and, another form $\Phi_2$, whose role is to encode transparently some other phonological generalization $\gamma_2$ not holding of $\phi_1$. Yawehman Yokuts furnishes us with a rule-interaction of this type, which, without reinterpretation, constitutes a counterexample to *Uniqueness*.

*Defeasibility* of $\mathcal{IF}$ predicts that *rule-sandwicking* interactions, where $\mathcal{P} > \mathcal{Q} > \mathcal{R}$, and $\mathcal{P}$ feeds/bleeds $\mathcal{Q}$, and $\mathcal{R}$ counterfeeds/counterbleeds (opacifies) $\mathcal{Q}$, are permitted. These predictions are crucially distinct from those made by Sympathy Theory. In particular,

- Sympathy Theory predicts, incorrectly, the absence of multiply opaque interactions.
- Sympathy predicts, incorrectly, the absence of rule-sandwicking interactions.

While the $\mathcal{I}$-form and the $\bullet$-Candidate are similar in terms of their roles in their respective theories, their formal properties are quite distinct, and, as a result, the two theories diverge in their empirical predictions. The logic of Sympathy Theory not only permits the selection of several $\bullet$-Candidates: multiple opacity *necessitates* the selection of multiple $\bullet$-candidates. Let us just take a moment to reflect why this is the case. The effect of the selector constraint $\mathcal{S}\Phi$ is to 'suppress' the activity of some specific highly-ranked markedness constraint, or set of markedness constraints, which have in common that they militate in some way against the specific element (the recuperatum $\varrho_i$) whose faithful preservation is required by $\mathcal{F}$. Crucially, $\mathcal{F}$ does not suppress the activity of other markedness constraints targeting structures other than $\varrho_i$. This is because the selector $\mathcal{S}\Phi$ picks the most harmonic candidate which satisfies $\mathcal{S}\Phi$, and completely ignores the well-being of recuperata other than $\varrho_i$. Other recuperata will feel the full force of whichever markedness constraints militate against them, with the result that Sympathy Theory imposes a limit of one recuperatum per sympathetic candidate. Multiple opacity in Sympathy Theory, then, implies the existence of multiple sympathetic candidates. This property turns out to be a liability of Sympathy Theory, since it turns out that multiple opacity of any kind is beyond the purview of the theory.

In Virtual Phonology, the role of the selector constraints of Sympathy Theory is played by Faithfulness constraints predicated on the correspondence between the input (3) and the virtual form (21). Unlike selector constraints, however, constraints on $\mathcal{IF}$ are fully integrated into the rest of the grammar: they are defeasible. This principle does not in actual fact add anything new to the theory of grammar as envisioned by OT. It is simply the OT tenet that constraints are violable, but we labour the point here to draw attention to the sharp distinction between Sympathy Theory and Virtual Phonology with respect to this property. The defeasibility of $\mathcal{IF}$ allows for much finer control of what goes into the virtual form compared with what goes into the $\bullet$-candidate. Selector constraints behave as if inviolable with respect to selecting the sympathetic candidate, with the result that, in the sympathetic candidate, preservation of the recuperatum required by $\mathcal{S}\Phi$ is tyrannically enforced, with
3.1. THE THEORY

the consequence that interesting interaction with other constraints is precluded in Sympathy Theory. The sympathetic candidate selected by \(*F\) must preserve \(\varphi_i\). This is not the case in Virtual Phonology, since the effects of the \(\mathcal{VS}\)-Faithfulness constraint will emerge only in the absence of conflict with any higher-ranked constraint. The result is that Virtual Phonology is capable of generating rule-sandwiching patterns which Sympathy Theory is unable to handle. The move has a conceptual pay-off as well, since it obviates the anomalous suspension of OT violability which pertains to selectship.

The relation between the virtual form and the surface form is governed by \(\mathcal{VS}\)-Faith. \(\mathcal{VS}\)-Faith thus plays the role of the \(\mathcal{O}\)-Faith or Cumulativity constraints of Sympathy Theory.

In tableaux, the terms of the array will be separated by the symbol ‘∼’. Occasionally, information content exceeds the available space. In this case, a notational variant of the single tableau will be used, in which the optimization of the virtual form is factored out. Thus, (49) and (50) are essentially equivalent as tableaux representing the optimization of the output array \([\text{def}!]? \sim \text{def}e\) from the input \(/\text{def}?/\) (cf. §2.6.1).

(49)

<table>
<thead>
<tr>
<th></th>
<th>/\text{def}?/</th>
<th>(\mathcal{VS})-Max-V</th>
<th>(\mathcal{O})-Max-C</th>
<th>\text{CodCond}</th>
<th>*Complex</th>
<th>(\mathcal{IS})-Max-C</th>
<th>(\mathcal{IS})-Dep-V</th>
<th>(\mathcal{IV})-Dep-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[\text{def}?!] \sim \text{def}</td>
<td>*! :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>(\Rightarrow) [\text{def}?!] \sim \text{def}</td>
<td>:Br</td>
<td>:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>[\text{def}?!] \sim \text{def}?</td>
<td>*! :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>[\text{def}?!] \sim \text{def}?!</td>
<td>:</td>
<td>**! :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>[\text{def}] \sim \text{def}</td>
<td>*! :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td>[\text{def}] \sim \text{def}?!</td>
<td>:</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td>[\text{def}] \sim \text{def}?</td>
<td>*! :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h.</td>
<td>[\text{def}] \sim \text{def}?!</td>
<td>:</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i.</td>
<td>[\text{def}] \sim \text{def}</td>
<td>:</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j.</td>
<td>[\text{def}] \sim \text{def}?!</td>
<td>:</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k.</td>
<td>[\text{def}] \sim \text{def}?</td>
<td>:</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l.</td>
<td>[\text{def}] \sim \text{def}?!</td>
<td>:</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m.</td>
<td>[\text{def}?!] \sim \text{def}</td>
<td>:</td>
<td>* : *!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n.</td>
<td>[\text{def}?!] \sim \text{def}</td>
<td>:</td>
<td>* : *!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o.</td>
<td>[\text{def}?!] \sim \text{def}?</td>
<td>:</td>
<td>*! : **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p.</td>
<td>[\text{def}?!] \sim \text{def}?</td>
<td>:</td>
<td>*! : *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2 Multiple Opacity

Sympathy Theory is ultimately unable to deal with multiple opacity. The latter claim may appear easily refuted by a cursory glance at the recent published literature: indeed McCarthy (1999) does deal with a case of multiple opacity in Yawelmani Yokuts. However, a close analysis reveals the account to be fundamentally flawed, and the claim that Sympathy Theory is inadequate for the task of dealing with multiple opacity stands.

There is a long tradition of generative studies of Yawelmani Yokuts, whose multiple opacity provides a rich testing ground for phonological theories. The original source on the language is Newman (1944), although various aspects of Yawelmani phonology have been treated by Kuroda (1967), Kisseberth (1969, 1970, 1973), Kenstowicz and Kisseberth (1977, 1979), Archangeli (1984, 1991), McCarthy (1999).4

After laying out the relevant aspects of the morphophonemics of the Yawelmani Yokuts verb, we will review McCarthy’s Sympathy-theoretic analysis of multiple opacity in Yawelmani Yokuts. Much of the activity in Yawelmani involves the interaction of three phonological processes. We will start by examining each of these in turn, beginning with Vowel Harmony (§3.2.1), Vowel Shortening (§3.2.2), and Long High Vowel Lowering (§3.2.3). §3.2.4 turns to

---

4Consonants transcribed with an underdot, e.g. ţ, are alveolar. The vowels /e/ and /o/ are phonetically mid-open [ɛ] and [ɔ].
3.2. MULTIPLE OPACITY

the opaque interaction between the three processes and reviews McCarthy’s recent published account in considerable detail. Finally, §3.2.5 proposes a Virtual Phonology account of the Yawelmani data.

3.2.1 Vowel Harmony

In Yawelmani, a suffixal vowel harmonizes in roundness to a preceding \([\alpha\text{high}]\) root vowel according to a rule of Vowel Harmony,

\[(51) \quad \text{Yawelmani Vowel Harmony} \]
\[
V_{[\alpha\text{high}]} \rightarrow \left[ +\text{round} +\text{back} \right] / \left[ +\text{round} \right] C_o
\]

A suffix containing a high vowel assimilates in roundness to the preceding root vowel, just in case the root vowel is also \([+\text{high}]\). For example, the nonfuture suffix alternates between -hin\(\sim\)hun, the -hun variant appearing following a root whose vowel is \(u\) (52d). Elsewhere, the -hin allomorph is found (52a–c).

\[(52) \]
\[
a. \quad \text{xat-hin} \quad '\text{eat-NONFUT}' \\
b. \quad \text{bok-hin} \quad '\text{find-NONFUT}' \\
c. \quad \text{xil-hin} \quad '\text{tangle-NONFUT}' \\
d. \quad \text{dub-hun} \quad '\text{lead by the hand -NONFUT}'
\]

Similar alternations are instantiated by the suffixes -mi\(\sim\)mu 'having . . . -en' and -nit\(\sim\)nut 'will be . . . -en', demonstrating the generality and productivity of the process.

Similarly, a suffix containing a non-high vowel assimilates in roundness to the preceding root vowel just in case the root vowel is also \([-\text{high}]\). The dubitative suffix, for example, alternates between -al\(\sim\)ol, the -ol variant occurring following a root whose vowel is \(o\) (53d). Elsewhere, the -al variant is found (53a–c).

\[(53) \]
\[
a. \quad \text{max-al} \quad '\text{procure-DUB}' \\
b. \quad \text{hud-al} \quad '\text{recognize-DUB}' \\
c. \quad \text{giy'-al} \quad '\text{touch-DUB}' \\
d. \quad \text{k'oP'-ol} \quad '\text{throw-DUB}'
\]

A similar pattern of alternation is instantiated by the suffixes -xa\(\sim\)xo 'let us . . . ', -k'a\(\sim\)k'o IMPERATIVE, and -taw\(\sim\)taw ' . . . -ing'.

Following Archangeli and Suzuki (1997: 207ff.) and McCarthy (1999), the pattern of harmony arises out of the interaction of the constraints in (54) to (56).

\[(54) \quad \text{ALIGN-COLOR} \]
\[
\text{Align(Color, R, Word, R)} \\
\text{Every instance of Color \([-\text{round, back}]\) is final in some word.}
\]

\[(55) \quad \text{RD/\alpha H} \]
\[
\text{Every path including \([\text{round}]\) includes \([\alpha\text{high}]\).} \\
\text{Every token of \([\text{round}]\) must be linked to vowels of the same height.}
\]
CHAPTER 3. VIRTUAL PHONOLOGY

(56) \textbf{Ident}[Color]

Two segments standing in IO correspondence have identical values for Color.

The basic interaction between these constraints is illustrated in (57) and (58). (57) illustrates the pattern for the input /dub+mi/, in which both root and suffix vowel are [+high].

(57)

<table>
<thead>
<tr>
<th>Input</th>
<th>Rd/αH1</th>
<th>ALIGN-Color</th>
<th>Ident[Color]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. dubmu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. dubmi</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

Where the root and suffixal vowel are the same height, Color may spread in satisfaction of ALIGN-Color, without incurring a violation of Rd/αH1. Ident[Color] is dominated by ALIGN-Color, since spreading occurs at the expense of preserving the input Color specification on the suffix.

(58) illustrates the same point for the input /bok'+al/, where both root and suffix vowel are [−high].

(58)

<table>
<thead>
<tr>
<th>Input</th>
<th>Rd/αH1</th>
<th>ALIGN-Color</th>
<th>Ident[Color]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. bok'ol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. bok'al</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

When the root and suffixal vowels differ in height, harmony is blocked, i.e. ALIGN-Color is violated in favour of satisfying the more highly ranked constraint Rd/αH1. In (59), the simultaneous satisfaction of Rd/αH1 and ALIGN-Color is impossible, because the root vowel is [−high], while the suffixal vowel is [+high]. The actual winner fails to implement Rounding Harmony, attesting to the higher priority of Rd/αH1.

(59)

<table>
<thead>
<tr>
<th>Input</th>
<th>Rd/αH1</th>
<th>ALIGN-Color</th>
<th>Ident[Color]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. bok'mi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. bok'mu</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

(60) illustrates the inverse situation. As in (59), the height specifications on root and suffix conflict, but here the root vowel is [+high] while the suffixal vowel is [−high].

(60)

<table>
<thead>
<tr>
<th>Input</th>
<th>Rd/αH1</th>
<th>ALIGN-Color</th>
<th>Ident[Color]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. hudal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. hudol</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>
3.2. MULTIPLE OPACITY

Top-ranked \text{RD}/\alpha\text{Hi} will also eliminate the rounded allomorphs when the
suffixed vowel is specified as [+round] in the input. This possibility must also
be considered under \textit{Richness of the Base}, since suffixed rounding is a non-
contrastive property, predictable from the properties of the root. The grammar
will, for example, neutralize the inputs /bok'+mu/ and /bok'+mi/ to \textit{bok'\text{mi}}.
Compare (61) with (59) above.

\begin{verbatim}
(61)
\begin{tabular}{|l|c|c|c|}
\hline
/bok'+mu/ & \text{RD}/\alpha\text{Hi} & \text{ALIGN-COLOR} & \text{IDENT[Color]} \\
\hline
a. & \# & \text{bok'\text{mi}} & *! \\
\hline
\end{tabular}
\end{verbatim}

3.2.2 Vowel Shortening

Root vowels alternate in length. Vowel Shortening applies in a closed syllable
according to a rule V:\text{--V/\text{C}]}\sigma.

\begin{verbatim}
(62) a. sa:p-al \text{`burn-dub'}
    dos-en \text{`report-fut'}
    pana:-hin \text{`arrive-nonfut'}
    c’uyo:-hun \text{`urinate-nonfut'}

b. panal \text{dub}
    pana:l \text{dub}
\end{verbatim}

Closed syllable shortening reflects the activity of a markedness constraint
\(\#[\mu\mu]\), which dominates the faithfulness constraint \text{MAX-\mu}.

\begin{verbatim}
(63) \#[\mu\mu]
\end{verbatim}

Trimoraic syllables are disallowed.

(64) illustrates the interaction between the two constraints.

\begin{verbatim}
(64)
\begin{tabular}{|l|c|c|}
\hline
/pana:+l/ & \#[\mu\mu] & \text{MAX-\mu} \\
\hline
a. & \# & \text{panal} & *! \\
\hline
\end{tabular}
\end{verbatim}

3.2.3 Long High Vowel Lowering

Long high vowels are lowered by a context-free rule \[+\text{high}] \text{--}[−\text{high}] in ac-
cordance with the markedness constraint \text{LONG/−HIGH} in (65).

\begin{verbatim}
(65) \text{LONG/−HIGH}
\end{verbatim}

Long vowels must be [−high].

(65) dominates \text{IDENT[high]}. The interaction of the two is illustrated in (66).

\begin{verbatim}
(66)
\begin{tabular}{|l|c|c|}
\hline
/\text{Trili}+hin/ & \text{LONG/−HIGH} & \text{IDENT[high]} \\
\hline
a. & \# & \text{\text{Trili}hin} & *! \\
\hline
\end{tabular}
\end{verbatim}
Note that long high vowels do not undergo repair by shortening. Hence, it must also be the case that Max-μ dominates Ident[high].

(67)

<table>
<thead>
<tr>
<th>Input</th>
<th>*/?ili:+hin/</th>
<th>Max-μ</th>
<th>Ident[high]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>✗  ?ile:hin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>✗  ?ili:hin</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

In rule-based terms, Lowering must precede Shortening. In (67), this interaction is transparent. Shortening and Lowering also interact opaquely, in counter-bleeding fashion. The derivation of ?ilel is shown in (68). The corresponding OT tableau is given in (69).

(68)

<table>
<thead>
<tr>
<th>Input</th>
<th>*/?ili:+l/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowering</td>
<td>?ile:l</td>
</tr>
<tr>
<td>Shortening</td>
<td>?ilel</td>
</tr>
<tr>
<td>Output</td>
<td>?ilel</td>
</tr>
</tbody>
</table>

(69)

<table>
<thead>
<tr>
<th>Input</th>
<th>*/?ili:+l/</th>
<th><em>[μμμ] Long/ /</em>$ ν$*</th>
<th>Max-μ</th>
<th>Ident[hi]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>✗  ?ilel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>✗  ?ili:ll</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>✗  ?ilel:ll</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>✗  ?ili:ll</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.4 The Opaque Interaction of Vowel Harmony and Vowel Shortening

As established in the preceding section, underlyingly long high vowels undergo Lowering. Nevertheless, with respect to Vowel Harmony, they pattern as if they were [+high]. This is the case with roots such as /c’um-/ ‘destroy’ and /mik’-/ ‘swallow’. In the dubitative, for example, /c’um-/ surfaces as c’ozmol. The form is non-surface-true: the conditions for the transparent application of Vowel Harmony (which would have given ✗c’ozmol) are fulfilled, yet Vowel Harmony fails to apply. Similarly, in the nonfuture, /c’um-/ surfaces as c’omhun. This form is non-surface-apparent: the suffix harmonizes in colour, but the reason for this is not apparent from the surface phonology, since the conditions for Vowel Harmony (sameness of vowel height) are not fulfilled.

Seen in terms of ordered rules, Vowel Harmony must precede Lowering. Sample derivations are given in (70).
3.2. **MULTIPLE OPACITY**

(70)

<table>
<thead>
<tr>
<th>Input</th>
<th>/c’um+al/</th>
<th>/c’um+it/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel Harmony</td>
<td>n/a</td>
<td>c’umut</td>
</tr>
<tr>
<td>Lowering</td>
<td>c’omal</td>
<td>c’omut</td>
</tr>
<tr>
<td>Output</td>
<td>[c’omal]</td>
<td>[c’omut]</td>
</tr>
</tbody>
</table>

In Sympathetic-theoretic terms, the sympathetic candidate has to be the one which preserves the height of the vowel. The following tableaux are adapted from McCarthy (1999). Whereas McCarthy uses intercandidate faithfulness, I have used $\mathbf{\ast Sym}$ here. (71) illustrates the non-surface-true type of opacity.

(71)

<table>
<thead>
<tr>
<th>/c’um+al/</th>
<th>LONG/$\mathbf{\ast Sym}$/</th>
<th>$\mathbf{\ast}$</th>
<th>Ident[hi]</th>
<th>Align-Color</th>
<th>Ident[col]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. $\ast$ c’omal</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. c’omol</td>
<td>**!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c. $\ast$ c’umal</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>d. c’umol</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

$\mathbf{\ast Ident[hi]}$ is (71c). Of the two candidates which obey $\mathbf{Ident[hi]}$, (71c) is the more harmonic, since its competitor, (71d) scores a mark on Rd/αHi.

(72) illustrates the non-surface-apparent form of opacity.

(72)

<table>
<thead>
<tr>
<th>/c’um+it/</th>
<th>LONG/$\mathbf{\ast Sym}$/</th>
<th>$\mathbf{\ast}$</th>
<th>Ident[hi]</th>
<th>Align-Color</th>
<th>Ident[col]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. $\ast$ c’umut</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. c’umit</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c. $\ast$ c’umut</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>d. c’umut</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Bearing this analysis in mind, let us move on to a multiple process in which all three processes (Vowel Harmony, Lowering and Shortening) interact. The central example is furnished by the derivation of *potshum* from underlying */ptu+n+hin/ *‘steal-NONFUT’. For the purposes of comparison, the derivation of transparent *doshin* from */dos+hin/ *‘report-NONFUT’ is included in (73).
The surface form \textit{Pothun} is thus doubly opaque. Vowel Harmony makes reference to the underlying \(+\text{high}\) specification on the root vowel. Shortening must also make reference to the underlying \textit{weight} of the vowel, otherwise the root vowel would shorten transparently to [u]. As a result, there must be two distinct selectors in the system, \textit{\#Ident[hi]} and \textit{\#Max-\textmu}.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Input & /\textit{Put}+hin/ & /\textit{dos}+hin/ \\
\hline
\textbf{Vowel Harmony} & \textit{Pothun} & n/a \\
\textbf{Lowering} & \textit{Pothun} & n/a \\
\textbf{Shortening} & \textit{Pothun} & doshin \\
\hline
\textbf{Output} & [\textit{Pothun}] & [doshin] \\
\hline
\end{tabular}
\end{table}

The sympathetic candidates are (74b), selected by \textit{\#Ident[hi]}, and (74c), selected by \textit{\#Max-\textmu}. (74b) incurs violations of \textit{Ident[hi]} and \textit{Ident[round]}. (74c) incurs a violation of \textit{Max-\textmu}. The optimal candidate, (74a) accumulates the faithfulness violations of both. Both of the transparent candidates, (74b) and (74d) fail to accumulate the faithfulness violations of both sympathy candidates, and therefore score an infinity of marks on \textit{\#Sym}. Notice that neither of the sympathetic candidates correspond to intermediate representations in the derivational analysis, emphasizing the parallel nature of the theory. That is neither of the sympathy candidates stand in a relationship of cumulative to the other.

Notice what has been done. \textit{Max-\textmu} has been crucially invoked as a selector. This has drastic consequences for the treatment of \textit{c\'o:mal} discussed earlier. Since the selector constraint is so specified for the entire grammar, we must now consider the effect of \textit{Max-\textmu}'s selectorship in the treatment of \textit{c\'o:mal}. As it turns out, incorporating \textit{Max-\textmu} as selector has disastrous consequences, as shown in (75).
3.2. MULTIPLE OPACITY

The other sympathetic candidate, (75c), is faithful to the input. Now let us evaluate the desired winner, (75a), in terms of its cumulative to each of the sympathetic candidates. (75a), is fully cumulative with respect to (75c), but not to (75b), since it crucially fails to accumulate violation of Ident[round]. By virtue of this, the desired winner will score an infinite number of marks on Sym. All candidates in fact fare equally abysmally on Sym. As a result, Sym is unable to do what is required of it, that is, favour the desired opaque candidate over the undesired transparent one. The consequence is that the transparent candidate, (75b), sails home victorious, precisely the result we don’t want. Splitting Sym into separate constraints, Cumul and Diff, won’t be of any help to us here either, since the undesired actual winner scores a subset of the desired winner’s marks on Diff.

In conclusion, the Sympathetic-theoretic account of multiple opacity turns out to be unworkable.

3.2.5 A Virtual Phonological Account of Opacity in Yawel-mani Yokuts

Sympathy Theory isn’t up to the task of dealing with the multiple opacity of Yawel-mani Yokuts. Can Virtual Phonology furnish us with a superior account?

At first blush, it seems that Virtual Phonology doesn’t hold out a great deal of promise either. The reason has to do with the kind of rule interaction apparently needed to account for opaque cases such as Potun: double counterbleeding on the environment. The rule-based account invokes two crucial intermediate steps on the way to the correct output form, and the Sympathy-theoretic account would seem to require two distinct sympathetic candidates. At some level

---

*I would like to thank Matt Goldrick for pointing this out to me.*
the root vowel has to be [+high], so that Vowel Harmony has a chance to apply. In the rule-based account, this is implemented by ordering Vowel Harmony prior to Lowering. In the Sympathy-theoretic account, *Ident[high] selects a sympathetic candidate, *Puthun, where Rounding Harmony applies transparently. At another level, however, the root vowel must be [−high], since at output, the root vowel shortens opaque to o, not transparently u. In rule-based terms, Shortening must be preceded by the Lowering of input /u:/ In the Sympathy-theoretic analysis, *Max−µ selects *Pothin. Either way, we are faced with apparently conflicting demands: the root vowel would appear to have to be [+high] on no less than two distinct levels, neither of them surface. Given Virtual Phonology’s insistence on the uniqueness of the virtual form, this kind of opacity would seem to be beyond its grasp. Since we only have one covert representation to play with, the vowel would have to be simultaneously [+high] (to get Vowel Harmony) and [−high] (to get the right quality of shortened vowel), clearly a paradox.

Relinquishing this uniqueness is an unappealing option, since this would be to give up any principled restriction on the number of opaque forms a grammar can support.

Yet all is not lost. In fact, there is a way of encoding both into a single virtual representation, but it involves a departure from the traditional analysis of Yawelmani in terms of Lowering. The idea is to build into the Ω-form a representation from which we can plausibly get both Vowel Harmony and Shortening at once. In rule-based terms, the Lowering rule would be replaced by a two-step process involving first a rule of Diphthongization which maps underlying /i: u:/ to intermediate ei ou, followed by a Monophthongization rule which maps intermediate ei ou onto e; o. Diphthongization is not motivated from the surface phonology of Yawelmani, but is nevertheless a plausible and natural intermediate stage. (76) illustrates how the rule-based account would work.

(76)

<table>
<thead>
<tr>
<th>Input</th>
<th>/?u:t+hin/</th>
<th>/c'um+al/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rounding Harmony</td>
<td>?uthun</td>
<td>n/a</td>
</tr>
<tr>
<td>Diphthongization</td>
<td>?outhun</td>
<td>c'oumal</td>
</tr>
<tr>
<td>Shortening</td>
<td>?othun</td>
<td>n/a</td>
</tr>
<tr>
<td>Monophthongization</td>
<td>n/a</td>
<td>c'oumal</td>
</tr>
<tr>
<td>Output</td>
<td>[?othun]</td>
<td>[c'oumal]</td>
</tr>
</tbody>
</table>

For the Virtual Phonology account, the diphthong would have the advantage of collapsing the trigger of Vowel Harmony and the target of Shortening into one. Because the weak mora of the diphthong is [+high], it will still be able to trigger Vowel Harmony on a following [+high] suffixal vowel and block it on a following [−high] vowel. Second, because the first component of the diphthong is [−high], Shortening can apply naturally to return a [−high] vowel.

We must also assume a Monophthongization rule which eliminates diphthongs on the surface by spreading the [−high] value of the first component of the diphthong onto the second component. In Virtual Phonology, this eliminates the necessity of positing more than one virtual form.

Let us suppose, then, that the output of the grammar given /?u:t+hin/ as input is the ⟨Ω, S⟩ pair ([?othun], [?othun]). Both the Ω-form and the
S-form satisfy Long/−High, but do so in different ways. The \( S \)-form satisfies Long/−High by diphthongizing, introducing a violation of Ident[high] in the strong mora of the vowel. I assume that Ident[high] is split according to whether the locus of the faithfulness violation is the strong or the weak mora of the underlying long vowel.

(77) \( S - \mu_s \text{-Ident}[\text{high}] \)
If the vowel \( V \) associated with \( \mu_s \) in the input is [\( \alpha \text{high} \)] in the input, then the output correspondent of \( V \) must be [\( \alpha \text{high} \)].

(78) \( S - \mu_w \text{-Ident}[\text{high}] \)
If the vowel \( V \) associated with \( \mu_w \) in the input is [\( \alpha \text{high} \)] in the input, then the output correspondent of \( V \) must be [\( \alpha \text{high} \)].

If underlying long high vowels are diphthongized virtually, we have to assume two things. First, \( S - \mu_s \text{-Ident}[\text{high}] \) must be dominated by Long/−High (giving us the diphthongization of /u:/ to ou) and, second, that \( S - \mu_w \text{-Ident}[\text{high}] \) dominates *Diphthong. The latter ranking prevents the monophthongization of [ou] to [o:] in the \( S \)-form. (79) and (80) illustrate the optimization of the \( VS \)-pair \([c\'oumal],[c\'omal]\).

(79)

<table>
<thead>
<tr>
<th>/c'um+al/</th>
<th>Long/−High</th>
<th>( S - \mu_s \text{-Ident}[\text{high}] )</th>
<th>( S - \mu_w \text{-Ident}[\text{high}] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[c'omal]</td>
<td>: *!</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>[c'omol]</td>
<td>: *!</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>[c'umal]</td>
<td>*! :</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>[c'umol]</td>
<td>*! :</td>
<td></td>
</tr>
</tbody>
</table>
| e. | [c\'oumal] | : * : | *
| f. | [c\'oumol] | : * : *! |

For the \( S \)-form, \( VS - \mu_s \text{-Ident}[\text{high}] \) must be top-ranked, but \( VS - \mu_w \text{-Ident}[\text{high}] \) must be subordinated to the requirements of *Diphthong, since in the surface form, the vowel is monophthongal.

(80)
The double opacity of \textit{c’oumal} thus follows from the high ranking of two VS-Faith constraints. VS-\(\mu\)-\textsc{ident}[high] demands that the surface form preserve values for [high] in strong moras (this generates the correct root vowel quality), and VS-\textsc{ident}[\(\mu\)] requires the suffix vowel of the surface form to preserve the value for [round] in the suffix of the virtual form.

Let us now turn to the more complex case of \textit{Tothun}. Assuming the correct \(\Omega\)-form for \textit{Tothun} is \{\textit{Tothun}\}, then it must be the case that \(\Omega\)-\textsc{max}\(-\mu\) dominates \[*\mu\mu\mu\], since closed syllable shortening does not occur in the \(\Omega\)-form. (81) and (82) illustrate the optimization of the \(\Omega\)-pair \{\textit{Tothun}\},\{\textit{Tothun}\}).
### 3.2. Multiple Opacity

For the surface form, however, *[\mu\mu\mu]* must dominate VS-Max-\(\mu\), since closed syllable shortening occurs on the surface.

(82)
This critique of Sympathy Theory has focused on the version which uses Cumulativity as the mechanism responsible for relating sympathetic candidates to actual outputs. We have to ask whether Sympathy Theory can be salvaged by reinstating the intercandidate Faithfulness constraint as the appropriate way to map the sympathetic candidate to the surface form. Given two \( \Phi \)-candidates, \( \Phi_1 \) Putun and \( \Phi_2 \) Potun, we see that the actual output Potun inherits its opaque suffixal vowel rounding from \( \Phi_1 \), but its opaque root vowel quality from \( \Phi_2 \). Any solution based on intercandidate correspondence would thus have to recognize as many dimensions of \( \Phi \)-O-Faithfulness as there are sympathetic candidates. However, this still doesn't address the problem dealt with in the following section, that Sympathy Theory cannot deal with rule-sandwiching.

### 3.3 Rule Sandwiching

A key claim of Sympathy Theory is that the selector constraint acts as if it were undominated for the purpose of selecting the sympathetic candidate. McCarthy states:

> Since selection is based on obedience to a faithfulness constraint, two processes that produce identical faithfulness violations are indistinguishable to the selector. [. . .] If A and B violate exactly the same faithfulness constraints, then they must act together in rendering a third process opaque.

\[(McCarthy\ 1999)\]

The kind of rule interaction that is specifically ruled out by this logic is rule-sandwiching. Consider a derivation consisting of three rules, \( P \), \( Q \), and \( R \), such that \( R \) opacifies \( Q \), but \( P \) does not opacify \( Q \). Imagine further that \( P \) and \( R \) correspond to violations of the same faithfulness constraint \( F \). Precisely this situation is impossible to model using Sympathy Theory, since if both \( P \) and \( R \) entail violations of \( F \), then both must opacify \( Q \). Nevertheless, cases exactly like this have been uncovered in the literature. We will examine three of them here. In §3.3.1 we will address a process of pre-glottal stop shortening in Yawelmani (Kisseberth 1973; Kenstonowicz and Kisseberth 1979), which is also discussed (and whose authenticity is rejected) by McCarthy. In §3.3.2, we will look at glottal deletion in Modern Hebrew (Levi 2000). §3.3.3 examines the interaction of epenthesis and stress assignment in Mohawk. We return to the topic of rule-sandwiching throughout Chapter 7 and again in §11.4.2 and §11.4.3, when we look at two rule-sandwiching processes in West Finnmark Saami.

### 3.3.1 Pre-glottal Vowel Shortening in Yawelmani

In addition to the process of closed syllable shortening discussed above in §3.2.2, Yawelmani also has a more restricted instantiation of the same process which is only triggered by a word-final \( ? \). Pre-glottal Shortening, however, is ordered before Lowering in order to account for mappings like /\( \text{?ll}:+?/ \to \text{?ll}? will fan/.

\[(83)\]

\[
\begin{align*}
\text{\( \text{pana}:+?/ \)} & \text{ pana?} & \text{'will arrive'} \\
\text{\( \text{?ll}:+?/ \)} & \text{?ll?} & \text{'will fan'} \\
\text{\( \text{hoyo}:+?/ \)} & \text{hoyo?} & \text{'will name'} \\
\text{\( \text{c'uyu}:+?/ \)} & \text{c'uyu?} & \text{'will urinate'}
\end{align*}
\]
3.3. RULE SANDWICHING

\begin{center}
\begin{tabular}{ll}
Input & /\text{iili}+?/ \\
Pre-glottal Shortening & ?ili? \\
Lowering & n/a \\
Output & [?ili?] \\
\end{tabular}
\end{center}

Since Max-µ is a selector, the sympathetic candidate will be the most harmonic candidate which preserves vowel weight. However, the sympathetically long vowel will be compelled to undergo Lowering, giving \text{?ile}? as the sympathetic candidate. The optimal candidate will then have to accumulate the violation of Ident[high], incorrectly predicting *?ile? as the actual output.

McCarth y acknowledges the challenge:

\begin{quote}
Obviously, there is no way to use a faithfulness constraint as a selector to give the same fineness of control over opacity that Kisseberth obtains by formulating and ordering two distinct closed-syllable shortening rules. Moreover, there’s no obvious modification of sympathy or faithfulness theory that would change this conclusion. Yavelmanwi, then challenges this prediction of sympathy theory: processes that produce identical violations should act together in rendering a third process opaque.
\end{quote}

\begin{flushright}
\text{(McCarthy 1999: 345)}
\end{flushright}

Nevertheless, McCarthy goes on to argue that the challenge is only apparent, and that Kisseberth’s analysis is actually “fundamentally dubious” for two reasons: McCarthy observes that the earlier rule is motivated by alternations involving just two suffixes, the future and the absolutive, which suggests, according to McCarthy, that we are dealing with something more of the nature of allomorphy rather than “true” phonology. He also sees the success of Kisseberth’s rule-based account as following from a highly arbitrary stipulation: both rules have identical structural changes and overlapping structural descriptions, and so miss the generalization that the two rules operate in the same closed-syllable environment.

Both of these claims may be challenged. As to the allomorphy point, it must be recognized that the alternation is productive. It is in the interests of generative phonology to reduce as much morphology as possible to regular phonology, and we should avoid prejudging where the boundary between the two goes.

As to the point that Kisseberth’s analysis misses an important generalization, it is not the case that the environments are identical. The preglottal environment is the more specific of the shortening environments (and, in the rule-based theory applies first, as required by the Elsewhere Principle), because in addition to the syllable having to be closed, it has to be word-final. Preglottal shortening is motivated quite independently of the closed versus open status of the syllable for many languages. Given this, there is no a priori reason why the more specific rule (Pre-glottal Shortening) and the more general rule (Closed Syllable Shortening) should act together.

Virtual Phonology has no problem dealing with the observed pattern, because the VV-Max-µ constraints are violable. VV-Max-µ is dominated by \text{*VV?}_{wd}.

\begin{flushright}
\text{(85) *VV?}_{wd}
\end{flushright}

A long vowel preceding a word-final glottal stop is disallowed.
(86) and (87) generate the Σp-S-pair (\[?ili?\],\[?ili?\]), illustrating how the opacity which manifests itself in other contexts may be defeased.

(86)

\[
\begin{array}{|c|c|c|c|}
\hline
\text{\[?ili:z\]} & \text{\Long Dash} & \text{\Lowest Dash} & \text{\Dependent[high]} \\
\hline
\text{a.} & \[?ile?\] & : & : & * & : & *! \\
\text{b.} & \text{\Long Dash} & \[?ili?\] & : & * & : & \\
\text{c.} & \[?ile:z\] & : & *! & : & * & \\
\text{d.} & \[?ili:z\] & *! & : & : & * & \\
\text{e.} & \[?ilei?\] & : & *! & : & * & \\
\hline
\end{array}
\]

(87)

\[
\begin{array}{|c|c|c|c|}
\hline
\text{\[?ili:z\]} & \text{\Long Dash} & \text{\Lowest Dash} & \text{\Dependent[high]} \\
\hline
\text{a.} & \[?ili?\] \sim \[?ile?\] & : & *! & : \\
\text{b.} & \text{\Long Dash} & \[?ili?\] \sim \[?ili?\] & : & : \\
\text{c.} & \[?ili?\] \sim \[?ile:z\] & : & *! & : \\
\text{d.} & \[?ili?\] \sim \[?ili:z\] & : & *! & : \\
\text{e.} & \[?ili?\] \sim \[?ilei?\] & : & *! & : \\
\hline
\end{array}
\]

3.3.2 Rule Sandwiching in Modern Hebrew

Like Yawelmani, Modern Hebrew evinces a rule-sandwiching effect which cannot be modeled by Sympathy Theory.

The Irpāʾel verb pattern displays an alternation between e and a. Underlying /e/ is lowered to a preceding a consonant cluster (Kenstowicz and Kisseberth 1979, Levi 2000) due to the activity of a constraint *eCC.

(88) *eCC

Do not allow [e] before a cluster of two consonants.
3.3. RULE SANDWICHING

The alternation is shown in (89). The data is drawn from Kenstowicz and Kisseberth (1979: 134). (89) illustrates the application of the Lowering rule ($e \rightarrow \epsilon / \text{CC}$).

(89)  

<table>
<thead>
<tr>
<th>1sg</th>
<th>3sg.m</th>
<th>3sg.f</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>itparasti</td>
<td>itparåô</td>
<td>itparåsu</td>
<td>‘earn’</td>
</tr>
<tr>
<td>itparångati</td>
<td>itparåsem</td>
<td>itparåsemu</td>
<td>‘become famous’</td>
</tr>
<tr>
<td>idbalbålit</td>
<td>idbalbådel</td>
<td>idbalbåelu</td>
<td>‘be confused’</td>
</tr>
</tbody>
</table>

/e/ is syncopated between two light syllables. This is exemplified in (90).

(90)  

<table>
<thead>
<tr>
<th>1sg</th>
<th>3sg.m</th>
<th>3sg.f</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ithomåkatì</td>
<td>ithomåkåk</td>
<td>ithomåku</td>
<td>‘turn away’</td>
</tr>
<tr>
<td>idkåbåditå</td>
<td>idkåbådåś</td>
<td>itkåpåśu</td>
<td>‘get dressed’</td>
</tr>
<tr>
<td>idbådåritå</td>
<td>idbådåder</td>
<td>idbådåru</td>
<td>‘make fun’</td>
</tr>
</tbody>
</table>

E-Lowering interacts with two consonant deletion rules. A rule of $\gamma$-Deletion ($\gamma \rightarrow \emptyset / \text{CC}$) is ordered before $e$-Lowering, bleeding it of its conditions of application. A rule of $\gamma$-Deletion is ordered after $e$-Lowering, thereby counterbleeding the $e$-Lowering rule. The relevant data is shown in (91).

(91)  

<table>
<thead>
<tr>
<th>1sg</th>
<th>3sg.m</th>
<th>3sg.f</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>itmaletåti</td>
<td>itmaletåle</td>
<td>itmaletåfu</td>
<td>‘become full’</td>
</tr>
<tr>
<td>itnåsetåti</td>
<td>itnåsetåse</td>
<td>itnåsetåsu</td>
<td>‘feel superior’</td>
</tr>
<tr>
<td>itpåtalåtåti</td>
<td>itpåtalåteåhå</td>
<td>itpåtalåthå</td>
<td>‘develop’</td>
</tr>
<tr>
<td>idgålaåhtåti</td>
<td>idgålaåhåthå</td>
<td>idgålaåhåhu</td>
<td>‘shave’</td>
</tr>
<tr>
<td>istågaåti</td>
<td>istågaåeqå</td>
<td>istågaåqåyu</td>
<td>‘become mad’</td>
</tr>
<tr>
<td>itparåtåritåti</td>
<td>itparåtåreå</td>
<td>itparåtåru</td>
<td>‘cause disorder’</td>
</tr>
</tbody>
</table>

As can be seen from (91), $e$-Lowering overapplies in forms like istågaåti and itparåtåti as if the underlying / útil/ was there. However, we do not find the same kind of overapplication in those forms whose third radical is a glottal stop. Thus, /mlP/ is mapped to itmaletåti in the 1sg, not *itmaletåti as expected.

Rule-based serialism has no problem dealing with the observed pattern. Consider the following derivations.

(92)  

<table>
<thead>
<tr>
<th>Input</th>
<th>/itpåleʔ+ti/</th>
<th>/itpåreʔ+ti/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal Deletion</td>
<td>itpåletåti</td>
<td>n/a</td>
</tr>
<tr>
<td>$e$-Lowering</td>
<td>n/a</td>
<td>itparåtåti</td>
</tr>
<tr>
<td>$\gamma$-Deletion</td>
<td>n/a</td>
<td>itparåtåti</td>
</tr>
</tbody>
</table>

As can be seen from (91), $e$-Lowering overapplies in forms like istågaåti and itparåtåti as if the underlying / útil/ was there. However, we do not find the same kind of overapplication in those forms whose third radical is a glottal stop. Thus, /mlP/ is mapped to itmaletåti in the 1sg, not *itmaletåti as expected.

Rule-based serialism has no problem dealing with the observed pattern. Consider the following derivations.

(92)  

<table>
<thead>
<tr>
<th>Input</th>
<th>/itpåleʔ+ti/</th>
<th>/itpåreʔ+ti/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal Deletion</td>
<td>itpåletåti</td>
<td>n/a</td>
</tr>
<tr>
<td>$e$-Lowering</td>
<td>n/a</td>
<td>itparåtåti</td>
</tr>
<tr>
<td>$\gamma$-Deletion</td>
<td>n/a</td>
<td>itparåtåti</td>
</tr>
</tbody>
</table>

The $e$-Lowering rule is ‘sandwiched’ inbetween the two deletion rules in such a way that the first deletion rule bleeds $e$-Lowering, while the second counterbleeds it. This situation is precluded from occurring in Sympathy. According

---

6 According to Levi (2000), the data represents the Mizrahi or Eastern dialect.
to Sympathy Theory, since both ?-Deletion and $\gamma$-Deletion violate exactly the same faithfulness constraint ($\text{Max-C}$), both processes must render $\epsilon$-Lowering opaque. Nevertheless, $\epsilon$-Lowering is only opaque with respect to the deletion of the voiced pharyngeal fricative, not the deletion of the glottal stop, disproving McCarthy’s claim. The problem, again, is the categorical nature of selectorship. A faithfulness constraint is either a selector or not. In the example at hand, we run up against conflicting evidence regarding the status of $\text{Max-C}$ as a selector. The behaviour of the voiced pharyngeal fricative indicates that $\text{Max-C}$ is a selector, but the behaviour of the glottal stop indicates that it cannot be, since a deleted glottal stop fails to trigger the overapplication of Lowering. Setting $\text{Max-C}$ as the selector would therefore predict, counterfactually, that $\epsilon$-Lowering should apply to forms such as *itmaeti as well (giving *itmakiti).

Let us briefly review the beginnings of the OT analysis. The other relevant constraints are shown below.

\begin{align*}
(93) \quad & *?_\sigma [\sigma] \\
& \text{Do not allow ? in coda.} \\
(94) \quad & *\gamma_\sigma [\sigma] \\
& \text{Do not allow \( \gamma \) in coda.} \\
(95) \quad & \text{IO-Max-C} \\
& \text{A consonant in the input must have a correspondent in the output.} \\
(96) \quad & \text{IO-Ident}[\text{low}] \\
& \text{If segment } \zeta \text{ is } [\text{low}] \text{ in the input then the output correspondent of } \zeta \text{ must be } [\text{low}].
\end{align*}

On the surface evidence, both $*?_\sigma$ and $*\gamma_\sigma$ must dominate IO-Max-C, because neither glottal stop nor the pharyngealized fricative ever occur in coda position on the surface. $\text{*CC}$ must dominate IO-Ident[low], since input $[-\text{low}]$ /e/ maps to $[+\text{low}]$ a on the surface. Sympathy Theory deals with the overapplication of Lowering in forms such as *itparati by positing Max-C as the selector. This generates a sympathetic candidate *itparati, which is the most harmonic candidate which satisfies $\text{*Max-C}$. Overapplication is enforced through preserving the value for [low] in the sympathetic candidate.

\begin{align*}
(97) & /\text{itpare}[^{\text{+ti}}] / \\
\begin{array}{|c|c|c|}
\hline
& *?_\sigma & *\text{CC} & \text{IO-Max-C} \\
\hline
a. & \text{itpare}[^{\text{ti}}] & *! : * : * \\
b. & \text{itpara}[^{\text{ti}}] & *! : * \\
c. & \text{itpareti} & : : *! \\
d. & \text{itparati} & : : *! \\
\hline
\end{array}
\end{align*}
While this works fine for inputs with an underlying voiced pharyngeal fricative, running inputs with a glottal stop through the same grammar fails to converge on the correct output form.

Max-C inappropriately selects (98b) as the sympathetic candidate. The consequence is that \(e\)-Lowering overapplies in the surface form, due to the high rank of Ident[low].

Virtual Phonology does not encounter these problems, since the \(\mathfrak{B}\)-Faith constraints are defeasible. The solution lies in seeing \(\mathfrak{B}\)-Max-C as dominated by \(\mathfrak{B}\)-Ident. Ranking \(\mathfrak{B}\)-Max-C below \(\mathfrak{B}\)-Ident[low] but above \(\mathfrak{B}\)-Ident will generate a virtual form in which the glottal stops are deleted in coda position but the pharyngeal fricatives are opaquely retained. If the glottal stops are deleted in the virtual form, they will of course be unable to condition \(e\)-Lowering.

(99) and (100) cover the case of transparent non-application of \(e\)-Lowering bled by Glottal Deletion, giving the output pair \([\text{itpaleti, itpaleti}]\).
CHAPTER 3. VIRTUAL PHONOLOGY

<table>
<thead>
<tr>
<th>/itpAle+ti/</th>
<th>VS-Ident[lo]</th>
<th>VS-Max-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [itpAle\text{t}i]~itpAle\text{t}i</td>
<td>*! : *</td>
<td></td>
</tr>
<tr>
<td>b. [itpAle\text{t}i]~itpAle\text{t}i</td>
<td>*! : *</td>
<td></td>
</tr>
<tr>
<td>c. [itpAle\text{t}i]~itpAle\text{t}i</td>
<td>*! : *</td>
<td></td>
</tr>
<tr>
<td>d. [itpAle\text{t}i]~itpAle\text{t}i</td>
<td>*! : *</td>
<td></td>
</tr>
</tbody>
</table>

(101) and (102) cover the case of opaque overapplication of e-Lowering countered by l-Deletion, giving the output pair ([itpara\text{t}i],[itpara\text{t}i]).

<table>
<thead>
<tr>
<th>/itpAre+ti/</th>
<th>*eCC</th>
<th>VS-Max-C</th>
<th>*\text{`a}</th>
<th>VS-Ident[lo]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [itpara\text{t}i]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. [itpara\text{t}i]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. [itpara\text{t}i]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. [itpara\text{t}i]</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Highly ranked VS-Ident[lo] ensures the overapplication of Lowering. *\text{\`a}* must outrank VS-Max-C (and IS-Max-C) to ensure that the pharyngeal glide doesn’t surface.

<table>
<thead>
<tr>
<th>/itpara\text{t}i+ti/</th>
<th>VS-Ident[lo]</th>
<th>VS-Max-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [itpara\text{t}i]~itpara\text{t}i</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b. [itpara\text{t}i]~itpara\text{t}i</td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>c. [itpara\text{t}i]~itpara\text{t}i</td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>d. [itpara\text{t}i]~itpara\text{t}i</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.3.3 Contextual Metrical Invisibility in Mohawk

In Mohawk, a Northern Iroquoian language spoken in New York State, Ontario, and Québec, epenthetic vowels are variably active metrically. The metrical visibility of epenthesis is a function of the phonotactic conditions which trigger the epenthesis.

In Mohawk, stress falls on the penultimate syllable, irrespective of weight. The stressed syllable must be heavy (either CVC or CV\text{\`a}). The basic pattern for
stress assignment is exemplified in (103). Examples are taken from Hagstrom (1997: 114).

(103) a. /wak-haratat-u-hatye/ wakharatatuhatye 'I go along lifting up'
b. /hra-kw-as/ raÌÊkwas 'he picks it'
c. /k-atirut-ha?/ katirutha? 'I pull it'
d. /k-ohar-ha?/ koharha? 'I attach it'
e. /k-atatkerahkw-ha?/ kataÌÊkerákwa? 'I float'
f. /k-owat-s/ kòÌÊkwats 'I dig'
g. /te-k-yak-s/ tekyak's 'I break it in two'
h. /wak-ashet-u/ wakashêtu 'I have counted it'
i. /k-akaÌÊrokew-as/ kakaÌÊroké:was 'I am dusting'
j. /k-hyatut-s/ khyatutus 'I write'
k. /k-haratat-s/ kharatats 'I am lifting it up a little (with a lever)'

Mohawk words are subject to a minimal bisyllabic requirement, which is enforced by adding a prothetic \( i \).

(104) a. /k-ya-s/ ëkya 'I put it'
b. /k-tat-s/ ëktats 'I offer it'
c. /k-ek-s/ ëkëks 'I eat'
d. /k-ya'k-s/ ëkya'ks 'I cut it'

In addition to prothetic \( i \), Mohawk also has an epenthetic \( e \) which is used to break up clusters of consonants. E-Epenthesis applies in three contexts:

- to break up a sequence of a consonant followed by a sonorant /n r w/,
- between a consonant and a word-final glottal stop,
- after a consonant followed by a consonant cluster: /CCC/ \( \rightarrow \) [CeCC], but /\( H \) CC/ \( \rightarrow \) [\( H \) CeC].

The effect of epenthesis is stress shift to the left. Epenthesizing once shifts the stress one position to the left of its canonical position. In (105), the epenthetic vowel surfaces in the penultimate position, the canonical position for stress. In these cases, stress shifts to the ante-penultimate syllable. (105e) shows that epenthesizing twice results in a second stress shift: stress surfaces in the pre-ante-penultimate syllable.

(105) a. /\( ë-\)k-r-\( ë-\)-\( ë-\)/ ëkerë 'I will put it into a container'
b. /te-\( ë-\)rk-s/ tekeriks 'I put them together'
c. /\( ë-\)kaÌÊahsutr-\( ë-\)/ takâhsûtàrä 'I will splice it'
d. /w-akra-s/ wàkeras 'it smells'
e. /wa'\( ë-\)-t-k-atat-nak-\( ë-\)/ wa'kataténakë 'I scratched myself'

(106) a. /ʌ-k-arat-ʔ/ ákäráteʔ  'I lay myself down'
b. /ro-kut-ot-ʔ/ rókútoteʔ  'he has a bump on his nose'
c. /t-ʌ-k-rík-ʔ/ tákéríkeʔ  'I'll put together side by side'
d. /o-nrah-ʔ/ ónéráhíteʔ  'leaf'
e. /t-ʌ-k-ähkw-ʔ/ tákéhkweʔ  'I'll lift it'

In (107), epenthesis into the CCC environment has no effect on stress placement. Stress surfaces in its canonical position regardless.

(107) a. /wá-k-nyak-s/ wákenyaks  'I get married'
b. /s-rhos/ serhos  'you coat it with something'
c. /te-k-ahsut-r-ʔ/ tek ahsutr-ʔ  'I splice it'
d. /s-k-ahkt-s/ skákhts  'I got back'
e. /sa-s-ahkt/ sásáhkts  'go back'

The data exemplify the familiar rule-sandwiching phenomenon. Stress Assignment is sandwiched between two distinct epenthesis rules: (a) a rule of epenthesis into a triconsonantal sequence, $∅ \rightarrow e/CC$ (Triconsonantal Epenthesis); $∅ \rightarrow e/\{n\ w\} CC$ and (b) a rule of epenthesis into a biconsonantal sequence of a consonant followed by a sonorant, $∅ \rightarrow e/CC\{n\ w\}$ (Biconsonantal Epenthesis). The latter rule counterbleeds normal stress assignment in the sense that stress does not appear on the penultimate syllable on the surface. (108) shows how this difference is derived in terms of an ordered sequence of rules.

(108)

<table>
<thead>
<tr>
<th>Input</th>
<th>/ʌ-k-r-ʌ-ʔ/</th>
<th>/wá-k-nyak-s/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triconsonantal Epenthesis</td>
<td>n/a</td>
<td>wákenyaks</td>
</tr>
<tr>
<td>Stress</td>
<td>ákéráʔ</td>
<td>wákenyaks</td>
</tr>
<tr>
<td>Biconsonantal Epenthesis</td>
<td>n/a</td>
<td>wákenyaks</td>
</tr>
<tr>
<td>Output</td>
<td>ákéráʔ</td>
<td>wákenyaks</td>
</tr>
</tbody>
</table>

Mohawk exhibits one other counterbleeding relationship. Prothesis takes place even though it is subsequently counterbled by epenthesis.

(109) a. /hs-ri-ht-∅/ śéerih  'cook'
b. /t-n-ehr-ʔ/ iténehreʔ  'you and I want'
c. /t-wa-ehr-ʔ/ iténehreʔ  'you and I want to'

(110)

<table>
<thead>
<tr>
<th>Input</th>
<th>/t-n-ehr-ʔ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prothesis</td>
<td>iténehreʔ</td>
</tr>
<tr>
<td>Stress</td>
<td>iténehreʔ</td>
</tr>
<tr>
<td>Epenthesis</td>
<td>iténehreʔ</td>
</tr>
<tr>
<td>Output</td>
<td>[iténehreʔ]</td>
</tr>
</tbody>
</table>
3.3. RULE SANDWICHING

Importantly, where epenthetic e is visible to stress placement, it is visible to
the minimal bisyllabicit y requirement as well. Consequently, words which
are underlyingly subminimal, but which apply Triconsonantal Epenthesis do
not evince Prothesis. Thus, /s-rho-s/ 'you coat it with something' undergoes
Triconsonantal Epenthesis to serhos. Here, epenthetic e is metrically active,
and prothesis would be gratuitous. Accordingly, *serhos is ungrammatical.

Let us now turn to the OT account. Epenthesis is invoked to avoid two
distinct configurat ions corresponding to two distinct sequential markedness con-
straints.

(111) *CSon
   The sequence C−Sonorant is disallowed.
(112) *CCC
   The sequence C−C−C is disallowed.

For underlying triconsonantal clusters C1C2C3, epenthesis will usually apply
between C1 and C2, e.g. /wak-nyak-s/ → wakñyaks. Clusters of hC or sC
cannot be broken up in this way: /sa-s-ahkt/ → sasāhkt.

Epenthesis into a triconsonantal cluster is enforced in virtual form. Epenthesis
into a cluster of C+Son, on the other hand, is a purely surface phenomenon
and is not represented virtually. This difference is captured by the ranking
of ƗB-Dep-V relative to (111) and (112). ƗB-Dep-V must be dominated by
*CCC, but must itself outrank *CSon in order to prevent virtual epenthesis in
the C−Son context. Consider the treatment of /æ-k-r-æ/ 'I will put it into a
container' (=105a). (113) and (114) deal with the optimization of the output
pair ⟨|ækræ|, [ækéra]|⟩.

In the case of the opaque surface form, faithfulness to the stress pattern
of the virtual form is achieved by ranking ƗB-Ident[Stress] above Stress.
Epenthesis is enforced at the surface by the ranking *CSon≫ƗB-Dep-V.

(113)

<table>
<thead>
<tr>
<th>/æ-k-r-æ/</th>
<th>*CCC</th>
<th>ƗB-Dep-V</th>
<th>*CSon</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ː*</td>
<td>[ækra]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>[ækera]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>[ækera]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the case of the opaque surface form, faithfulness to the stress pattern
of the virtual form is achieved by ranking ƗB-Ident[Stress] above Stress.
Epenthesis is enforced at the surface by the ranking *CSon≫ƗB-Dep-V.

(114)

<table>
<thead>
<tr>
<th>/æ-k-r-æ/</th>
<th>*CCC</th>
<th>ƗB-Dep-V</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ː*</td>
<td>[ækra] ~ [ækra]</td>
<td>:</td>
<td>*!</td>
</tr>
<tr>
<td>b. ː*</td>
<td>[ækra] ~ [ækera]</td>
<td>:</td>
<td></td>
</tr>
<tr>
<td>c. ː*</td>
<td>[ækra] ~ [ækera]</td>
<td>:</td>
<td>*!</td>
</tr>
</tbody>
</table>
CHAPTER 3. VIRTUAL PHONOLOGY

Compare this with *wákenyaks. The metrical visibility of the epenthetic vowel in this case shows that *CCC must dominate \( \text{V}_2\)-Dep-V. (115) and (116) deal with the optimization of the output pair \{wákenyaks, wákenyaks\}.

\[(115)
\]

<table>
<thead>
<tr>
<th></th>
<th>/wak-nyak-s/</th>
<th>*CCC</th>
<th>( \text{V}_2)-Dep-V</th>
<th>*CSon</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>wákenyaks</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>wákenyaks</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>wákenyaks</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

\[(116)
\]

The difference in behaviour cannot be modeled using Sympathy Theory. Using \( *\text{Dep-V} \) as selector would select the non-epenthesized candidate in both cases, i.e. \( *\text{wákenyaks} \), and, erroneously, \( *\text{wákenyaks} \). The selection of \( *\text{wákenyaks} \) as sympathetic candidate is infelicitous, since it predicts, counterfactually, that stress should surface on the antepenultimate syllable, giving *wákenyaks.

Before leaving Mohawk, let us consolidate the result by deriving the overapplication of Biconsonantal Epenthesis ‘after’ Prothesis, as exemplified in (109). WdMin is undominated. (117), (118), and (119) deal with the optimization of the output pair \{srih\t, srih\t\}.

\[(117)
\]

As it stands, (117) generates the wrong output form, since it predicts that, of Prothesis and Biconsonantal Epenthesis, the latter will be least costly. This is because Biconsonantal Epenthesis can be invoked to satisfy both WdMin and
*CS on in one fell swoop. This is not what we want. In fact, we may rank 33-Contig in place of 33-Dep-V, and achieve the desired result, without any loss of empirical coverage.

(118)

<table>
<thead>
<tr>
<th>/hs-ri-ht-∅/</th>
<th>*CS → DM → VS Contig</th>
<th>*CS → DM → VS Coress</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [sũriht]</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>b. [sũriht]</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>c. [sũriht]</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>d. [ũsũriht]</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>e. [ũsũriht]</td>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

(119)

<table>
<thead>
<tr>
<th>/hs-ri-ht-∅/</th>
<th>*CS → DM → VS Contig</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [ũsũriht] ~ũsũriht</td>
<td>:</td>
</tr>
<tr>
<td>b. [ũsũriht] ~ũsũriht</td>
<td>:</td>
</tr>
<tr>
<td>c. [ũsũriht] ~ũsũriht</td>
<td>:</td>
</tr>
<tr>
<td>d. [ũsũriht] ~ũsũriht</td>
<td>:</td>
</tr>
<tr>
<td>e. [ũsũriht] ~ũsũriht</td>
<td>:</td>
</tr>
</tbody>
</table>

Mohawk is not alone in the non-uniformity of its behaviour with respect to stress-epenthesis interaction. Alderete (1999a) cites Spanish as identical in form with the Mohawk case. In Spanish, epenthesis into initial sC clusters is ignored for the purposes of stress. Epenthesis to break up a triconsonantal cluster, on the other hand, is metrically active. Similarly, Kiparsky (1998: 76f.) discusses an exactly parallel case from Palestinian Arabic under the rubric of a lexical vs. postlexical distinction between two epenthesis rules.

3.3.4 Discussion

Rule sandwiching does not appear to be a common phenomenon. The reasons for its low statistical frequency are not hard to understand from a learning perspective. Since opaque generalizations are transparent only at the level of
the abstract virtual form, their learnability is greatly enhanced if they are exceptionless at that level. Anything less, and the cues for them will tend to be attenuated to the point where they can no longer be reliably learnt by a new generation of speakers. Learning will favour opaque generalizations which are exceptionless, i.e. grammars in which the relevant P\(\Box\)-FAITHFULNESS constraints are undominated. This is precisely the situation modeled by Sympathy Theory, but as we have seen, Sympathy Theory is not refined enough to capture the range of attested opaque patterns. The Yawelmani, Modern Hebrew and Mohawk examples drive this point home.

3.4 Virtual Markedness

A number of cases in the literature have been observed where language-specific restrictions on the input apparently have to be invoked. This is incompatible with the central OT tenet of the Richness of the Base. Here we elaborate the claim that the virtual form is the appropriate locus in which to encode apparent language-specific restrictions on the input in terms of virtual markedness constraints whose activity is relativized to the virtual form.

3.4.1 Nasalized Vowels in American English

Speakers of American English normally pronounce *say*, *sate*, *sane*, and *saint* as [\(\text{sei}\)], [\(\text{sei}\ t\)], [\(\text{s}\ e\ n\)], and [\(\text{s}\ e\ n\ t\)] respectively. Vowel nasalization is always the result of being followed by a nasal. However, in the last example, nasalization is opaque: the input nasal which conditions nasalization on the preceding vowel has itself been deleted due to the following voiceless obstruent (cf. [\(\text{sei}\ n\ d\)], the pronunciation of *seined*). In rule-based terms, the rule for Nasalization (V→V/N) must precede the rule for Nasal Deletion (N→∅[\(\text{−}\text{son}\ −\text{voice}\ −\text{cont}\ )]), as shown in (120).

\begin{align*}
\text{Input} & /\text{sei}\ t\ n/ /\text{sei}\ n/ \\
\text{Nasalization} & \text{\(\text{sei}\ t\)} \quad \text{\(\text{sei}\ n\)} \\
\text{Nasal Deletion} & \text{\(\text{sei}\ t\)} \quad \text{\(\text{n/}\)} \\
\text{Output} & [\text{\(\text{sei}\ t\)}] [\text{\(\text{sei}\ n\)}] \\
\end{align*}

The problem is that there is no possible item [\(\text{sei}\)]. Nasalized vowels are obviously possible on the surface in the absence of a following nasal, as grammatical forms such as [\(\text{sei}\)] attest. However, under Richness of the Base, we have to allow for the possibility of nasalized vowels in the input. Given the well-formedness of [\(\text{sei}\)], there is no way of generating the correct result that [\(\text{sei}\)] is ill-formed.\(^8\) Rule-based theory dealt with cases of this sort precisely by invoking language-specific restrictions on the input, in the form of MSC's. In rule-based

\(^8\)One logically possible way using output-oriented constraints is to use a constraint prohibiting nasalized vowels in any syllable except one closed with a final voiceless obstruent. Such a constraint is patently absurd, and does violence to the requirement that OT constraints be universal and functionally grounded.
3.4. **VIRTUAL MARKEDNESS**

terms, there is an MSC which holds of underlying representations in English: underlyingly, vowels must be [−nasal]. All surface nasal vowels in English are thus systematically unfaithful, corresponding to underlying sequences of V+N. How do we deal with the gap in OT? At the level of the virtual form in English, vowel nasalization is only ‘licensed’ by the presence of a following nasal.

(121) **Lic-NasV**

Nasalized vowels must be licensed by a following nasal.

(121) interacts with another markedness constraint, *NT, which militates against a nasal preceding a voiceless stop.

(122) ***NT**

The sequence of nasal followed by a voiceless stop is disallowed.

Given an input /sɛɨt/, the domination of *NT by [Lic-NasV] ensures that the optimal virtual form is [sɛɨn], with unfaithful epenthesis of a homorganic nasal. On the surface, however, Lic-NasV must be dominated by the constraint demanding faithfulness to virtual nasalization. The surface form [sɛɨt] displays the overapplication of nasalization in the absence of a nasal triggering environment. Thus, the inputs /sɛɨt/, /sɛɨnt/, /sɛɨnt/ are neutralized to the same output pair ([sɛɨnt], [sɛɨnt]). Only candidates which obey *Voral and I-Ident[nas] are shown.

(123)

| /sɛɨnt/, /sɛɨnt/, /sɛɨnt| | Lic-NasV | *NT | Lic-NasV |
|--------------------------|----------|------|---------|
| a. ✕ | sɛɨnt ~ [sɛɨt] | ✗! | ✗! |
| b. | sɛɨt ~ [sɛɨt] | ✗! | ✗! |
| c. | sɛɨnt ~ [sɛɨnt] | ✗! | ✗! |
| d. | sɛɨt ~ [sɛɨnt] | ✗! | ✗! |

When the final voiceless stop is absent, performance on *NT is irrelevant. Inputs /sɛɪn/, /sɛɪn/, and /sɛɪ/ are all neutralized to the output pair ([sɛɪn], [sɛɪn]).

(124)

<table>
<thead>
<tr>
<th>/sɛɪn/, /sɛɪn/, /sɛɪ/</th>
<th></th>
<th>Lic-NasV</th>
<th>*NT</th>
<th>Lic-Nas-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ✕</td>
<td>sɛɪn ~ [sɛɪ]</td>
<td>✗!</td>
<td>✗!</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>sɛɪ ~ [sɛɪ]</td>
<td>✗!</td>
<td>✗!</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>sɛɪn ~ [sɛɪn]</td>
<td>✗!</td>
<td>✗!</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>sɛɪ ~ [sɛɪn]</td>
<td>✗!</td>
<td>✗!</td>
<td></td>
</tr>
</tbody>
</table>

3.4.2 Palato-alveolar Affricates in Abersoch Welsh

In Welsh, Soft Mutation or Lenition applies to spirantize an underlying /b d g/ to [v ʌ ʊ] in certain morphological contexts (Willis 1982). Some dialects add a fourth consonant ʤ to the voiced obstruent series, whose behaviour under Soft Mutation varies dialectally in significant and interesting ways. In the

⁹ /g/ lenites to ʊ via an ‘intermediate’ v.
Llanbrynmair dialect of Welsh, spoken in eastern Mid Wales (Sommerfelt 1925),
the outcome of leniting /dZ/ is Z. Yet, in Abersoch Welsh, /dZ/ lenites to D in
the same environment, e.g. djawl ‘devil’~i djawl ‘his devil’. It is possible to
interpret this as reflecting a restriction on the input specific to Welsh to the
effect that surface dZ must itself be derived from input /dj/, and that /dZ/ is
not a permitted input. This is so, since there is no natural or obvious way to
derive Dj from /dZ/ by spirantization alone. On the other hand, Dj is a natural
and transparent product of the spirantization of /dj/. Here we interpret this
as a restriction on the V-form. In virtual form, inputs /dZ/ and /dj/ must be
neutralized into a single virtual cluster |dj|. (127) shows the optimization of
the mapping /dj/→|dj|. The pattern emerges from the interaction of two basic
markedness constraints, shown in (125) and (126).

(125) *PALAIV
   * [ +coronal ]
      −anterior
      +strident

(126) *Tj
The sequence of a coronal consonant followed by the glide j is disallowed.

On the surface evidence, (126) dominates (125). Yet, judging from the evi-
dence provided by the alternation of dZ with Dj under Soft Mutation, the reverse
must be true of the virtual form. (126) must be dominated by the virtual in-
carnation of (125), |*PALAIV|.

(127)

<table>
<thead>
<tr>
<th>Input</th>
<th>*PALAIV</th>
<th>*Tj</th>
<th>*PALAIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>/djawl/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. djawl~[djawl]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Djawl~[djawl]</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. djawl~[djawl]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. [djawl]~[djawl]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(128) shows that the same virtual output is optimized under the assumption
that the input is /dZ/.

(128)

<table>
<thead>
<tr>
<th>Input</th>
<th>*PALAIV</th>
<th>*Tj</th>
<th>*PALAIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>/djawl/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. djawl~[djawl]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Djawl~[djawl]</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. djawl~[djawl]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. [djawl]~[djawl]</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The outcome of the lenition of djawl, Djawl, is thus more faithful to the
virtual base [djawl] than it is to the surface base [djawl], since Djawl differs
from [djawl] only in terms of Ident[cont]. In relation to [djawl], on the other
hand, Djawl is unfaithful with respect to a host of other features in addition to
its violation of Ident[cont]. The affricate [dZ] is [+strident, −anterior], whereas
the fricative [Z] is [−strident, +anterior]. For present purposes, we may collapse
these violations into a single violation of Ident[Place].
What is interesting about the Welsh case is that it seems to necessitate positing faithfulness constraints on the correspondence relations holding between a virtual base or underlying form [djaw] and a surface sandhi form (which occurs in the Lenition context) i djaw. The latter is a morphological superset of the virtual base form.

(129)

<table>
<thead>
<tr>
<th>/i djaw/</th>
<th>*S-Ident</th>
<th>PalAlv</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [djaw]~i djaw</td>
<td>!</td>
<td>*</td>
</tr>
<tr>
<td>b. [djaw]~i djaw</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. [djaw]~i jaw</td>
<td>!</td>
<td>*</td>
</tr>
<tr>
<td>d. ☞ [djaw]~i djaw</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

3.4.3 Ga-gyō Variation in Tokyo Japanese

Tokyo Japanese has an alternation between a voiced velar stop g and a velar nasal y (Itô and Mester 1997). On initial impressions, the alternation bears all the hallmarks of classical allophony. The two sounds are in complementary distribution, with g restricted to initial position in the PrWd, and y occurring medially.

(130) a. geta *yeta ‘clogs’
go *no ‘game of Go’
b. *kagi kagi ‘key’
*tokage tokage ‘lizard’

The distribution of both allophones is captured by the interaction of the markedness constraints *[y] and *g with the low-ranked Ident[nas].

(131)

<table>
<thead>
<tr>
<th>/kagi/</th>
<th>*[y]</th>
<th>*g</th>
<th>Ident[nas]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. kagi</td>
<td>!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. ☞ kagi</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The grammar is indifferent to the presence of y or g in the input, and the correct output is generated in robust fashion.

(132)

<table>
<thead>
<tr>
<th>/yeta/</th>
<th>*[y]</th>
<th>*g</th>
<th>Ident[nas]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. yeta</td>
<td>!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. ☞ yeta</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Nevertheless, Tokyo Japanese evinces an apparent input restriction to the
effect that whenever \( \gamma \) surfaces, /\( \gamma \)/ is underlying. The evidence for this comes
from the interaction of \( qa-gy\ddot{a} \) allomorphy with a process of Sequential Voicing
(\( \text{Rendaku} \)). \( \text{Rendaku} \) voices the initial segment of the second component of a
compound PrWd. Examples are shown in (133), the sequentially voiced consonant underlined.

(133) tama ‘ball’ \underline{teppoodama} ‘bullet’
    tana ‘shelf’ \underline{garasudama} ‘glass shelf’

The observed pattern reflects the activity of the markedness constraint \( \text{SeqVoi} \) in (134). The informal statement of the constraint is due to Itô and Mester

(134) \( \text{SeqVoi} \):
In \([\text{Wd} X_1 X_2]\), \( X_2 \) begins with a [\(+\text{voi}\)] segment.

(134) must dominate \( \text{Ident}[\text{voi}] \).

\begin{tabular}{|c|c|c|}
\hline
 & /teppoo+tama/ & \text{SeqVoi} & \text{Ident}[\text{voi}] \\
\hline
a. & teppootama & *! & \\
\hline
b. & \underline{teppoodama} & & \\
\hline
\end{tabular}

\( \text{Rendaku} \) is blocked by an incarnation of the OCP known as \( \text{Lyman’s Law} \),
which bans consecutive voiced obstruents in the same PrWd. Consider the
examples in (136).

(136) taba ‘bundle’ satsutaba ‘wad of bills’
    \( *\)satsudaba
    tade ‘knotweed’ harutade ‘redshank’
    \( *\)harudade

Following Itô and Mester (1997), I assume the blocking effect is due to the
domination of \( \text{SeqVoi} \) by the OCP.

\begin{tabular}{|c|c|c|c|}
\hline
 & /satsu+taba/ & \text{OCP} & \text{SeqVoi} & \text{Ident}[\text{voi}] \\
\hline
a. & \underline{satsutaba} & & *! & \\
\hline
b. & satsudaba & & *! & \\
\hline
\end{tabular}

Crucial to note here is that only consecutive voiced \textit{obstruents} are subject to
\( \text{Lyman’s Law} \). Sonorants do not block \( \text{Rendaku} \) in an initial obstruent. Hence
\underline{teppoodama} is fine. The twist comes with the behaviour of \( \gamma \): \( \gamma \) behaves as if it
were a voiced obstruent \( \gamma \), blocking \( \text{Rendaku} \).

(138) toqi ‘sharpen’ hasamitogi ‘knife grinder’
    \( *\)hasamidogi
    toge ‘thorn’ \underline{barato\d{e}} ‘rose thorn’
    \( *\)barad\d{e}
3.4. VIRTUAL MARKEDNESS

Under the standard assumptions of rule-based serialism, this may be taken as evidence that the underlying segment must be /g/. However, under Richness of the Base, we have to consider inputs with /ŋ/. The danger is of course that /ŋ/ in the input will fail to block Rendaku. This predicts a two-ways split in lexical items with medial ɣ: those whose surface ɣ derives from underlying /g/ will block Rendaku, while those deriving from underlying /ŋ/ will not. No such split exists in the Tokyo Japanese lexicon: all instances of surface ɣ pattern like underlying /g/. Essentially, input /ŋ/ has to be forced to pattern like an obstructor by undergoing virtual denasalization to ɣ. We can use the virtual form to model this apparent restriction on the input in a robust and output-oriented way. By letting the V-form-relativized constraint [*_ŋ] dominate [nas]-Ident, we can force the mapping /ŋ/ → /g/.

(139)

<table>
<thead>
<tr>
<th>/bara+toNe/</th>
<th>[*_ŋ]</th>
<th>CCS</th>
<th>[nas]-Ident[vol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [baratoge]</td>
<td>*! :</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. [baradoqe]</td>
<td>*! :</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ☞ [baratoge]</td>
<td>:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. [baradoqe]</td>
<td>: *!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On the surface, virtual [ɣ] is (re)nasalized to ŋ because of the highly ranked markedness constraint [*_ŋ]. [*_ŋ], of course, has no jurisdiction at the surface. The unexpected blocking effect of surface ɣ is then accounted for by highly-ranked [nas]-Ident[vol]. Performance on this constraint is crucial in securing the victory of the opaque candidate (140a) over the transparent candidate (140b).

(140)

<table>
<thead>
<tr>
<th>/bara+toNe/</th>
<th>[nas]-Ident[vol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ☞ [baratoge]~baratoge</td>
<td>∼ baratoge</td>
</tr>
<tr>
<td>b. [baratoge]~baradoqe</td>
<td>*! :  baradoqe</td>
</tr>
<tr>
<td>c. [baratoge]~baratoge</td>
<td>: *!  baratoge</td>
</tr>
<tr>
<td>d. [baratoge]~baradoqe</td>
<td>: *!  baradoqe</td>
</tr>
</tbody>
</table>

There is an alternative analysis of these facts which deploys an Input-Output chain shift. We could, instead of invoking a Duke-of-York mapping of /ŋ/ → /g/ → [ŋ],
neutralize /n/ to some other output, say [n]. This would circumvent the objection that the mapping is non-cumulative. The problem with the IO-chainshift scenario is lack of evidence. In fact, the hypothesis that Tokyo Japanese speakers map input /n/ to [n] is easily testable. The equivalence of a ‘Wug’ test (Berko 1958) using some non-occurring but phonologically possible Japanese word, e.g. kooge, should reveal that this word patterns with toge ‘thorn’ in blocking Rendaku in the compound context.

The present section, as well as sections 3.4.1 and 3.4.2, has dealt with apparent language-specific restrictions on the underlying segment inventory. In Classical OT, the lexical, or underlying representation is completely epiphenomenal, and plays no role in the grammar of the language. The evidence of American English, Abersoch Welsh, and Tokyo Japanese, on the other hand, suggest it is entirely sensible, as well as necessary to incorporate covert restrictions on underlying forms as well. This move doesn’t compromise the Richness of the Base: the reason is that the input and the underlying form are still crucially distinct in OT. While the input is universal, the underlying form is a language specific construct: it is a covert output of the grammar. The difference between the present proposal and the Classical OT line on underlying forms is that, on the approach advocated here, underlying forms are not merely epiphenomenal. There are situations in which the grammar must crucially refer to them.

3.4.4 Disjoint Metrical Tiers in Huariapano

Up until now, we have reviewed three cases where virtual markedness may be used to account for apparent language-specific restrictions on the input. Virtual forms may be subject to restrictions beyond those imposed on surface forms, and then influence the surface form through virtual-surface identity, but not all of these have the character of input restrictions. This is clearest when those restrictions are metrical in nature.

Two cases of virtual metrical restriction are analyzed in this and the subsequent section.

Huariapano, a Panoan language formerly spoken in Peru (Parker 1994, 1998), has a process of Rhythmic Coda Epenthesis. Abstracting away from the intricate details, a coda is assigned to an odd-numbered syllable going from left to right across the word. Nevertheless, Parker observes that there is also an exceptional, but robustly attested pattern in which secondary stress iterates from right to left in many words (about one third of the vocabulary). This lexically exceptional rhythmic pattern of assignment is inconsistent with the pattern rhythmic coda epenthesis.

In Huariapano, main stress is assigned by constructing a moraic trochee at the right edge of the word. By default, secondary stress iterates from the left edge of the word in quantity-insensitive fashion (i.e. syllabic trochees are constructed). (141) illustrates the basic pattern of stress assignment.\(^{10}\) (141a) illustrates the default pattern of penultimate primary stress. (141b) shows that stress shifts to the final syllable when the word ends in a closed (heavy) syllable in quantity-sensitive pattern. (141c) shows the iterative assignment of secondary stress. (141d) shows that the assignment of secondary stress is quantity-sensitive.

\(^{10}\)In Parker's transcription, /θ/ is a retroflexed alveopalatal fricative; /β/ is a voiced bilabial obstruent whose phonetic implementation fluctuates between stop, fricative, and glide; /ɨ/ is a high back unrounded vowel; nasalization is transcribed with the Polish Hook, e.g. [ŋ].
insensitive, since heavy syllables may end up in weak metrical positions. (141e) demonstrates that the default direction of secondary stress assignment is left-to-right. Finally, (141f) illustrates the pattern of right-to-left iteration of secondary stress found in many words.

(141) a. *ka(notí) (kósh)ni* `bow (weapon)`
   b. *lön(tís) ša(šú)ni* `claw; fingernail` `beard`
   c. *(nósš) (rána) (kýo)(kánkì)* `we` `they finished`
   d. *(yómú)(ránò)(škì)* `he is going to hunt`
   e. *(nihay)(šáškì)* `we` `they washed`
   f. *(milömú)(rána) ih(kásčañ)(kátì)* `you (plural)` `you would shake with fear`

Coda epentheses of *h* applies to any syllable in an odd-numbered position, provided that certain conditions are fulfilled. The examples in (142a) illustrate the application of the process in the word-initial syllable, preceding primary stress. In (142b), Coda Epenthesis takes place in the initial, secondary-stressed syllable. (142c) exemplifies the process in the third syllable (with or without secondary stress), and (142d) exemplifies the process in the fifth syllable (again, with or without the presence of secondary stress).

(142) a. *nolpóś NHpá (pášiki) píšši* `snail` `we washed` `sleeping mat`
   b. *hikompána ciškinamáñ* `rattlesnake` `corner`
   c. *yómirahkáno (hayáyikáñki) bójinkáy* `let’s go hunting` `they possessed` `they will take/carry`
   d. *bímakanóshikáy yómiránóshikáy yómíráhkañúkkáy* `they will look for` `they will hunt` `they hunted`

Although rhythmic, coda epenthesis fails completely to correlate with the presence versus absence of stress on the epenthesized syllable. The epenthized syllable may thus be stressed (as in *čiškinamáñ* and *yómíráhkañúkkáy*), or unstressed (as in *nolpóś, hayáyikáñki, and bímakanóshikáy*).

Parker accounts for the mismatch in representational terms, proposing that Huariapano has two distinct metrical planes: a stress tier used purely for the

11 We abstract away from these conditions in our analysis, since they are not strictly germane to the opacity question posed by the Huariapano data. Epenthesis of *h* is blocked preceding a sonorant or consonant cluster. It is also prevented from occurring in the first syllable of the word if it is also main-stressed. For details, the reader is referred to Parker (1994, 1998).
CHAPTER 3. VIRTUAL PHONOLOGY

placement of stress, and a rhythm tier used solely to calculate eligibility for Coda Epenthesis. On the rhythm tier, left-to-right parsing of syllabic trochees is rigidly enforced, and Coda Epenthesis is a response to the requirement that foot heads be bimoraic. The exigencies of parsing on the rhythm tier thus clash with the surface stress pattern in two respects. First, there are those cases where Coda Epenthesis takes place in the absence of any surface foot being constructed over the relevant syllable due to a surface observance of *Clash. (142a) gives examples of this, as does the form bànakânošt#ký in (142d). Second, there are cases where the directionality of Rhythmic Coda Epenthesis is at odds with the directionality of surface stress assignment. This is due essentially to the large number of stems which exceptionally specify right-to-left iteration of secondary stress, of which hayáyi#káqki in (142c) is an example.

Parker’s representational solution is probably best abandoned in favour of a solution in terms of Virtual Phonology. Let’s sketch the outlines of an analysis. In virtual-phonological terms, Huariapano words are subject to an undominated virtual constraint enforcing left-to-right parsing into syllabic trochees. This is more accurately seen in terms of the interaction of ALIGN (Ft.L,PrWd.L) — ‘Every foot in PrWd must be left edge’ — and Parse-σ. In short, Parse-σ dominates the virtual markedness constraint [ALIGN-Ft-L]. This interaction is abbreviated [Stress] in the tableau in (143).

Virtual trochees are further subject to a constraint requiring bimoraicity of the foot head. This mora is implemented by epenthizing an unmarked segment, h. For example, the virtual correspondent of pih(káti) ‘I ate’ must be ![p(hk)a]ti]. The surface form obliterates the virtual stress pattern, but traces of it survive in the form of the preservation of virtually-assigned h.

(143)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/pikati/</td>
<td>Stress</td>
<td>Hdbin-Dep-C</td>
</tr>
<tr>
<td>a.</td>
<td>![p(hk)a]ti]</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>![pika]ti]</td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>![pí(káti)]</td>
<td>*</td>
</tr>
<tr>
<td>d.</td>
<td>![pí(káhtí)]</td>
<td>*</td>
</tr>
<tr>
<td>e.</td>
<td>![pih(káti)]</td>
<td>*</td>
</tr>
</tbody>
</table>

In order to prevent Coda Epenthesis in any surface-stressed syllable (i.e. even positions), IS-Dep-C must outrank Hdbin.

(144)
3.4. VIRTUAL MARKEDNESS

The locus of main stress in Cairene Classical Arabic is explained by constructing trochaic feet from left to right, and then placing main stress on the rightmost available foot except when this would give a final stressed syllable (McCarthy 1979a; de Lacy 1998b). Cairene Classical Arabic is opaque because there is actually no evidence of iterating secondary stress (and hence foot-structure) on the surface except the placement of main stress. Rule-based theories handled this by constructing foot-structure to function as a scaffold for the correct placement of main stress, and subsequently deleting this structure in a process called ‘conflation’. The examples are taken from de Lacy (1998b). The first column illustrates the (abstract) fully-footed forms. The middle column gives the surface stress pattern.

<table>
<thead>
<tr>
<th>Footed form</th>
<th>Surface stress</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (kā.ta)(bā)</td>
<td>[kātaba]</td>
<td>‘he wrote’</td>
</tr>
<tr>
<td>(fā.ḍa)(rā.tu)(hū)</td>
<td>[fadžratuhu]</td>
<td>‘his tree’</td>
</tr>
<tr>
<td>b. (kā):(tāba)</td>
<td>[kataba]</td>
<td>‘to keep up’</td>
</tr>
<tr>
<td>(mūn)(tā.xa)(bā)</td>
<td>[muntāxaba]</td>
<td>‘a correspondence’</td>
</tr>
<tr>
<td>c. (kā)(tā)(tā)</td>
<td>[katātā]</td>
<td>‘you (m.sg) wrote’</td>
</tr>
<tr>
<td>(mū)(gā):(tāla)</td>
<td>[mugatāla]</td>
<td>‘fighter (pausal)’</td>
</tr>
<tr>
<td>d. (kā)(tāb)</td>
<td>[katab]</td>
<td>‘write’</td>
</tr>
<tr>
<td>(fā.ḍa)(rā.tu)(hū)(māc)</td>
<td>[fadžratuhumā]</td>
<td>‘their (du) tree’</td>
</tr>
<tr>
<td>(?ād)(wi.ja)(tū.hu)(māc)</td>
<td>[wdwijatūhumā]</td>
<td>‘their (du) drugs’</td>
</tr>
</tbody>
</table>

In de Lacy’s analysis, surface opacity results from the activity of the constraint $O-Ident-σ$, which requires that the surface form bear the main stress in exactly the same position as the fully-footed sympathetic candidate. The sympathetic selector is the markedness constraint $Parse-σ$, which behaves as if undominated for the purposes of selecting the sympathetic candidate. Abstracting away from the details of stress assignment, the sympathetic account of stress assignment in Cairene Classical Arabic looks like in (146). Simplifying slightly, I will take $*Ft$ to penalize secondary stress feet only.

---

12 The literature on this dialect is quite extensive and I will not pretend to cite all of it here. See the most recent piece of work on Cairene (de Lacy 1998b) for further references.
It is simple enough to translate this vision into Virtual Phonology. Instead of a markedness selector, the $\mathcal{V}$-relativized $\mathcal{P}ars\mathcal{E}-\sigma$ dominates $^*\mathcal{F}t$. $\mathcal{V}\mathcal{S}-\mathcal{I}dent-\delta$ replaces $\mathcal{O}-\mathcal{I}dent-\delta$ in the optimization of the actual output stress pattern.

### 3.4.6 Yidi Deep Phonotactics

The Australian language Yidi has two putative input restrictions on the form of the stem. Neither restriction makes sense on the surface, because both configurations are possible on the surface. The restrictions concern the distribution of long vowels in the language. It is argued here that the deep phonotactic properties of Yidi essentially boil down to the virtual restriction that all long vowels are sequences of vowel+glide in $\mathcal{V}$-form.

The most productive source of long vowels in Yidi is a rule of Penultimate Lengthening. In any word with an odd number of syllables, the penultimate vowel is lengthened. For example, /mudjam/ ‘mother’ appears unsuxed as mudjam in the absolutive case, but when suffixed with the purposive suffix -gu, undergoes Penultimate Lengthening to mudjamgu. Similarly, the trisyllabic stems /gudaga/ ‘dog’ and /yabulam/ ‘loya-cane sp.’ undergo lengthening to
3.4. VIRTUAL MARKEDNESS

gudaŋa and yabuldam respectively in the absolutive case. Suffixation of purpose-
ful -gu renders the form parasyllabic, and hence ineligible for Penultimate
Lengthening, giving gudaŋāgu and yabuldamgu. The basis of the alternation
is the variation of foot form as a function of syllable count: parasyllabic words are
parsed into trochees, while imparasyllabic words are parsed iambically. Pen-
ultimate Lengthening thus targets the final iambic foot of the word, and applies
only when there is a stray syllable following this foot. 

The crux of the problem is that the rule produces configurations which,
if permitted underlingly, predicts the existence of alternations which are un-
grammatical and unattested. Restrictions on the input apparently have to be
invoked, and the Richness of the Base must once again be pulled from the mine.

Metrical parsing and Penultimate Lengthening interact opaquely with a pro-
cess of Apocope (Dixon’s ‘Final Syllable Deletion’), which deletes a final un-
footed vowel under certain conditions. Thus only the final syllable of iambic
(σ|σ)σ is eligible for Apocope — the final syllables of parasyllabic (σ|σ) or (σ|σ|σ)
are not. The conditions on Apocope are complex. For one thing, it applies only
variably to lexical roots. Dixon classifies V-final trisyllabic lexical roots into
reducing and non-reducing (Dixon 1977: 59). The Yidi dictionary contains
115 items which are in principle eligible for Apocope. Of these, a clear majority
of 81 (just over 70%) are reducing. The remaining 30% must be diacritically
marked as ‘resistant’, subject to Max-V.

Apocope is blocked by phonological and morphological factors. Apocope
must yield a legal final consonant of Yidi, one of the set /l r y m n n y/. Thus,
gudaŋa/ is non-reducing, for purely phonological reasons, because the
result would be *gudāŋ, with illicit final *ŋ13. Word-final clusters are repaired by Cluster Simplication, e.g. Penultimate
Lengthening and Apocope on /bupa+ngu/ gives intermediate [bupaŋg]. The
final consonant cluster is repaired on the surface, giving [bupaŋ]. The exam-
ple illustrates that Af-Max-C is dominated. This is shown in (148) for the
absolutive and ergative forms of /bupa/ ‘woman’ and /ğunangara/ ‘whale’.

(148)

<table>
<thead>
<tr>
<th>Input</th>
<th>/bupa/</th>
<th>/bupa+ngu/</th>
<th>/günangara/</th>
<th>/günangara+ngu/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>(bupa)</td>
<td>(bupaŋg)gu</td>
<td>(günangara)</td>
<td>(günangaraŋg)</td>
</tr>
<tr>
<td>Penult Lengthening</td>
<td>n/a</td>
<td>(bupaŋg)gu</td>
<td>n/a</td>
<td>(günangara)gu</td>
</tr>
<tr>
<td>Apocope</td>
<td>n/a</td>
<td>bupaŋg</td>
<td>n/a</td>
<td>günangaraŋg</td>
</tr>
<tr>
<td>Cluster Simplication</td>
<td>n/a</td>
<td>bupaŋg</td>
<td>n/a</td>
<td>günangaraŋg</td>
</tr>
<tr>
<td>Output</td>
<td>[bupa]</td>
<td>[bupaŋg]</td>
<td>günangara</td>
<td>günangaraŋg</td>
</tr>
</tbody>
</table>

There is one final situation in which Apocope is blocked. This is a situation
in which Apocope would yield a cluster eligible for repair by Cluster Simplicit-
cation, as in bupaŋg, but where Apocope is blocked by the demands of Mor-
phReal. For example, in the genitive, /bupa+ni/ gives surface [bupaŋn], the
suffixal vowel apocopating. However, added to a consonant-final stem /guygal/
‘bandicoot’, we get guygadni, without Apocope. Of the two other relevant can-
didates, *guygadni is out because of the illegal word-final cluster, and *guygad,

13 Mappings such as /gudaŋa/ → gudāŋ or /gudaga/ → gada must also be ruled out. This can
be achieved by ranking both R-Ident[has] and R-Max-C high.
with Cluster Simplification, is bad because the genitive has no phonological exponent, in violation of MorphReal (see Samek-Lodovici 1996).

(149) MorphReal

Every morpheme must have some non-null exponent.

Full nominal declensions are given here in (150) for marqu- ‘grey possum’, mulari- ‘initiated man’, and gindanu- ‘moon’. The trisyllabic stems mulari- and gindanu- differ with respect to the behaviour of the final vowel. The former is a non-reducing stem; the latter reducing. The suffixes with participate in the Apocope process are shown underlined.¹⁴ As an aid to the reader, stress has been retranscribed into Dixon’s examples.

(150)

<table>
<thead>
<tr>
<th>Case</th>
<th>marqu</th>
<th>mulari</th>
<th>gindan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolutive</td>
<td>marqu</td>
<td>mulari</td>
<td>gindan</td>
</tr>
<tr>
<td>Ergative</td>
<td>marqu</td>
<td>mulari</td>
<td>gindan</td>
</tr>
<tr>
<td>Dative</td>
<td>marqui</td>
<td>mulari</td>
<td>gindan</td>
</tr>
<tr>
<td>Purpose</td>
<td>marqu</td>
<td>mulari</td>
<td>gindan</td>
</tr>
<tr>
<td>Locative</td>
<td>marqu</td>
<td>mulari</td>
<td>gindan</td>
</tr>
<tr>
<td>Abative</td>
<td>marqu</td>
<td>mulari</td>
<td>gindan</td>
</tr>
<tr>
<td>Comitative</td>
<td>marqu</td>
<td>mulari</td>
<td>gindan</td>
</tr>
<tr>
<td>Genitive</td>
<td>marqu</td>
<td>mulari</td>
<td>gindan</td>
</tr>
</tbody>
</table>

Generally speaking, bisyllabic stems evince trochaic parsing in the absolute, and iambic parsing with Penultimate Lengthening in the oblique cases (which are virtually trisyllabic). Trisyllabic stems evince iambic parsing with Penultimate Lengthening in the absolute, and trochaic parsing in the oblique cases.

So far, surface long vowels are all products of the application of the Penultimate Lengthening rule. There are other sources of long vowels in Yidj, however. There are three sources of invariant V:

1. Certain ‘prelengthening’ suffixes cause the vowel of the preceding syllable to surface as long.

¹⁴ Some of the case suffixes do not alternate, or fail to alternate in the expected way. These are briefly enumerated here.

1. The purpose suffix -gu is not predicted to alternate because g is not a legal final consonant of Yidj.

2. The locative suffix -la alternates with the presence of vowel lengthening on the final syllable of the bisyllabic stem. For the locative of a bisyllabic, we may suppose a virtual representation [margul:]. On the surface, Apocope will apply, but there is also ad hoc deletion of the consonant of the suffix, giving margu: rather than the expected, but ungrammatical form *margul.

3. The comitative suffix fails to trigger Penultimate Lengthening in a parasyllabic stem. This implies that the suffixal vowel of -mu is absent in the virtual form. This may be explained in terms of a sequential markedness constraint *mu. Ranked above V Max-V, the vowel will be absent in the V form and unable to condition the expected Penultimate Lengthening. This is an example of a rule-sandwiching effect.

4. Finally, the dative suffix -nda fails to exhibit a V-∅ alternation. This is simply a lexically idiosyncratic property of this particular suffix.
2. Some surface instances of [i:] result from the monophthongization of underlying /iy/.

3. There is a class of parasyllabic stems in which vowel length is an invariant property of the stem.

An example of a prelengthening suffix is the antipassive -zdī. Adding this to the stem meaning ‘see, look’ along with the past tense ending -ṇu (to get a parasyllabic word), we get wawāzdīṇu. The antipassive suffix causes lengthening on the second syllable.

Historically at least, this lengthening was most probably compensatory, arising from the deletion of a segment. Crowhurst and Hewitt (1995) analyze these suffixes as having a floating mora in the underlying representation, which docks on a preceding vowel.

There are sixteen morphemes with invariant length, exemplified in (151).

(151)  
durgu:    'mopoke owl'
yibu:    'miming in dance routine'
galambara:    'march fly'
wāpintaːraː:    'what’s the matter?’
wāpabuga    'white apple tree'  

All invariant long vowels occur in odd-numbered positions. This is readily explainable on metrical grounds. Long vowels are permitted exclusively in the head of an iambic foot. Long vowels cannot appear in trochaic words.

Two systematic gaps in the Yidi phonology appear to have to be stated as constraints on underlying representations (Hayes 1999b: 184):

(152)  
a. There are no trisyllabic stems with a long vowel.
b. There are no long vowels in closed syllables.

There are thus no alternations of the form *CVCV:C~CVCV:C-CV or *CVCV:CV~CVCV:CV-CV. Seen from the perspective of the surface phonology of Yidi, these restrictions make little sense: Penultimate Lengthening is so productive in Yidi as to be virtually exceptionless. Similarly, the underlying ban on long vowels in closed syllables finds just as little surface motivation: Penultimate Lengthening applies in closed syllables as it does in open (e.g. mudāmgu). The approach taken to each of these restrictions here is different. The absence of *CVCV:CV~CVCV:CV-CV alternations is attributed to a constraint militating against CVCV:CV-CV. Specifically, it will be shown that the structure in question is defined by a unique conjunction of phonological and morphological properties, which distinguishes it from other phonologically identical (but grammatically possible) structures. The absence of alternations of the type *CVCV:C~CVCV:C-CV is reinterpreted here as a ban on roots of the form *CVCV:C. It is proposed that long vowels which are not the product of Penultimate Lengthening are underlyingly clusters of vowel + glide, and that the absence of *CVCV:C reflects an independently motivated and undominated constraint in Yidi phonology against word-final consonant clusters, because *CVCV:C has to be construed as CVCVGC underlyingly.
How can we ensure that alternations of the type *ginda:n*–*gindan uNgu* are always the result of Penultimate Lengthening? The most expedient way to effect this result is to ban long vowels in `underlying' or virtual forms using a constraint \(^*V^\). However, this doesn't jive with the fact that there are invariant long vowels in the surface phonology of Yidiŋ, whose source is not Penultimate Lengthening. It would initially appear difficult to square the existence of invariant long vowels with the apparent ban on underlying vowels in virtual forms. Nevertheless, a solution is forthcoming. A clue is provided by Dixon's proposal (Dixon 1977: 77) that some instances of [i] result from the monophthongization of underlying /iy/. For example, *galbiː* is inflected in precisely the same way as the y-final stem *d'anguy* `possum sp.', selecting allomorphs which are otherwise only found on y-final stems. This suggests that *galbiː* is derived from the input /galbiːy-/. Crucially, this explains why *galbiː* behaves as if it were consonant-final with respect to Apocope: Apocope is blocked as if to avoid a final consonant cluster. (153) illustrates the declension of *d'anguy* and *galbiː*.

(153)

<table>
<thead>
<tr>
<th>Case</th>
<th>d'anguy</th>
<th>galbiː</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute</td>
<td>galbiː</td>
<td>galbiː</td>
</tr>
<tr>
<td>Ergative</td>
<td>galbiːa</td>
<td>galbiːa</td>
</tr>
<tr>
<td>Dative</td>
<td>galbiːa</td>
<td>galbiːa</td>
</tr>
<tr>
<td>purposive</td>
<td>galbiːa</td>
<td>galbiːa</td>
</tr>
<tr>
<td>Locative</td>
<td>galbiːa</td>
<td>galbiːa</td>
</tr>
<tr>
<td>ablative</td>
<td>galbiːa</td>
<td>galbiːa</td>
</tr>
<tr>
<td>Comitative</td>
<td>galbiːi</td>
<td>galbiːi</td>
</tr>
<tr>
<td>genitive</td>
<td>galbiːi</td>
<td>galbiːi</td>
</tr>
</tbody>
</table>

We may maintain the ban on long vowels in virtual forms if all invariant long vowels are the result of compensatory lengthening on the surface through the loss of some abstract segment, e.g. a homorganic glide. Thus surface forms such as *durguː* arise from underlying /durguːy/. Any potential vowel length contrast is neutralized to short in any position except the penultimate syllable of an imparisyllabic word. Thus, \(^*V^\) must dominate \(\mathbb{Wt}^\)-Ident, but the constraint driving Penultimate Lengthening (call it \(\text{PenLength}\)) must in turn dominate \(^*V^\): (Penultimate Lengthening must take place in the virtual form, since, on the surface, the process is frequently rendered non-surface-apparent because of Apocope, e.g. /gindam/-→*gindam*.) It doesn't matter whether we choose /gindam/ or /gindam/ as the input: both converge robustly on the output-pair ⟨[gindam],[gindam]⟩.

(154)
3.4. VIRTUAL MARKEDNESS

The corresponding surface form, *gindan* is apocopated, and, consequently, the lengthened vowel is no longer penultimate. The non-surface-apparent of the Penultimate Lengthening in the surface form must follow from the undominated status of VS-Wt-Ident.

All long vowels in positions other than penultimate in an imparisyllabic word must derive from virtual vowel+glide. Thus, the inputs */durqu:/ and */durquw:/ must both neutralize to virtual */durquw:/.\(^{15}\) This means that long vowels in non-penultimate position have to be diphthongized in the virtual form. I will take this to reflect the ranking of *V over the faithfulness constraint VS-µ-Ident\[approximant\]. Surface *durqu:* can only be the product of Monophthongization on the surface.

\(^{15}\) Again, we have the option of neutralizing the rogue input using an IO-chain-shift to, say, *durqu:* IO-chain shifts were considered and rejected in §3.4.3.
A third candidate, durgu, not included in (157), evinces entirely gratuitous lengthening of the penultimate syllable: gratuitous because the condition of imparisyllabic is not satisfied. Hence, the candidate fares no better on Pen-Length than (157a), merely adding an unnecessary violation of Wt-Ident.

Invariant long vowels are not represented as such in virtual form: rather, they are represented as sequences of vowel+homorganic glide. These glides are preserved at virtual form, reflecting the relatively high rank of $\mu_w$-Ident[approximate] over *VG.

(158) *VG
A vowel followed by a homorganic glide is disallowed.

For the input /durguw/, this gives (159).

(159)
The remaining problem is to explain why there are no invariant long vowels in trisyllabic stems. By hypothesis, invariant long vowels are the result of compensatory lengthening through the loss of a homorganic glide. Thus, the problem boils down to why there are no stems of the form /CV(C)VGCVC/. Hypothetical /gugulu+ngu/ cannot be permitted to map to *guguluangu. In purely phonological terms, there is nothing wrong with *guguluangu, since precisely the same length pattern can be found in examples such as wawa:djuw 'saw', duqunula 'mopoke owl-GEN-LOC', and monomorphemic waqaduq 'white apple tree'. However, in non-occurring *guguluangu, the relationship between metrical and morphological structure is crucially different: *(gugu:(luNgu)). (Underlining of the lexical stem makes this relationship clear.) Iambicity is tolerated so long as the iambic foot is not split by a stem boundary. Hence the grammaticality of (wawa:djuw), (duqunula), and (waqaduq). Essentially, an iambic foot cannot contain a stem boundary. We cannot make this constraint hold of feet in general, since, of course the corresponding trochaic feet are fine, e.g. (ginda)(mingu).

(161) **Right-Anchor (Stem, Foot)**
Every stem boundary must coincide with a foot boundary.

(162) **Trochee**
Feet must be trochaic.

The ungrammaticality of forms such as *(gugu:(luNgu)) arises due to the violation of the local conjunction of (161) and (162) within the domain of the foot. Thus, an apparent underlying restriction — the absence of trisyllabic stems with invariant long vowels in the second syllable — may be reanalyzed as a surface restriction.

The second question, why there are no alternations of the type CVVC:C∼CVVC:C- CV must boil down to the impossibility of /CVVC:C/ as a possible stem. Under the assumption that invariant long vowels (vowels in any other position than penultimate in an imparisyllabic word) are derived from vowel+glide, the gap has a simple phonotactic explanation: there are no word-final consonant clusters (Dixon 1977: 35), reflecting a highly ranked constraint *CC_Wd. ¹⁰ *CC_Wd must

¹⁰The ban does not extend to syllable-final position in general. Yidi permits complex codas word-internally, as attested by such examples as balmip ‘grasshopper’, dahlkay ‘white cedar tree’, muyggi ‘cicatrices’, and wurmbo ‘asleep’.

<table>
<thead>
<tr>
<th>/durguw/</th>
<th>*VS-Mask≠Ident[approx]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>![durgu]~durguw</td>
</tr>
<tr>
<td>b. ☞</td>
<td>![durgu]~durgu:</td>
</tr>
<tr>
<td>c.</td>
<td>![durgu]~durgu</td>
</tr>
</tbody>
</table>
dominate Ψ-Max-C. Hypothetical /guquwl/ must map to [gugu] in virtual form.

Still, some housekeeping remains. The issue is complicated by the niceties of allomorph selection. The idea proposed here, that long vowels are underlyingly sequences of vowel+glide is not new. In fact, Dixon briefly entertains the hypothesis but ends up rejecting it. On analogy of the behaviour of nouns with final /iy/, Dixon (1977: 83) experiments with the idea of representing roots like durgu: underlyingly as /durgu w/, but rejects the idea on the following grounds. Words such as galli‘:‘catfish’ behave as if consonant-final for the purposes of allomorph selection (Dixon 1977: 77). Specifically, Apocope is blocked, as if to avoid a final consonant cluster, which is impermissible in Yidiŋ. Thus, the genitive of galli‘: is galli‘ni. In this, galli‘: patterns similarly to a word like d’anguy ‘possum sp.’, whose genitive form is d’anguyi. Roots like durgu:, on the other hand, permit Apocope in the regular way, and in doing so are behaving as if vowel-final. Thus, instead of the expected genitive form *durgu:n, we actually find apocopated durgu:n. Put differently, while clusters of y+C block the application of Apocope, clusters of w+C do not. (163) tabulates the declension of durgu: against that of galli‘:

(163)

<table>
<thead>
<tr>
<th></th>
<th>durgu:</th>
<th>galli‘:</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolutive</td>
<td>durgu:</td>
<td>galli‘:</td>
</tr>
<tr>
<td>ergative</td>
<td>durguŋ</td>
<td>galliŋŋu</td>
</tr>
<tr>
<td>dative</td>
<td>durgunda</td>
<td>gallbinda</td>
</tr>
<tr>
<td>purposive</td>
<td>durgunu</td>
<td>gallbugu</td>
</tr>
<tr>
<td>locative</td>
<td>durgu:</td>
<td>galliŋḍa</td>
</tr>
<tr>
<td>ablative</td>
<td>durgum</td>
<td>gallinu</td>
</tr>
<tr>
<td>comitative</td>
<td>durguy</td>
<td>galliŋḍi</td>
</tr>
<tr>
<td>genitive</td>
<td>durgu:n</td>
<td>gallini</td>
</tr>
</tbody>
</table>

It is not unknown to find such splits in the glide system, where one glide might behave more like a consonant or vowel than the other at a different place of articulation. Glides phonologically have a foot in both the consonantal and the vocalic camp. Isthmus Nahua, for example, has a a rule of domain-final Sonorant Devoicing which illustrates the point (Kenstowicz and Kisseberth 1979). The glides /y/ and /w/ are both eligible for devoicing, but the conditions under which devoicing occurs differs depending on the glide. /y/ devoices syllable finally, except when followed by a voiced consonant. /w/ is much more resistant to devoicing, undergoing the rule only when in word-final position, and only optionally at that. In this respect, /w/ behaves much more like a vowel. /y/ on the other hand exhibits behaviour which is much more akin to the other, clearly [+consonantal] sonorant /l/, which always devoices syllable finally. We find exactly the same split in the Yidiŋ example: clusters of w+C fail to block Apocope simply because /w/ is ‘more vocalic’ than /y/.

3.4.7 Vowel Harmony in Finnish

In Finnish, there is a process of Vowel Harmony which spreads the value for [back] from the initial, main-stressed syllable of the word to the right edge of the word domain.

The surface vowel system of Finnish is given in (164).
What is interesting about the Finnish system is the absence of the [+back] counterparts of /i/ and /e/: /u/ and /y/ are systematically absent in surface forms of Finnish. This gap correlates with the fact that /i/ and /e/ are neutral with respect to Vowel Harmony: backness spreads across these vowels leaving them unaffected.

Abstracting away from the neutrality of /i/ and /e/, all Finnish words must harmonize for [±back]. Finnish suffixes thus come in a [-back] and a [+back] shape, depending on the [back] specification of the stem.

(165) a. tyhmæ-stæ ‘stupid-ILL’
    b. tuhma-stæ ‘naughty-ILL’

Spreading of backness ‘skips’ the neutral vowels /i/ and /e/. In a [+back] word, a new [+back] domain resumes in the syllable following the neutral vowel.

(166) illustrates the resumption of the harmony domain following a syllable containing /i/, in this case, introduced by the possessive suffix ni ‘my’. Examples are from Walker (1998).

(166) a. værttinæ-lle-ni-hæn ‘spinning wheel-ADESS-1SG.POSS-as you know’
    b. paltti-na-li-ni-han ‘linen cloth-ADESS-1SG.POSS-as you know’
    c. ljo-da-kse-ni-kø ‘hit-INF-TRANS1-1SG.POSS-Q’
    d. ljo-da-kse-ni-kø ‘create-INF-TRANS1-1SG.POSS-Q’

Following recent research by Chiosáin and Padgett (1997), Gafos (1996), Walker (1998), and others, I will assume that feature spreading is segmentally strictly local. There are in reality no gapped configurations or discontinuous feature domains interrupted by neutral segments as assumed in much autosegmental phonology. The neutrality of /i/ and /e/ arise out of the conflict between the constraint *[+back, −round], which militates against /u/ and /y/, and R-Align[back], which attempts to drive every instance of [±back] over to the right edge of the word. It is clear that in words with neutral vowels, R-Align[back] is going to be violated, because the word will be split into more than one harmonic domain. As a consequence, any feature domain which precedes a neutral vowel will incur violations of R-Align[back] by however many syllables it stands separated from the right edge of the domain. For example, paltti-nallanihan consists of three distinct [+back] feature domains, the second and third resuming following a [−back] neutral vowel i. The first [+back] domain thus stands misaligned from the right edge of the word by five syllables, and the second by two. R-Align[back] is thus obviously not surface-true. The markedness constraints *[u] and *[y], by contrast, are surface-true in Finnish. However, at the virtual level, it must be the case that R-Align[back], or its Θ-relativized counterpart, R-Align[back] must be undominated. In this way, [±back] spreads right to form a single feature domain, forming surface-ill-formed back unrounded vowels *[u] and *[y] in the process. Feature domains are represented here by underlining.
CHAPTER 3. VIRTUAL PHONOLOGY

(167)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [palttinallanihan]</td>
<td>*! ****, **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ☞ [paltumallanuhan]</td>
<td>**</td>
<td>!***</td>
<td></td>
</tr>
<tr>
<td>c. [palttinellañhæn]</td>
<td><em>!</em>***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On the surface, R-ALIGN[back] is dominated by *[+back, −round]. However, reference is made to the [back] value of the harmonizing vowels in the V-form by way of VS-Ident[back]. VS-Ident[back] must itself be dominated by *[+back, −round], otherwise back unrounded vowels would be preserved in surface forms. VS-Ident[back] must also dominate R-ALIGN[back] to permit the formation of more than one feature domain. The result of permuting the ranking of these two constraints would be the blocking of spreading by i and e.

(168)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. palttinallanihan</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ☞ paltumallanuhan</td>
<td>!*</td>
<td>!***</td>
<td></td>
</tr>
<tr>
<td>c. palttinellañhæn</td>
<td>*<strong>!</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4.8 Duke of York Effects

All of the examples in sections §3.4.1 to §3.4.6 instantiate the Duke of York gambit, at least potentially. The reason for this is the non-uniqueness of the input. For example, the Tokyo Japanese word baratoNe `rose thorn' is consistent in its phonological behaviour with either /bara+toNe/ or /bara+toge/ in the input. Only /bara+toge/ stands in a non-cumulative relationship to the actual surface form. Not all of them are however non-cumulative in the sense of McCarthy (1999). This is because cumulativity is defined in terms of shared faithfulness violations (see also McCarthy 1998a: 33ff.). Not all Duke-of-York effects involve Cumulativity violations because not all involve violations of Faithfulness. The Cairene Classical Arabic example is a case in point: since faithfulness constraints on foot structure arguably do not exist, the selection of a sympathetic candidate with foot structure does not compromise Cumulativity. Huariapano, however, does violate Cumulativity, since the relationship between virtual form and surface form involves not the elimination of stresses, but the shifting of stresses to other positions, in violation of Ident[Stress].
3.5 Learnability issues

McCarthy (1999) claims a major empirical advantage for Sympathy Theory over rule-based theory in that Sympathy Theory in principle excludes derivations of the type /A/ → B → A, in which A is mapped to some non-A expression and subsequently mapped back to A by a later rule, and, indeed, McCarthy presents a strong case for excluding such Duke-of-York derivations on empirical grounds: they are arguably unattested in natural language.

A recent paper by Kiparsky (1998), however, takes issue with McCarthy's empirical claim, and argues that there are in fact genuine, but perfectly benign examples of Duke-of-York rule interactions. Kiparsky argues that these Duke-of-York effects arise as a consequence of the serial ordering of lexical-phonological levels. The example considered by Kiparsky is of exactly the benign kind we have been considering here: they arise as an artefact of the Richness of the Base. An example is furnished by Palestinian Arabic (Brame 1974), where there is a cross-cycle underapplication of a syncopation rule. i-Syncopation deletes an unstressed i preceding an open syllable. Stress is assigned essentially according to the Latin Stress Rule (stress a heavy penult, else the antepenult). Suffixation of subject agreement suffixes alters the metrical profile of the word, triggering i-Syncopation in different locations as a function of where the open syllables fall in the stress domain. The normal application of the process is illustrated in (169).

(169) a. /fīhim/ fīhim 'he understood'
   b. /fīhim-na/ fhīmna 'we understood'
   c. /fīhim-u/ fīhu 'they understood'

The verb possesses a further morphological layer inhabited by suffixes encoding object agreement. The paradigm for 'he understood X' is shown in (170).

(170) a. /fīhim/ fīhim 'he understood'
   b. /fīhim-ak/ fhīmak 'he understood you (masc)'
   /fīhim-ik/ fhīnik 'he understood you (fem)'
   /fīhim-u/ fīhu 'he understood him'
   c. /fīhim-ni/ fihīmni 'he understood me'
   /fīhim-ha/ fīhinha 'he understood her'
   /fīhim-na/ fihīmna 'he understood us'

In (170c), stress is assigned according to the canonical pattern, but i-Syncopation underapplies — the unstressed i, shown underlined, surfaces even though the conditions for its deletion are surface-true. On the addition of the object suffix -na 'us' to the base form fīhim, the stress shifts to the resulting heavy syllable, giving fihīmna 'he understood us'. Since, as Kiparsky observes, in the input /fīhim/, the initial vowel is unstressed (or at least can be under the Richness of the Base), we are dealing with a Duke-of-York effect, albeit one of a perfectly harmless kind. Moreover, it is one which poses no problems for the learner. The reason it is harmless, I suggest, is that we are dealing with a cross-cycle Duke-of-York effect. It merely undoes the phonology of a 'previous' cycle, but the phonology of the inner morphological layer is learnable independently of the phonology of the outer morphological layer — the learner has robust access to the i-Syncopation rule due to the existence of paradigms as in (169) lacking an object suffix.
CHAPTER 3. VIRTUAL PHONOLOGY

The undesirable cases involve the undoing of some process within one and the same cycle. Kiparsky (1998: 70) invites us to imagine an example of such a case three-process Duke-of-York derivation as in (171). (The same example is used by McCarthy 1999: 378f. to argue for the correctness of the cumulativity-based approach to Sympathy over the inter-candidate faithfulness approach).

(171) Hypothetical three-process Duke-of-York derivation

<table>
<thead>
<tr>
<th>Input</th>
<th>/mat/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epenthesis</td>
<td>∅ → (CV:C)σ</td>
</tr>
<tr>
<td>Palatalization</td>
<td>t → /i</td>
</tr>
<tr>
<td>Apocope</td>
<td>V → /∅</td>
</tr>
<tr>
<td>Shortening</td>
<td>V → /ʃ</td>
</tr>
<tr>
<td>Output</td>
<td>[mαːʃ]</td>
</tr>
</tbody>
</table>

In this example, epenthesis of /i/ applies first to repair a superheavy syllable. This in turn triggers the crucial bit of phonological business, namely the palatalization of the preceding /t/. Subsequently, though, the epenthetic /i/ deletes by Apocope, but the palatalization it conditioned remains. At the last stage of the derivation, the superheavy syllable is repaired yet again, this time by the shortening of the syllable nucleus. Obviously, the interaction is distinctly dubious: we don’t want to be able to predict a language which permits this pairing of input and output. McCarthy argues that the spuriousness of the example is to be attributed to the Duke-of-York effect itself: what’s wrong with this derivation is that Epenthesis is undone by Apocope at a later stage. McCarthy shows that this process is perfectly tractable given an inter-candidate faithfulness perspective to Sympathy. Every process in (171) is natural, and can be approximated with the following constraint rankings (McCarthy 1999: 379).

(172)

*µµµσ ≫ Dep-V Trimoraic syllables repairable by epenthesis.
*µµµσ ≫ Max-µ Trimoraic syllables repairable by shortening.
Dep-V ≫ Max-µ Shortening overrides epenthesis.
*σµ ≫ Ident[high] Palatalization.

Taking *Max-µ to be the selector, the tableau in (173) shows that the interaction in (171) can be modeled easily within Sympathy Theory based on intercandidate faithfulness. The point carries over directly to Virtual Phonology as well. In the tableaux, mαːʃ is precisely the candidate we don’t want to win. The desired winner is the transparent candidate mat. A new constraint, *Final-V, has been added to take care of apocope.

(173)
3.5. **LEARNABILITY ISSUES**

<table>
<thead>
<tr>
<th>/maat/</th>
<th><em>FINN</em></th>
<th><em>SYM</em></th>
<th><em>Dep-V</em></th>
<th><em>MAX-µ</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. -Encoding mat</td>
<td>: : : : !</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.  Maaçi</td>
<td>: ! : :</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.  Nucl mač</td>
<td>: : : :</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.  Maat</td>
<td>: ! : :</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.  Maati</td>
<td>: ! : * :</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.  Maač</td>
<td>: ! : :</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The undesired candidate (173c) can clearly be optimized given the inter-candidate faithfulness approach. The cumulativity-based approach, on the other hand, cannot derive this result, which, according to McCarthy is precisely as desired.

(174)

<table>
<thead>
<tr>
<th>/maat/</th>
<th><em>FINN</em></th>
<th><em>SYM</em></th>
<th><em>Dep-V</em></th>
<th><em>MAX-µ</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.  Encoding mat</td>
<td>: : : : !</td>
<td>:</td>
<td>(\infty)'s</td>
<td></td>
</tr>
<tr>
<td>b.  Maaçi</td>
<td>: ! : :</td>
<td>:</td>
<td>(\infty)'s</td>
<td>!</td>
</tr>
<tr>
<td>c.  Nucl mač</td>
<td>: : : :</td>
<td>:</td>
<td>(\infty)'s</td>
<td>!</td>
</tr>
<tr>
<td>d.  Maat</td>
<td>: ! : :</td>
<td>:</td>
<td>(\infty)'s</td>
<td></td>
</tr>
<tr>
<td>e.  Maati</td>
<td>: ! : * :</td>
<td>:</td>
<td>(\infty)'s</td>
<td></td>
</tr>
<tr>
<td>f.  Maač</td>
<td>: ! : :</td>
<td>:</td>
<td>(\infty)'s</td>
<td></td>
</tr>
</tbody>
</table>

On the cumulativity-based approach, the candidate which is actually optimized is in fact the transparent candidate (173a), exactly as McCarthy desires. Neither the transparent nor the optimal candidate have a superset of the candidate's unfaithful mappings in this case, but when the decision is passed down to the lower ranked constraints, the transparent candidate is favoured, as it inevitably will be when it scores the same marks as the opaque candidate.

Kiparsky argues that a formal approach to Duke-of-York effects of the more malign variety is misconceived: we can’t rule out the malign variety by formal means without also eliminating those which are merely a benign artefact of the Richness of the Base. According to Kiparsky, the correct explanation for the conspicuous absence of such dubious derivations lies, rather, in considerations of learnability. In the case at hand, the fault is more appropriately seen as lying,
not in its non-cumulativity, but in the tenuousness of the relationship between the surface form and the input. What can be excluded on grounds of learnability need not, therefore, be written into the formalism of the grammar, in Kiparsky's view. A similar point has been made by Orgun (1996: 119). For both authors, the interest in learnability stems from the need to restrict the degree of variation between cophonologies in the grammar. Kiparsky's project is the reworking of Lexical Phonology within an OT framework. However, his approach is a serialist one, and he proposes that the levels recognized in Lexical Phonology (Stem, Word, Postlexical) are constituted by independent and complete OT grammars (rankings of CON), or 'cophonologies' such that the output of the Stem Level serves as the input to the Word Level, and the output of the Word Level serves in turn as the input to the Post-lexical Level. Serial OT has reaped considerable criticism from Benua (1998) and McCarthy (1999) on the grounds that cophonologies may differ from each other to an arbitrary degree, and that the theory of grammar it defines is insufficiently restrictive. Orgun explicitly addresses the problem of cophonology proliferation, but rejects the notion that arbitrary differences mitigate the use of any formal universal constraints on their variation, concluding similarly that the appropriate explanatory locus is acquisition.

This criticism is not merely relevant for serialist approaches such as the one proposed by Kiparsky, but it may also be brought to bear on Virtual Phonology despite its avowed parallelism. Since Virtual Phonology in principle allows any permutation of the IV- and VS-Faithfulness constraints, it ends up overgenerating from the point of view of McCarthy's and Benua's critique. Adding the set of virtual markedness constraints into the mix would appear to make the situation even worse, since it allows in principle for an arbitrary degree of difference between the virtual form and the surface form. But the same considerations of learnability can be made to neutralize the argument. The reason is that the only evidence the learner has for opaque phonological generalizations is misapplication in the surface form. Unless the phonological generalization holding of the virtual form has some surface reflex, the generalization cannot be learnt.

Imagine a language, Tiberian Hebrew', which is identical to Tiberian Hebrew\(\textsuperscript{\dag}\) in terms of the phonology of the virtual form, that is word-final glottal stops are preserved and they trigger epenthesis, e.g. \[\text{deP}\]. In Tiberian Hebrew', IV-Max-C dominates CODA\textsuperscript{\dag} exactly as in real Tiberian Hebrew. Suppose further that, on the surface, what we find is transparency, i.e. the surface form is transparent \text{def}. IV-Max-C and VS-Max-C are dominated by CODA\textsuperscript{\dag}, giving word-final deletion of the glottal stop, and VS-Def-V dominates VS-Max-V, thereby enforcing the transparency: epenthesis fails to apply because it is not surface-motivated. This is a Duke-of-York effect. Formally, there is nothing to prevent such a language from being generated. Nevertheless, the mini-grammar we have just sketched is hardly a serious candidate as a psychologically real model of the Tiberian Hebrew' mapping /\text{def}?/ →\text{def}.

But the reason is not to be traced to any architectural flaw in the theory of Virtual Phonology. There is merely a simpler grammar which describes the mapping in question, a grammar which avoids postulating the operation of a phonological process which the learner can have no rationale for positing given the available surface evidence. Hypothetical Tiberian Hebrew' instantiates what we might dub vacuous opacity. In conclusion, this type of overgeneration cannot
be deemed a liability of the theory.

### 3.6 Residual Problems

Sympathy Theory has recently come under attack by Odden (2000) for being too powerful in other respects. Odden draws attention to a kind of transitivity violation, unattested in natural language, which Sympathy Theory predicts, but which cannot be generated under standard linear ordering of rules. The criticism is addressed here because in this case, Virtual Phonology makes the same predictions as Sympathy Theory.

In rule-based theory, if rule $P$ precedes $Q$, and $Q$ precedes $R$, then it follows from transitivity that $P$ precedes $R$. There can be no language in which $P$ precedes $Q$, and $Q$ precedes $R$, but $R$ precedes $P$, since this would violate transitivity. Odden shows that Sympathy Theory can generate a language with precisely this property, and illustrates the point using an imaginary language, Kalaba, with the following rules. First, the language has a rule deleting a glottal stop preceding a consonant.

$$(175) \quad ? \rightarrow \emptyset/\_C$$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>lim</td>
<td>'tongue'</td>
<td>lim-pa</td>
<td>'tongues'</td>
</tr>
<tr>
<td>to</td>
<td>'child'</td>
<td>topa</td>
<td>'children'</td>
</tr>
<tr>
<td>la?</td>
<td>'fish'</td>
<td>la-pa</td>
<td>'fishes'</td>
</tr>
</tbody>
</table>

Second, there is a syncope rule which deletes high vowels in the context VC CV.

$$(176) \quad V \rightarrow \emptyset/VC\_CV$$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>wali</td>
<td>'rice'</td>
<td>wal-pa</td>
<td>'rices'</td>
</tr>
<tr>
<td>kapu</td>
<td>'basket'</td>
<td>kap-pa</td>
<td>'baskets'</td>
</tr>
<tr>
<td>ugli</td>
<td>'porridge'</td>
<td>ugli-pa</td>
<td>'porridges'</td>
</tr>
<tr>
<td>tungu</td>
<td>'onion'</td>
<td>tungu-pa</td>
<td>'onions'</td>
</tr>
<tr>
<td>neno</td>
<td>'word'</td>
<td>neno-pa</td>
<td>'words'</td>
</tr>
</tbody>
</table>

Finally, there is regressive voicing assimilation between adjacent obstruents.

$$(177) \quad C[\_son] \rightarrow [\_voice]/C[\_voice]$$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>nam</td>
<td>'meat'</td>
<td>nam-gu</td>
<td>'my meat'</td>
</tr>
<tr>
<td>nam</td>
<td>'meat'</td>
<td>nam-pa</td>
<td>'meats'</td>
</tr>
<tr>
<td>naz</td>
<td>'coconut'</td>
<td>nas-pa</td>
<td>'coconuts'</td>
</tr>
<tr>
<td>zig</td>
<td>'load'</td>
<td>zik-pa</td>
<td>'loads'</td>
</tr>
<tr>
<td>tot</td>
<td>'infant'</td>
<td>tod-gu</td>
<td>'my infant'</td>
</tr>
</tbody>
</table>
CHAPTER 3. VIRTUAL PHONOLOGY

The rules are ordered pairwise such that Glottal Deletion precedes Syncope, Syncope precedes Voicing Assimilation, and Voicing Assimilation precedes Glottal Deletion. This last local ordering results in a violation of transitivity, which cannot be modeled under the standard assumptions of rule-based serialism.

(178) **Glottal Deletion > Syncope** ($P > Q$)

<table>
<thead>
<tr>
<th>English</th>
<th>Phonetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>'child'</td>
<td>to-pi 'little child'</td>
</tr>
<tr>
<td>la? 'fish'</td>
<td>la-pi 'little fish'</td>
</tr>
</tbody>
</table>

(179) **Syncope > Voicing Assimilation** ($Q > R$)

<table>
<thead>
<tr>
<th>English</th>
<th>Phonetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>kapu 'basket'</td>
<td>kab-gu 'my basket'</td>
</tr>
<tr>
<td>diizi 'banana'</td>
<td>dis-pa 'bananas'</td>
</tr>
</tbody>
</table>

Voicing Assimilation precedes and is counterfaced by Glottal Deletion. Voicing Assimilation between obstruents is blocked if they are separated underlingly by a glottal stop.

(180) **Voicing Assimilation > Glottal Deletion** ($R > P$)

<table>
<thead>
<tr>
<th>English</th>
<th>Phonetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>'child'</td>
<td>to-? 'old child'</td>
</tr>
<tr>
<td>tot 'infant'</td>
<td>tot-? 'old infant'</td>
</tr>
<tr>
<td>naz 'coconut'</td>
<td>naz-? 'old coconuts'</td>
</tr>
</tbody>
</table>

Olden's point is that Kalaba can be modeled easily using Sympathy Theory, and that since languages like Kalaba are not attested, this should count as a strike against parallelist OT.

The pattern involves the interaction of three markedness constraints in (181) to (183).

(181) *HetVoi
Adjacent obstruents must agree in voicing.

(182) **Syncope**
Short unstressed vowels are disallowed.

(183) *C
A glottal stop is disallowed before a consonant.

Normal application, shown in (184) and (185), provides no evidence for any crucial rankings between the three markedness constraints.

(184)
3.6. RESIDUAL PROBLEMS

<table>
<thead>
<tr>
<th>/lapipa/</th>
<th>*HetVoi : Sync : *C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. lapipa</td>
<td>: *! :</td>
</tr>
<tr>
<td>b. lapipa</td>
<td>: : *!</td>
</tr>
<tr>
<td>c. *lappa</td>
<td>: :</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/lapugu/</th>
<th>*HetVoi : Sync : *C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. lapugu</td>
<td>: *! :</td>
</tr>
<tr>
<td>b. lapugu</td>
<td>: : *!</td>
</tr>
<tr>
<td>c. *labgu</td>
<td>: :</td>
</tr>
</tbody>
</table>

(185)

Symphathy provides a ready way to model the opacity of the interaction between Glottal Deletion and Voicing Assimilation, the locus of the transitivity violation, according to (180). The output is faithful in voicing to the sympathetic candidate which maximizes input consonants, i.e. Max-C is the selector. (186) shows that we can derive the output totgu from input /tot-?-qu/. This cannot be replicated in a rule-based theory. Since transitivity violations like this are not found, this is taken by Odden as evidence in favour of the derivational approach to phonology.

(186)

<table>
<thead>
<tr>
<th>/tot-?-qu/</th>
<th>*Ident[voi]</th>
<th>*C</th>
<th>*HetVoi</th>
<th>Max-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. totqu</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. *totqu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. todqu</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are some problematic aspects to Odden’s constructed example. First, notice that in the opaque case, the learner has to be able to posit a crucial ranking between two constraints, *HetVoi and *C, which do not conflict. Nevertheless, it is crucial for the optimization of the desired winner, (186b), that *C dominate *HetVoi. It is an interesting question whether learners are capable of learning orderings between constraints which do not in fact conflict or not. The problem is worsened by the fact that this ranking is only needed in the opaque case, to ensure the victory of the desired winner over the sympathetic (and fully faithful) candidate (186a). Languages like Kalaba might turn out to be unlearnable for this reason alone.

Suppose we admit the absence of languages like Kalaba as a counterexample to the predictions of Sympathy Theory and Virtual Phonology. We would have conceded Odden’s point, that the theory overgenerates, predicting a type of language which is unattested. We would then have to argue that the explanation for the gap lay elsewhere. We have addressed the point that some gaps may be due to acquisition: there are simply some regions of the variation space which are beyond the reach of the learning algorithm. We can apply similar reasoning to this case.
There may also be a diachronic reason for the gap. Halle (1962) proposed the rule-based theory of diachronic change. New rules are added to the end of the grammar (i.e., they are ordered last). Thus, rules are accumulated by a gradual process of 'sedimentation'. It is entirely possible to subscribe to the accretive nature of diachronic change without endorsing the view that rules are the appropriate way in which to encode the synchronic grammar. If new processes are appended to the end rather than 'intruded' into the body of the grammar, then, as a matter of diachronic course, languages exhibiting transitivity violations of the 'local ordering' variety will not occur. The reason is of course that if process \( P \) is appended to the grammar before process \( Q \), diachronically speaking, and process \( Q \) is appended before process \( R \), then it cannot be the case that rule \( R \) was added before rule \( P \). The gap would have a diachronic rather than an architectural explanation, but it is not surprising that languages like Kalaba are so conspicuously absent. If it is correct that diachronic considerations rule out languages of this type, then there should be no burden on OT, as a model of the *synchronic* grammar to rule them out.
Part II

Quantity Alternation in North Saami
Chapter 4

Grade Alternation: An Overview

This chapter offers a brief overview over of Grade Alternation and the other quantitative processes with which it interacts — Coda Maximization and Balance. Some basic descriptive terminology is introduced, and the place of the Saami gradation system within Balto-Finnic and Finno-Ugric in general is also discussed.

4.1 Grade Alternation in Saami

North Saami is one of the few languages in the world which evinces a three-way length opposition in consonants.\(^1\) In addition to the contrast between singleton and geminate, geminates themselves may be plain or 'overlong' (or 'hyperchronous') (Sammallahti 1977, 1984, 1998a,b; Magga 1984; Korhonen 1988a,b). This contrast is found in both alternating and non-alternating forms, exemplified in (187) and (188).

\(^1\) To my knowledge, Estonian is the only other language which evinces this property. See Hint (1997) as well as Bye (1997a,b) and references therein. Overlength as a property of vowels, on the other hand, is more frequently attested, and is found in Estonian, the Applecross dialect of Scots Gaelic Ternes (1973), and St. Lawrence Island Yupik (Hayes 1995: 241).
Let me briefly clarify the terminology and system of transcription used. Traditional labels have been adopted to refer to quantity distinctions in consonants. A non-moraic consonant is said to be in Quantity 1 (Q1); a monomoraic consonant is Q2; and a bimoraic consonant is Q3 (also 'overlong' or 'hyperchronous'). In general, the moraicity of a consonant, C, understood as the number of moras to which C is linked, is equal to its traditional quantity minus one.

Plain geminates are transcribed by doubling the symbol used for the corresponding singleton, e.g. /ss/ = short geminate (Q2) 's'. Geminates in Q3 (overlong geminates) are transcribed using the IPA length mark [], e.g. /ss/ = geminate 's' in Q3.

Similarly, clusters whose initial segment is in Q3 are written with the IPA length mark following the Q3 segment, e.g. [sk] = 'sk' in which only 's' is overlong. Clusters whose initial segment is in Q2 will be written without the length mark, e.g. [sk]. In sum, the IPA colon [.] is reserved exclusively for Q3 contexts.

The three contrastive quantities shown in (187) interact in an intricate system of Consonant Gradation, or Grade Alternation, which targets foot-medial consonants and consonant clusters (the consonant centre in traditional terminology). I will use Grade Alternation as a process neutral term to refer to the surface pattern of alternation evinced in surface forms. Consonant Gradation will be used to refer to the process itself.

The process underlying Grade Alternation may be seen as either Lenition (reduction in quantity) or Fortition (enhancement of quantity). The data is consistent with either interpretation, and for the time being I will not prejudge the issue by deciding in favour of one or the other.

Grade Alternation relates a pair of alternating terms known traditionally as the 'Strong' (=unlenited or fortitioned) and 'Weak' (=lenited or unfortitioned) grades. Idealizing somewhat, the Strong Grade is found preceding an open syllable, while the Weak Grade is found preceding a closed syllable resulting on suffixation.

For single segments in alternation the relationship between grade and quantity is transparent. If the Strong Grade form has a moraicity of \( n \), then the corresponding Weak Grade form has a moraicity of \( n - 1 \). This is shown in (189) and (190). (189) illustrates the alternation between monomoraic (Q2) and amoraic (Q1) consonants. Subscript 's' and 'w' will be used to designate the Strong and Weak Grades respectively.

(189) \( Q3_s \sim Q2_w \)

a. kaaffe 'coffee, NOM.SG' kaafes LOC.SG
b. ruoTTa 'Sweden' ruoTas LOC.SG
c. vivvaa 'son-in-law, NOM.SG' vivaah NOM.PL

(190) shows the corresponding alternation between a Q3 (bimoraic) and a Q2 (monomoraic) consonant centre.

(190) \( Q2_s \sim Q1_w \)
a. koli:lii 'gold, NOM.SG' kollies LOC.SG
4.1. GRADE ALTERNATION IN SAAMI

b. hear:raa 'lord, NOM.SG' hearraah NOM.PL

c. tuv:vaa 'dove, NOM.SG' tuvvaah NOM.PL

Note that while a single segment in Q3 can only represent the Strong Grade, and a single segment in Q1 can only represent the Weak Grade, a single segment in Q2 can represent either the Strong Grade or the Weak Grade. A single segment in Q2 represents the Weak Grade if it alternates with Q3, and the Strong Grade if it alternates with Q1. This morphological ambiguity will be designated by a superscript 's/w' where relevant.

Grade Alternation is fundamentally quantitative: on the evidence of simple segments, Grade Alternation consists in the manipulation of quantity alone. Grade Alternation also fundamentally involves a chain shift, i.e. while Q3 and Q2 alternate and Q2 and Q1 alternate, Q3 and Q1 do not alternate directly (but see below).

Consonant Gradation interacts with several other rules which also manipulate symbolic representations of quantity, and frequently this interaction is opaque. The two most important rules of this type are *Coda Maximization* and *Balance*.

We have already commented on the chain-shifting aspect of Grade Alternation. *Ceteris paribus*, this implies that there should be no surface paradigms in which Q3 and Q1 alternate directly. This prediction turns out to be false due to the refractory influence of *Balance*. The conditions for the application of Balance are entirely different from that of Consonant Gradation. While Grade Alternation is conditioned by the open vs. closed prosody of the following syllable, Balance is triggered by a long vowel in the following syllable. The combination of Grade Alternation and Balance yields surface alternations between Q3 and Q1, as exemplified in (191). Applied in the Strong Grade, Balance equates to Sammallahti’s ‘primary lengthening’ (Sammallahti 1998b: 49).

(191) $Q_3^s \sim Q_1^w$

a. jah:kkii ‘year, NOM.SG’ jakii ACC/GEN.SG
b. pal:luu ‘fear, NOM.SG’ paluu ACC/GEN.SG
c. joh:taa ‘travel, 3SG.PRES’ joo aan 1SG.PRES

Consonant clusters bring added complexities in their wake due to a combination of Balance and an additional rule, *Coda Maximization*, which lengthens the second member of a consonant cluster if the first member of the cluster is a sonorant.

Grade Alternation in clusters has varying results depending on the applicability of these rules. In the Strong Grade, it is always the first component which is overlong (e.g. $C_1:C_2$ or, if it’s a triliteral cluster $C_1:C_2:C_3$). If $C_1$ is an obstruent (most commonly, a sibilant $s$ or $\theta$, but also $\theta$), we find a two-way length contrast, and a simple $Q_3^s \sim Q_2^w$ alternation with $C_1:C_2$ as the Weak Grade.

(192) $C_1:C_2^s \sim C_1:C_2^w$

---

2 The one exception to this statement is a class of non-alternating bisyllabic stems in Q1 — all of these have the additional property that they either have a short mid vowel in the latus or a low vowel, e.g. peedo ‘gable’, spile ‘wife’s brother’, piro ‘devil’, saave ‘large tub’, pela ‘niece; nephew’. 
CHAPTER 4. GRADE ALTERTATION

Yet, when $C_1$ is a sonorant, we get other quantative effects. In this case we get at least Coda Maximization, with Balance applying if the ambient prosodic conditions require it. When the vowel of the second syllable is short, we get a surface alternation between $C_1C_2$ and $C_1C_2C_2$, with $C_2$ geminated. (193) illustrates the pattern.

\[(193) \ C_1:C_2 \sim C_1C_2:C_2\]

Balance applies preceding a long vowel. So, in a sonorant-initial cluster preceding a long vowel, we get the application of both Coda Maximization and Balance, to produce surface alternations of the form $C_1:C_2 \sim C_1C_2C_2$. In alternations of this sort, there is no net decrease in the moraicity of the cluster, simply a redistribution of moras to consonants: the Strong Grade and the Weak Grade are, to coin a term, isomorphic. (194) illustrates the isomorphic alternation.

\[(194) \ C_1:C_2 \sim C_1C_2:C_2\]

Applied in the Weak Grade, Balance equates with what Sammallahti dubs ‘secondary lengthening’, although I am treating ‘primary’ and ‘secondary lengthening’ as instantiations of one and the same Balance rule. Notice though that while ‘primary lengthening’ is associated exclusively with the Strong Grade, ‘secondary lengthening’ is associated exclusively with the Weak Grade.

4.2 Grade Alternation in other languages

Typologically speaking, Consonant Gradation of the kind displayed by the Saami languages is extremely infrequently attested amongst the languages of the world. This typological scarcity has immediate implications for an OT approach.

Consonant Gradation bears a passing resemblance to the Lenition systems of the Celtic languages, at least in the sense that these languages also operate with essentially ternary ‘grade’ systems.

For example, Soft Mutation, or Lenition, in Welsh (Pyatt 1997) involves a mapping between three series of consonants. In the relevant environment for Soft Mutation, input voiceless aspirated stops /pʰ tʰ kʰ/ are ‘lenited’ to voiceless unaspirated stops /p t k/. At the same time, input voiceless unaspirated stops are mutated to the corresponding voiced fricatives /v ð ɣ/ in the
The Celtic system shares with the Saami the property of ternarity and its chain-shifting aspect. Nevertheless, at the level of the driving phonological constraints, the systems are very different. Lenition in Celtic is driven, at least historically, by the post-vocalic context: deaspiration and spirantization are responses to constraints militating against, respectively, sequences of vowel + voiceless aspirated stop and vowel + voiceless unaspirated stop. The post-vocalic environment is an extremely common lenition context cross-linguistically, and so the universal status of the constraints involved is hardly in doubt. See Kirchner (1998) for an overview.

Consonant Gradation is driven by entirely different constraints. Moreover, there is a major difference, in terms of the frequency and universality of the two kinds of process. While consonant ‘reduction’ processes such as post-vocalic voicing and spirantization are cross-linguistically frequent, the kind of process instantiated by Consonant Gradation is apparently restricted to the Uralic language family, where, in addition to Saami, it is a characteristic of the Balto-Finnic languages (Barbera 1993), and the rather distantly related Samoyedic languages Nganasan (previously known as ‘Tavgi’) and Selkup (formerly known as ‘Ostyak’). A similar phenomenon is found in Eskimo (Ulving 1953).

Even within the Uralic family, though, its manifestations are diverse. Within this diversity, however, we may extract a single property which is common to all Consonant Gradation systems: consonant gradation is always associated (at least historically) with the open vs. closed prosody of the following syllable. (195) illustrates the Finnish pattern (Hammarberg 1974; Keyser and Kiparsky 1984; Karlsson 1987; Cathey 1992; Bye 1998b; McCartney 1998; Harrikari 1999b).

(195) Consonant Gradation in Finnish

| a. | seppä ‘smith’ | sepän ‘of a smith’ |
|    | matto ‘rug’ | maton ‘of a rug’ |
|    | kukka ‘flower’ | kukän ‘of a flower’ |
| b. | kylpy ‘bath’ | kylvyn ‘of a bath’ |
|    | mato ‘worm’ | madow ‘of a worm’ |
|    | puku ‘suit’ | puvun ‘of a suit’ |
|    | soldi ‘buckle’ | soljen ‘of a buckle’ |
|    | joki ‘river’ | joen ‘of a river’ |

In (195a), geminates undergo degemination preceding a closed syllable, while singletons (195b) undergo voicing and spirantization, or else delete.

Consonant Gradation in the Balto-Finnic languages differs from the Saami system on two counts. For one thing, Consonant Gradation in Balto-Finnic is restricted to stops, whereas in Saami, fricatives and sonorants also participate in the alternation. The same restriction holds in the Samoyedic language Nganasan, one of the only other Uralic languages with a comparable Grade Alternation system (Hajdu 1962: 50).

3 This is an oversimplification: although \( y \) can be reconstructed as the lenited reflex of /k/ in Old Welsh, all modern Welsh dialects lenite /k/ to \( \emptyset \).

4 Of the 650 common-Samoyedic root morphemes surveyed by Janhunen (1977), only about 150 are proto-Uralic in origin. Given this small proportion, Janhunen puts the primary split between Finno-Ugric and Samoyedic as having taken place at least six thousand years ago, and possibly much earlier (Janhunen 1992).
Second, the Balto-Finnic system is essentially mixed, combining quantitative reduction of geminates to singletons preceding a closed syllable, as in (195a), and voicing and/or spirantization of singleton stops in the same environment, as in (195b).

The Grade Alternation system of the Saami languages, on the other hand, is fundamentally and thoroughly quantitatively in nature. In fact, the only other attested purely quantitative system is that found in the Ket, Natskoo-Pumpokolek, and Tschulym dialects of Selkup (Hajdu 1962: 43).

Based on the scanty evidence available, there would seem to be an implicative relationship between the nature of the system, (quantitative as opposed to ‘mixed’) and the set of consonants subject to the process. There are no purely quantitative systems where Consonant Gradation does not also extend to sonorants (and fricatives) as well. Despite being based on so little evidence, there may be something to this claim. Since Grade Alternation developed independently in Samoyedic (Selkup) and Finnic (Saami), it is striking that both developed systems which combined these two properties completely independent of each other.

Nevertheless, given the scarcity of Consonant Gradation, and quantity-based Consonant Gradation systems in particular, it is very difficult to reach firm typological conclusions about them. It is unfortunate that the Selkup systems, which most closely approximate to that found in Saami, are so poorly documented and little understood, and so there is obviously much scope here for future research.

The infrequency of Consonant Gradation raises another problem, which is especially urgent from the perspective of OT’s avowed commitment to universality. The universality of constraints is usually justified in terms of their cross-linguistic utility. Unfortunately, the constraints driving Grade Alternation are poorly attested. Consonant gradation cannot be universal in the same sense as a process such as post-vocalic spirantization is universal, which is extraordinarily frequently attested amongst the world’s languages. Still, very similar Consonant Gradation systems have developed as much as three times independently of each other. Granted, this has occurred within the same language family, Uralic, but it may well be that Consonant Gradation represents the grammaticalization of a substratum of tempo-related phonetic tendencies and ‘habits of speech’, to use the phrase of Jackson (1953), which characterize the Uralic languages as a whole and have done so for thousands of years. Sammallahti suggests something similar, at least to explain the striking commonalities between Saamic and Balto-Finnic consonant gradation.

Proto-Saamic gradation arose through strengthening of the consonant center [. . . ] before open syllables in contradistinction to the Finnic gradation weakens stops (by lenition and shortening) before a closed syllable [sic!].

Both gradations, however, have the same preconditions in the trochee stress and quantity structures of Finno-Saamic.

(Sammallahti 1998b: 191)

While many constraints are grounded in functional drives which assert themselves in all linguistic human beings, others may be grounded in the more or less parochial paralinguistic conventions of a particular speech community, culture, or Sprachbunt. These paralinguistic conventions include things like tempo,
notions such as 'syllable-timed' and 'stress timed', and global 'articulatory settings' (cf. Laver 1994).

Some aspects of phonetic experience are thus universal; others may have a more contingent status. The constraints governing patterns of intervocalic stop voicing and final stop devoicing, discussed by Hayes, would appear to be grounded in experience of precisely such a universal nature. The constraints governing Consonant Gradation, on the other hand, may emerge only against the background of other paralinguistic givens.
Chapter 5

Segment Inventory

This chapter addresses the phonological inventory of Saami. Since the inventory is inseparably connected to metrical structure, we will address stress assignment first, in §5.1.

§5.2 presents a traditional system of terminology for describing the (metrical) positions within the word in Saami.

§5.3 briefly lays out the vowel inventory and sub-inventories associated particular metrical positions.

§5.4 describes the consonant inventory in greater detail.

The bulk of the material discussed here is drawn from sources on two closely related dialects of West Finnmark Saami: East Enontekiö (Nuorta Eanodat) and Kautokeino (Guovdageaidnu).\(^1\)

The East Enontekiö dialect is one of the most thoroughly documented varieties of Saami, and forms the basis of the descriptions in Sammallahti (1971, 1977, 1998a,b).

Magga (1984) contains a valuable fund of phonemicized data for Kautokeino (Guovdageaidnu), and Sammallahti (1984) is also a useful source for this dialect. The differences between the two dialects are sufficiently minor to warrant treating them as being of a piece. Differences will be explicitly noted where important.

Other sources which have been of particular use in collecting data have been Nielsen (1926), Bergslund (1976), Nickel (1994) and Kåven et al. (1995).

5.1 Stress assignment

Stress is assigned by constructing syllabic trochees from left to right (Bye 1997b). Main stress is on the initial syllable (196).

(196) **Stress assignment in Northern Saami**

a. (t̪e̊h.t̪i̊h) ‘to know’

\(^1\) Although Finnmark is a province of Norway, the term West Finnmark Saami includes varieties of Saami spoken on both the Norwegian and Finnish sides of the border. In addition to the dialects spoken in East Enontekiö (in Finland) and Kautokeino (in Norway), the western group also includes the varieties spoken, on the Norwegian side, in Alta (Álahaadjtu), and, on the Finnish side, Sodankylä (Soodegillii) (Sammallahti 1998b: 12).
Lexical words are subject to a minimal bisyllabic requirement. This minimality requirement is revoked in what may loosely be defined as the class of functional words, which includes both inflected items and particles. Hence, pronouns (such as *maa* 'I', *maat* 'what') and particles (e.g. *ka* 'I suppose') are exempt, as are some forms of the verb *leah* 'to be' and the inflected negative, e.g. *iin* 'I don’t', *eaj* 'he/she/it doesn’t'.

Saami abounds in phonological processes which make reference to metrical positions in one way or another. For example, the domain of Grade Alternation is the foot, restricted to foot-medial position.

The possibilities for contrast among segments is also a function of their prosodic position. For vowels, the relevant parameters include: (a) whether the vowel is in the stressed or unstressed syllable of the foot, (b) whether the vowel is parsed into a foot or not, and (c) whether that foot is the head of its PrWd or not. For example, the head foot of the word affords greater possibilities of vowel contrast than a non-head foot. The system of vowel contrasts in unstressed syllables is similarly impoverished relative to the corresponding system in stressed syllables.

Consonants are also sensitive to prosodic structure. The relevant parameters for consonants are: (a) position within the foot (initial, medial, or final), and (b) whether the consonant is parsed into the head foot of the PrWd or not. In particular, we find that the foot-medial position is strikingly richer in consonant contrasts than foot-peripheral positions.

### 5.2 Traditional terminology for positions within the word

The positions within the Saami word are traditionally described in terms of the system of nomenclature developed by Bergsland (1948), also described in Sammallahiti (1998b: 39). The system has some genuine descriptive utility, and so an overview is provided here.

The basic unit of rhythmic organization is taken by Bergsland and Sammallahiti to be the stress group, which is a sequence of syllables spanning the beginning of one stressed syllable to the beginning of another stressed syllable. It thus approximates to the notion foot, except that a stress group is not necessarily binary and may include the final syllable of an imparisyllabic word. A trisyllabic word is taken here to be footed as (σσ)σ, i.e. the final syllable is not part of the foot. On the conception of Bergsland and Sammallahiti, however, the final unfooted syllable in a domain does form part of the stress group.

The nucleus of the stressed syllable is known as the vowel centre. The vowel centre is rich in contrasts, and supports a system of monophthongs and diphthongs, both of which may appear long and short.

The consonant, or span of consonants, immediately following the vowel centre is known as the consonant centre. Because the consonant centre is the focus
of Grade Alternation, regular descriptive use will be made of this term.

The consonant centre is quite extraordinarily rich in the distinctions it permits. Both singleton consonants and clusters of up to three consonants are permitted in this position, and singletons and clusters alike may appear in any one of (up to) three distinct degrees of length.

The unstressed syllable following the consonant centre contains the latus, which refers to the unstressed nucleus of the foot. The set of vowels permitted in the latus is significantly reduced compared with that of the vowel centre: diphthongs are disallowed, and length and height are no longer contrastive in the non-low vowels.

In trisyllabic words, the third and final syllable is unstressed. The vowel in this syllable is referred to as the vowel margin.

The Bergsland system recognizes three further positions. They are included here for the sake of completeness, although no use will be made of them here.

The finis designates the final consonant in the stress group, and the initium refers to the initial consonant or consonant cluster in the stress group.

Finally, the consonant or consonant cluster at the juncture between two stress groups is known as the limes.

Each of these positions are associated with distributional restrictions on the segments that can occur in them, and we shall examine these more closely in the coming sections.

5.3 Vowels

In the first and main-stressed syllable of a word, the possibilities for vowel contrast are markedly greater than elsewhere. The vowel centre admits both long and short monophthongs, as shown in (197).

(197) Monophthongs in main-stressed σ (vowel centre)

\[
\begin{array}{cccc}
i & u & ii & uu \\
e & o & ee & oo \\
a & a & aa \\
\end{array}
\]

Two sets of rising diphthongs are also distinguished. The short rising diphthongs are contextually determined variants of the long, appearing preceding a Q3 consonant or cluster. When bimoraic, the head of the diphthong is the first vowel. When short, however, the first component becomes a vocalic onglide (transcribed here with a breve accent, e.g. V), and the second component is accorded the greater relative prominence. In Sammallähtti’s terminology, the bimoraic diphthongs are known as ‘initially stressed’, the short diphthongs as ‘finally stressed’ (Sammallahti 1998b: 40).

When stress is transcribed, an acute accent will be placed over the first component of the diphthong when bimoraic (e.g. úö), and over the second component when monomoraic (e.g. úö).²

(198) Diphthongs in main-stressed σ (vowel centre)

²The inventory abstracts away from the so-called allegro form, which permits some vowel types not permitted in the corresponding largo forms, e.g. the ‘finally stressed’ /iá/, and the short initially prominent rising diphthongs /iê uê öê aê/; see Bye (1998a) for analysis.
I will assume the difference is represented phonologically as in (199). In the short rising diphthong, the second component of the diphthong is the head. This is shown graphically by a vertical line linking the mora to the relevant vowel melody. The non-head is associated to the mora with a slanting line.

\[(199)\] Length contrast in rising diphthongs
\[
\begin{array}{ll}
\mu & \mu \\
\mu & \\
\mu & \\
\end{array}
\]

Latus vowels are drawn from a reduced set, shown in (200). Diphthongs are disallowed in this position. There is also an interaction between length and vowel height in the non-low vowels. In the short series, the feature specification \([-\text{low}]\) redundantly implies \([-\text{high}]\). Long \([-\text{low}]\) vowels in the latus, however, are redundantly \([+\text{high}]\).

\[(200)\] Vowels in latus
\[
\begin{array}{ll}
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
\mu & \\
5.4. CONSONANTS

The dental fricatives $\theta, \delta$ are not found word-initially, although $\delta$ may surface as a contextually determined variant of /d/ following a vowel.

Initially, clusters of a sibilant followed by a stop, nasal, or liquid are also permitted. These are shown in (203).

\[(203)\quad \text{Word-initial clusters}
\begin{align*}
\text{sp} & \quad \text{st} & \quad \text{sk} \\
\text{sm} & \quad \text{sn} & \quad \text{sp} & \quad \text{fp} \\
\text{sl} & \\
\text{sr} & \\
\text{jl}
\end{align*}\]

5.4.2 Consonants in marginal and final position

Marginally and finally, the sets of consonants are reduced. In marginal position, i.e. at the right boundary of the foot, the following consonants are possible.

\[(204)\quad \text{Marginal consonants}
\begin{align*}
p & \quad \text{ts} & \quad \text{tf} & \quad \text{k} \\
\text{v} & \quad \delta \\
\text{m} & \quad \text{n} \\
\text{l} & \\
\text{j} & \quad \text{h}
\end{align*}\]

In addition, the marginal position permits the following clusters.

\[(205)\quad \text{Marginal consonant clusters}
\begin{align*}
\text{st} & \quad \text{sk} \\
\text{j} \\
\text{lt} & \quad \text{lk} & \quad \text{lm} \\
\text{rt} & \quad \text{rs} \\
\text{jk} & \quad \text{jst} & \quad \text{jm} & \quad \text{jn} & \quad \text{j}\delta
\end{align*}\]

Finally, contrast is still more reduced: labials and dorsals are precluded from occurring finally in the word, and only coronals can occur.\(^3\)

\(^3\)The Pama-Nyungan language Lardil has an almost identical constraint. See Prince and Smolensky (1993: 97ff.) for discussion and references.
CHAPTER 5. SEGMENT INVENTORY

(206) Final consonants

\[
\begin{array}{c}
t \\
s \ f \\
n \\
l \ r \\
j \\
h \\
\end{array}
\]

In addition, the final position permits the following clusters.

(207) Final consonant clusters

\[
\begin{array}{c}
lh \\
rh \\
rs \\
js \ jn \ jh \\
\end{array}
\]

5.4.3 Consonants in foot-medial position

The foot-medial position is exceptionally rich in contrast. The foot-medial position licenses three contrastive degrees of length for singleton consonants, and up to three contrastive degrees of length for clusters.\(^4\)

The set of consonants which occur foot-medially in West Finmark Saami is shown in (208).

(208) Foot-medial consonants

\(^4\)The overview of the system of contrasts presented is incomplete. It omits reference entirely to the so-called allegro system. The allegro (Sammallahti 1977: 100-108) is essentially a form of quantitative truncation (Bye 1998a), which is implemented by various shortening processes. It is the obligatory realization of a number of morphological categories, including the first PrWd of a compound PrWd, the imperative form of the verb, and so on, and cannot be reduced to a matter of phonetic implementation. For example, long vowels in the latus are shortened, giving an allegro inventory /e o a/. In the head syllable of the foot, diphthongs are shortened to monomoraic, but initially prominent, rising diphthongs /i a u o a/ — these are not in the inventory of the corresponding non-truncated (‘largo’) forms. Another dimension of the quantitative truncation process is the optional syncopation of Q3 to Q2. Syncope in sonorant-initial clusters, however, is associated with the underapplication of certain processes which otherwise would be expected to apply in Q2, giving clusters which are unattested in the corresponding non-truncated forms. For example, aaktua ‘female reindeer’ undergoes quantitative truncation as the first PrWd of a compound PrWd to adita- and, optionally, syncopated to adita- (e.g. adita-piellua ‘bell on female reindeer’). In the latter form, however, we would naively expect Coda Maximization to apply, yet it does not, due to high-ranking constraints on identity between the truncated form and its base. The cluster /lt/ is not permissible under normal circumstances (in the phonology of base forms), yet it is permitted in the allegro. The consequence of this underapplication is a four-way length contrast in sonorant-initial clusters, once the phonology of the allegro is taken into account. I abstract away from the phonology of the allegro here.
Foot-medially, both single segments and clusters are permitted. In practice, it can be hard to tell whether a particular consonant centre should be classified as a single segment or a cluster, since certain expressions, which are featurally bisegmental, pattern in essential respects like single segments. I have based my primary taxonomic division into single segment vs. cluster on a single diagnostic. If the series has a Q1 alternant, I have classified the series as monosegmental. If the series lacks a Q1 alternant, I have classified the series with the clusters.

**Single segments**

The following series classify as monosegmental under the diagnostic introduced above.

- Fricatives and sonorants
- Preaspirated stops
- Prestopped nasals

**Fricatives and sonorants**

The fricatives and sonorants contrast in all three quantities, and this is true of the voiceless sonorants as well. The system is summarized in (209).

(209) Three-way length contrast in fricatives and sonorants

\[
\begin{array}{ccccccccccc}
Q_{3b} & ff & θθ & ss & jʃ & vʃ & ʃʃ & ʃʃ & vʃ & ʃʃ \\
Q_{2s/w} & ff & θθ & ss & jʃ & vʃ & ʃʃ & ʃʃ & vʃ & ʃʃ \\
Q_{1w} & f & θ & s & j & v & ʃ & ʃ & ʃ & ʃ \\
\end{array}
\]

Representative examples of the contrast are shown in (210).

(210)
CHAPTER 5. SEGMENT INVENTORY

nan'neh mannos junis
‘to fortify’ ‘may he go’ ‘lower spur of mountain’

passaas passaah pasii
‘strictly observant of the Sabbath’ ‘wash’ ‘he roasted’

noodu naa hiis
‘ball of thread’ ‘handle’ ‘it appeared’

poldluu tolaa hulil
‘round wooden cup’ ‘fire’ ‘in the south’

skerrii vierru suoro
‘dwarf-birch’ ‘habit’ ‘bull’s penis’

.tsu^h^aji tsu^h^laah tsu^h^aj
‘mumbler’ ‘to mumble’ ‘he mumbled’

sku^h^rii sku^h^raah sku^h^raj
‘snorer’ ‘to snore’ ‘he snored’

Note that the palatal nasal /n/ and palatal lateral /l/ have no Q1 congener. It is possible that this gap betrays an underlyingly bisegmentalism, i.e. /nj/ and /lj/.

Preaspirated stops

The preaspirated stops pattern in many respects like consonant clusters. Preaspirated stops may occur as syllable onsets, but not word-initially. Post-vocically and foot-medially, preaspirated stops are always in either Q2 or Q3, and alternate with plain unaspirated stops in Q1, except at the dental place of articulation, ht alternates with a voiced fricative ʔ in Q1. (211) summarizes the system of contrast.

(211) Preaspirated stop series

\[
\begin{array}{cccccc}
\text{Q3}\_w & h_p & h_t & h_t\_s & h_t\_j & h_k \\
\text{Q2}\_w & h_p & h_t & h_t\_s & h_t\_j & h_k \\
\text{Q1}\_w & p & ʔ & ts & tf & k \\
\end{array}
\]

(212) furnishes representative examples of the contrast.

(212)

to:h:pa stohpuj sapeh
‘sheath’ ‘living room-ILL-SG’ ‘ski’

pih:taa ihue iöö
‘bit’ ‘they requested’ ‘it appeared’

tohkeh tohko teko
‘to be fit’ ‘thither’ ‘as it were’
Prestopped nasals

The prestopped nasal series evinces a four-way contrast. Prestopping is generally accompanied by glottalization, except perhaps in the older generation of speakers (Berit-Anne Bals, p.c.), and there is rich variation in the realization of the prestop. In the Strong Grade, the prestop may be absent entirely, the prestopped and glottalized nasal being replaced by an overlong nasal accompanied by a medial glottal closure. This suggests the possibility of an alternative analysis of the series as simply glottalized nasals.

In the Weak Grade, the glottalized prestop may also be realized as a plain stop or a bare glottal stop (Summallahti 1998b: 51).

(213)  **Prestopped nasal series**

\[
\begin{array}{cccc}
Q_3 & b & m & \sim & p \\
Q_2_{s/w} & pm & \sim & p \\
Q_3_{s/w} & pm & \sim & p \\
Q_1 & m & n & \sim & p \\
\end{array}
\]

Representative examples of the contrast are shown in (214).

(214)

\[
\begin{array}{lll}
& \text{get beaten} & \text{miry hole-ILL.SG} & \text{I hit} \\
\text{tsaab} & \text{moh} & \text{op} & \text{maj} & \text{tsaap} & \text{maan} \\
\text{fid} & \text{ni} & \text{rut} & \text{noh} & \text{kaat} & \text{niis} \\
\text{tj ñc} & \text{ñaah} & \text{poc} & \text{ñaah} & \text{pocac} & \text{niis} \\
\text{Su} & \text{sm} & \text{SuSm} & \text{i} & \text{SuSmiih} & \text{heels} \\
\text{ku} & \text{ta} & \text{ku} & \text{staah} & \text{kuStaah} & \text{brushes} \\
\end{array}
\]

Clusters

The main taxonomic division in clusters is that between obstruent-initial and sonorant-initial clusters. Sonorant-initial clusters may further be divided according to the number of components by which they are comprised, giving a distinction between biliteral and triliteral. Obstruent-initial clusters are arguably always biliteral.

Obstruent-initial clusters

Clusters of obstruent+consonant have two quantities, as exemplified by the (near) minimal pairs in (216). Coda Maximization and Balance do not apply.

(215)  **Obstruent+C series**

\[
\begin{array}{cccc}
Q_3 & sp & st & sk & sm \\
Q_2_w & sp & st & sk & sm \\
\end{array}
\]

(216)  **Clusters of obs+C**

\[
\begin{array}{lll}
\text{ósKunuh} & \text{to believe} & \text{ósKunun} & \text{I believe} \\
\text{huspii} & \text{‘outlet} & \text{húsPiil} & \text{‘outlets} \\
\text{fuSmií} & \text{‘heel} & \text{fuSmiil} & \text{‘heels} \\
\text{kuSta} & \text{‘brush} & \text{kuStaah} & \text{‘brushes} \\
\end{array}
\]
Sonorant-initial clusters

Sonorant-initial clusters evince a three-way contrast in quantity. They obligatorily undergo Coda Maximization, and Balance applies under the appropriate conditions, yielding a three-way contrast. The three-way contrast in quantity is shown (218) for l-initial clusters only. Similar possibilities exist for clusters whose initial sonorant is drawn from the set /β δ j m n η r/, although there may be additional phonotactic restrictions on C2. According to Nickel (1994), the possible clusters are as in (217).

(217) Sonorant+C clusters

\[
\begin{array}{cccccccc}
\beta \beta s \beta j \beta k \beta s \beta j & \beta l \beta r \\
\delta p & \delta k & \delta j \\
jp & jt & jts & jk & js & jv & jl & jr \\
mp & ns & mj \\
nt & nts & ntj & \\
lp & lt & lts & ltj & lk & ls & lj & \\
rp & rt & rts & rtj & rk & rf & rs & rf & rv & rj
\end{array}
\]

(218) Quantitative contrast in sonorant-initial biliteral clusters

\[
\begin{align*}
Q_3^s & : l:p \sim l.l@.p \quad l:t \sim l.t \sim l:ts \sim l:tS \sim l:k \\
Q_2^w & : lpp \quad ltt \quad lts \quad ltj \quad llk \\
Q_3^w & : lpp \quad ltt \quad lts \quad ltj \quad llk
\end{align*}
\]

\[
\begin{align*}
Q_3^w & : l:f \sim l.l@.f \quad l:s \sim l.l@.s \quad l:j \sim l.l@.j \quad lv \sim l.lv \quad l:v
\end{align*}
\]

A salient property of sonorant-initial clusters is the presence of an epenthetic vowel in the Strong Grade. This epenthetic vowel is absent in homorganic clusters, however.

Representative examples of the contrast are shown in (219).

(219)

\[
\begin{align*}
\text{multuj} & & \text{palttii} & & \text{jal.ttiil} \\
\text{‘mould-ILL.sg’} & & \text{‘he frightened’} & & \text{‘bridges’} \\
\text{aar.ro.pah} & & \text{oorppef} & & \text{taa.rp.pujt} \\
\text{‘inherit’} & & \text{‘ring-finger’} & & \text{‘necessity-ACC.PL’} \\
\text{toaŋ.kko} & & \text{toaŋ.kkuu} & & \text{toaŋ.kkuj} \\
\text{‘it doesn’t} & & \text{‘it becomes} & & \text{‘it became} \\
\text{become inelastic’} & & \text{inelastic’} & & \text{inelastic’}
\end{align*}
\]

Clusters of voiced stop+voiceless stop

The voiced stops /b d j g/ pattern as [+sonorant] in Saami (cf. Rice 1993). Clusters of voiced stop+voiceless stop are therefore classified here alongside the other sonorant-initial clusters.
There is one major twist in the phonology of these clusters, though. In the Weak Grade, the voiced stop is deleted, leaving only the voiceless stop behind on the surface. Historically these clusters are derived from clusters of nasal+stop by a process of denasalization.

(220) *Prevoiced stop series*

<table>
<thead>
<tr>
<th>Q3v</th>
<th>b:p</th>
<th>d:t</th>
<th>d:ts</th>
<th>d:tf</th>
<th>p:t</th>
<th>q:k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q3w</td>
<td>p:t</td>
<td>t:t</td>
<td>t:ts</td>
<td>t:tf</td>
<td>c:k</td>
<td>c:k</td>
</tr>
</tbody>
</table>

Representative examples of the system of contrasts shown in (220) are shown in (221) below.

(221)

<table>
<thead>
<tr>
<th>tsub:poh</th>
<th>tfippiit</th>
<th>top:piil</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘frogs’</td>
<td>‘knee-ACC.PL’</td>
<td>‘in that direction’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>f:dd:tuu</th>
<th>t:jotaa</th>
<th>k:ott:uur</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘growth’</td>
<td>‘throat’</td>
<td>‘four legs of slaughtered animal’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>p:og:kaa</th>
<th>kokko</th>
<th>r:ikkii</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘it lies’</td>
<td>‘by which way’</td>
<td>‘rich’</td>
</tr>
<tr>
<td>short and fat</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Historically, clusters of voiced + voiceless stop derive from clusters of nasal + stop (Korhonen 1981), and have thus never had a Q1 alternant. This is also the case synchronically.

The distribution of voicing in this series is of considerable interest. Strikingly, the Weak Grade is completely devoiced. Yet, in the Strong Grade, these clusters are at least partially voiced. The behaviour of the voiced + voiceless stop clusters is schizophrenic: in the Strong Grade, they seem to pattern as true clusters, while in the Weak Grade (whether in Q2 or in Q3 as a result of Balance), they behave as monosegmental. Which of these conflicting behaviours is the more basic?

Sammallahti opts for the second approach, and analyses these clusters as monosegmental /b:b, d:d, . . . /, and so on. Phonetically, however, these segments are only partially voiced. For Sammallahti, this is a fact of phonetic implementation, since in the Q3 Strong Grade, voicing fades away due to the articulatory incompatibility of voicing stops with prolonged duration, giving phonetic [b:p, d:t, . . . ].

There are two serious problems with Sammallahti’s analysis. For one thing, the assumption of monosegmentalism renders the fact that the series lacks a Q1 alternant very puzzling, since all other monosegmental consonant centres alternate in all three lengths, making the ‘voiced stop’ series a curious anomaly. Second, the monosegmental hypothesis renders the distribution of voicing in the series inexplicable. The aerodynamic incompatibility of sustained voicing with long obstructive stop articulations is well known, and this incompatibility increases, the longer the stop. Granted this explains the ‘fading out’ of voicing in the Strong Grade, but it is still more than mildly surprising to find that voicing *disappears entirely* in the shorter Q2 geminates but is *retained* in the longer
Q3 geminates. As far as I know, there is no plausible functional reason for this distributional difference on the assumption that we are dealing with a monosegmental representation. It is for these reasons that Sammallahti’s ‘voiced stop’ series is reanalyzed here as a series of homorganic clusters of voiced+voiceless stop.5

Clusters of sonorant+nasal

Sonorant-initial clusters whose second component is a nasal (prestopped or plain) fall into two main types: (a) r-initial, and (b) the rest. In the r-initial series, the nasal always surfaces prestopped irrespective of the grade. In the remaining clusters of the sonorant+nasal type, the nasal only surfaces as prestopped in the Strong Grade. In the Weak Grade, it surfaces without preStopping.

(222) R+Nasal series

\[
\begin{array}{cccc}
Q3_r & r \cdot r \cdot p ^\text{im} & r \cdot r \cdot t ^\text{in} & r \cdot r \cdot c ^\text{ip} \\
Q2_r & r \cdot p ^\text{im} & r ^\text{in} & r ^\text{ip} \\
Q3_w & r \cdot p ^\text{in} & r ^\text{in} & r ^\text{ip} \\
\end{array}
\]

If the liquid is drawn from the non-rhotic set /β δ j l/, there is no stop epenthesis in the Weak Grade.

(223) Liquid+Nasal series

\[
\begin{array}{cccc}
Q3_l & l \cdot p ^\text{im} \sim l \cdot l \cdot p ^\text{im} & l \cdot r ^\text{in} \sim l \cdot l \cdot r ^\text{in} & l \cdot c ^\text{ip} \sim l \cdot l \cdot c ^\text{ip} \\
Q2_l & l ^\text{im} & l ^\text{in} & l ^\text{ip} \\
Q3_w & l ^\text{im} & l ^\text{in} & l ^\text{ip} \\
\end{array}
\]

(224) exemplifies the contrast.

(224)

aarr.\cdot r \cdot p ^\text{imu} sorr^\text{miit} kur\cdot p ^\text{maah}

‘mercy’ ‘accidental’ ‘gorm’

deh\text{-}\text{ACCP.L}.

kol\.\cdot l \cdot p ^\text{ima} kalm\text{maas} al\text{mmuus}

‘three’ ‘cold’ ‘public’

5.5 Triliteral clusters

Triliteral clusters fall into two types: (a) those whose medial component is /s/, and (b) those whose medial component is /h/. The possibilities for contrast in s-medial triliteral clusters are illustrated in (225).

5Phonetically, the voiced portion of these clusters is apparently sometimes accompanied by laryngealization, thus [b\cdot r, d\cdot t, . . . ] (Sammallahti 1977: 23). This may be understood as an instance of facilitative enhancement of voicing: laryngeal creak implies a slower rate of glottal pulsing. This has the effect of decreasing the rate of pressure build-up in the oral cavity, deferring the moment at which the transglottal pressure differential is reduced to zero, inhibiting voicing.
5.5. **Triliteral Clusters**

(225) **Triliteral clusters with medial /s/**

\[
\begin{align*}
Q^s & : jsk & jst & sk & nsk & bst & rsk & rsst & bsk \\
Q^w & : jsk & jst & mjk & nsk & rst & bst & \\
Q^w & : jsk & jst & mjk & nsk & rst & bst & \\
\end{align*}
\]

(226) exemplifies the contrast.

(226)

raamskii raamskas raamskiil
'a mess' 'totally disarranged' 'totally disarranged'
maajstii maaajstii maaajstaan
'to taste' 'he tasted' 'I taste'

Similar possibilities for contrast exist for h-medial clusters. Note that, in cases where \(C_1\) and \(C_3\) are homorganic, \(/h/\) is realized as an aspirated variant of \(C_1\). Similar possibilities exist for clusters whose initial sonorant is drawn from the set \(/B j m n N r/\). According to Nickel (1994), the possible clusters are as in (227).

(227) **H-medial clusters**

\[
\begin{align*}
\betaht & \betahts & \betahtf & \betah & \\
\jhp & \jht & \jhts & \jhtf & \jhk & \jhn & \jhl \\
\mmp & \nnt & \nnts & \nntf & \ynk & \\
\lhp & \llt & \llts & \lltf & \llhk & \\
\rhp & \rht & \rhts & \rhtf & \rhl & \\
\end{align*}
\]

(228) **Triliteral clusters with medial /h/**

\[
\begin{align*}
Q^s & : lhp ~ llo.hp & llt & llts & lltf & lhk & llo.lhk \\
Q^w & : lhp & llt & llts & lltf & lhk & \\
Q^w & : lhp & llt & llts & lltf & lhk & \\
\end{align*}
\]

(229) shows how the three-way length contrast works out for l-initial h-medial clusters.

(229)

limm:piii kummp:peh limmnpajt
'he made lumps' 'wolves' 'lumps-ACC.PL'

mur ra.likuu nirkhol skarhkaan

'winter fog' 'poison-NOM.PL' 'I make rattling sounds'

There is one further variety of cluster, the k-initial cluster, which doesn’t fall neatly into the taxonomy proposed here. In the Strong Grade, these are arguably triliteral, with a medial /h/ or /s/. The naive expectation is that they pattern with the obstruent-initial clusters. However, in the Weak Grade, /k/ undergoes Voicing and Spirantization with a shift in place of articulation to the bilabial approximant [β] or [βh] ([ββ]) resulting in a sonorant-initial cluster. A three-way contrast is the result.
5.6 Conclusions

The importance of rhythmic structure in Saami phonology cannot be overemphasized. Metrical positions (main-stressed syllable, stressed syllable, unstressed footed syllable, unfooted syllable, and so on) play a major role in shaping the inventory.

The foot-medial position, or consonant centre, is the prime locus for consonant contrast in the language. In particular, there is a rich system of quantity contrast, due to the interaction of several processes, Grade Alternation, Coda Maximization, and Balance.

Both single segments and clusters may appear in the foot-medial position in one of up to three distinct quantity degrees.

A taxonomy of consonants and consonant clusters was proposed. Clusters may be classified into obstruent-initial and sonorant-initial. Obstruent-initial clusters contrast in only two quantities: Q3 is Strong Grade and Q2 is Weak Grade.

For sonorant-initial clusters, however, this pristine relationship between grade and quantity is refracted. We cannot equate Q3 with the Strong Grade, because sonorant-initial clusters are subject to additional rules which manipulate quantity. Coda Maximization feeds a rule of Balance, which lengthens a consonant centre to Q3 preceding a long vowel. The Weak Grade may thus be realized as Q2 or Q3, depending on the ambient phonological conditions.
Chapter 6

Historical origins of gradation

In this chapter we will address the diachronic origin of Consonant Gradation, endorsing the essential correctness (with some modifications) of the view espoused by (Sammallahti 1998b), that Consonant Gradation in Saami is a separate Proto-Saamic development with origins in fortition.

Sammallahti's view represents an audacious break with earlier thinking. Gradation has generally been held to be a common Finno-Saamic development, and to have arisen in Proto-Finno-Saamic, the ancestor language of both Saami and the Balto-Finnic languages (Finnish, Estonian, and so on). This line of thought is represented by Wildlund (1896, 1915), Itkonen (1969), Décsy (1965), Décsy (1969a,b), Korhonen (1981, 1988a,b), Tauli (1954), Pikamäe (1957), Steinitz (1952), Bergsland (1945), and Ravila (1960).¹

In Proto-Finno-Saamic, there was an opposition between singleton and geminate obstruent stops. In the traditional finno-centric account, gradation applied originally to stops only, although subsequently it spread, in Saami, to include the other non-stop consonants by "analogy". Gradation is thus traditionally seen as a lenition or shortening process applying before a closed unstressed syllable (Korhonen 1988b: 275). Before a closed unstressed syllable, both the singletons and the geminates were shortened. Subsequently, the shortened reflexes of the singleton stops became either voiced stops (Ravila 1900) or voiced spirants (Décsy 1965; Hakulinen 1961; Pikamäe 1957). Setälä (1896), Tauli (1954), and Gordon (1997) are dissenters from this received view. Adopting what might be dubbed a 'lappocentric' view, they argue that gradation originally affected all consonants in Proto-Finno-Saamic. Subsequently, Balto-Finnic lost gradation in all consonants but the obstruent stops, while most of the Saami dialects retained it. Whether or not consonant gradation originally only affected the stops and was subsequently generalized to other segments, or originally affected all segments and subsequently restricted to the stops, all of these researchers share the belief that Balto-Finnic and Saami consonant gradation spring from a common historical source in the Finno-Saamic ancestor language.

A third view is represented by Sammallahti (1998b), who sees gradation in Saami as a separate Proto-Saamic development. This is quite a daring hy-

¹The existence of similar consonant gradation phenomena in Samoyed has led some earlier scholars (Setälä 1896, 1912a,b) to argue that gradation was a proto-Uralic feature. This view no longer has any currency. The Samoyed languages developed gradation later independently of Balto-Finnic influence.
CHAPTER 6. HISTORICAL ORIGINS OF GRADATION

hypothesis, given the fact that Consonant Gradation is found in the Balto-Finnic languages as well. Naively, one would expect the two kinds of Consonant Gradation to have a common origin. For Sammallahti, though, the Saamic and Balto-Finnic developments have to be diachronically unconnected, at least at the level of phonologization. It is nevertheless probable that Saamic and Balto-Finnic Consonant Gradation have a common source at some level. It is probable that Saamic and Balto-Finnic independently phonologized durational tendencies which had obtained in the Finno-Saamic languages generally and for quite some time, and which probably still do. A parallel with the development of Consonant Mutation in the Celtic languages is apposite here. Lenition developed apparently independently in the Brythonic and the Goidelic branches of Celtic. While in Brythonic, lenition took the form of intervocalic voicing or deaspiration of aspirated voiceless stops, in Goidelic, it took the form of intervocalic spirantization. Obviously, spirantization and deaspiration have nothing in common, and therefore, Brythonic lenition and Goidelic lenition cannot spring from the same source. Having said this, the environments for lenition in Brythonic and Goidelic are identical (intervocalic), and the parallel development of lenition in both branches of Celtic must have its roots in phonetic tendencies which characterized the Celtic languages as a whole.

6.1 Gradation as Historical Fortition

According to Sammallahti, Consonant Gradation arose in Proto-Saamic through the fortition of a foot-medial consonant or consonant cluster when the final (unstressed) syllable of the foot was open (Sammallahti 1998b; Korhonen 1988b: 191). Fortition can be broken down into two, possibly three, sub-processes.

- Singletons (C) undergo fortition to geminate (CC) preceding an open syllable (Singleton Fortition).
- Geminates (CC) undergo fortition to overlong geminates (C:C) preceding an open syllable (Geminate Fortition), and
- Clusters (C₁C₂) undergo fortition to overlong clusters (C₁:C₂) preceding an open syllable (Cluster Fortition).

For Sammallahti, though, all of these processes are in fact one coherent process. We shall see that there are some reasons to doubt their coherence, however, similar as they appear to be. One of these reasons is theoretical: in OT, since processes are driven by the requirement to satisfy specific output targets, the homogeneity of the process does not carry over to the descriptive level of the constraint. As output targets, geminates and overlong geminates are not the same, and this immediately casts doubt on the idea that Singleton Fortition and Geminate Fortition were necessarily implemented simultaneously as a single process as Sammallahti claims. The other reason is a typological one. While there seems to be no typological evidence bearing on the distinction between Singleton and Geminate Fortition, there would appear to be typological evidence on the distinction between Geminate Fortition and Cluster Fortition. It merely happens that in North Saami, clusters undergo a type of Fortition which is procedurally very similar to Geminate Fortition — the insertion of a
mora on $C_1$, to give $C_1C_2$. In other dialects of Saami, namely Pite and Ume Saami, clusters pattern differently. Instead of mora insertion, $C_2$ undergoes Coda Maximization, giving $C_1C_2C_2$, generating a system in which clusters, arguably, pattern more on a par with the singletons. Examples of the Ume alternation in clusters are shown in (232).

(232)  Grade Alternation in Clusters in Ume Saami

a. $kìjöø$-$kìjöuv$  ‘spring, Nom.sg~Acc.sg’
b. $ak$ta-$ak$ton  ‘one, Nom.sg~Acc.sg’
c. muj$h$tìt-$muj$h$tìaw  ‘remember, Inf~1sg.Pres’

In North Saami, these alternations become respectively: $kìdåa$-$kìdåa$, $ok$:ta-$oBhtaa$, muj$h$tìh$tih$~muj$h$ttìaan.

Representative examples of Singleton Fortition are given in (233) below. Sammallahti reconstructs the first (prephonemic) stage of fortition as lengthening to a ‘half-long’ stop. I will argue presently that a distinction between half-long and long (i.e. truly geminate) is not warranted. Nonetheless, following Sammallahti, the fortitioned output of input singletons is designated with the half-length sign [\] here.

(233)  Fortition of singletons in Proto-Saamic

$*C > *C/V_\_V$  *mene$-$ > *mene$-$  ‘go’
*kae$e$ > *kæe$e$  ‘hand’
*kala > *køol$a$  ‘fish’

A cursory glance at the data might suggest the driving constraint was a requirement that stressed syllables be bimoraic, as found in the Scandinavian languages (Prokorsch 1939; Árnason 1980; Riad 1992). This hypothesis cannot be correct, since it predicts that Singleton Fortition would not apply to a syllable containing a long vowel, the reasoning being that, if the syllable is already long, Singleton Fortition would be gratuitous. However, Singleton Fortition does apply in syllables containing long vowels, to give a hypercharacterized CVVC-syllable, sinking the Stress-to-Weight hypothesis. Nevertheless, the fortition occurred in a stressed syllable. The relevant constraint driving Singleton Fortition, SingFort, effectively demands that the onset of the second syllable of the foot be the coda of the first.

In what Sammallahti sees as a parallel development, geminates and clusters were lengthened in the same environment, as shown in (234).

(234)  Fortition of geminates and clusters in Proto-Saamic

$*CC > *C/C/V_\_V$  *appe > *oøpp$e$  ‘father-in-law’
*aj$ma$ > *aj$ma$  ‘needle’
*kølpe > *køl$pe$  ‘shield’

Because the outputs of Singleton Fortition and Geminant Fortition are structurally distinct, it entails that they cannot be driven by the same constraint, i.e. they cannot be coherent. For the changes in (234), the driving constraint
cannot be SingF ort. Since the medial consonants in the base are geminate, they are already codas in $\sigma_1$, thereby satisfying SingF ort. Singleton F ortion can only apply vacuously.

Lenition in Celtic is demonstrably non-coherent, and the analogy lends some support to seeing F ortion in Saami as non-coherent as well. Lenition in Brythonic can be shown to have developed in two distinct waves (Thomas 1990; Sims-Williams 1990; Pyatt 1997). Jackson (1967) also shows that some dialects of Breton are extending lenition to fricatives, in yet another wave of intervocalic lenition hundreds of years after consonant mutation was first grammaticalized.

It is perhaps tempting to see the reason for the lengthening of the geminates in (234) as grounded in the preservation of the lexical contrast between geminate and singleton. It is true that the failure of geminates to shift to overlong would have resulted in massive neutralization of quantity contrast. Nevertheless, it is generally agreed that phonological constraints cannot refer globally to non-phonological states of affairs. This consideration requires us to posit an independent phonological constraint to drive fortition in geminates.

If Singleton F ortion and Geminate F ortion are indeed responses to distinct constraints, then Sammallahiti’s assumption that they applied simultaneously is moot. It is more likely that singletons and geminates, and clusters too, underwent F ortion at different times. This raises the question whether it is Singleton F ortion or Geminate F ortion that is diachronically prior. Put differently, did the present pattern arise as a result of a push chain (singletons underwent fortition first, followed by the geminates), or a drag chain (geminates underwent fortition first, followed by the singletons)?

At least two considerations would argue for the correctness of the drag chain account. If singletons underwent fortition before geminates, the result would have in the first place been neutralization of the singleton/geminate distinction in the open-syllable environment. Given that in the modern language, the underlying contrast is not neutralized, it is more natural to assume that the geminates underwent fortition at a stage prior to the singletons. Another consideration which favours this conclusion is that genuine push chains such as this are arguably unattested anyway (King 1969). Therefore I propose that open-syllable fortition applied in two stages: Geminate F ortition applied first, followed by a subsequent stage during which singletons underwent an analogous development in the same environment. (235) summarizes the proposed development. Stage I represents the system prior to the introduction of F ortion. At Stage II, Geminate F ortion applies. At Stage III, Singleton F ortion applies, completing the drag chain.

(235) Stages in the Development of F ortion

---

2The case is reminiscent of the three-way length contrast in vowels in iambic languages such as St. Lawrence Island Yupik, Choctaw, and Chickasaw. See Hayes (1995: 209ff., 241) and references cited therein.

3For these concepts, see Martinet (1952, 1955) and Bynon (1977)
6.2. HISTORICAL MORPHOLOGIZATION

| Stage | V|FtCC| CC C | V|FtCC| CC C | V|FtCC| CC C |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Stage I | V|Ft | CC C | V|Ft | CC C | V|Ft | CC C |
| Stage II | V|Ft | C:C | CC C | V|Ft | CC C | V|Ft | CC C |
| Stage III | V|Ft | C:C | CC C | V|Ft | CC C | V|Ft | CC C |

At Stage II, the system which resulted from the initial round of Geminate Fortition, Q3 would have alternated with Q2, but Q1 would not have alternated with Q2. Preceding a closed syllable, we would have had a relatively unmarked contrast between geminate and singleton. Preceding an open syllable, on the other hand, we would have had a typologically marked contrast between an over-long geminate and a singleton. Such a contrast is marked, not only because it involves a marked phonological configuration (attested examples of overlength are not exactly thick on the ground), but because there is a gap at Q2 preceding open syllables. The introduction of a rule geminating singletons in this environment eliminated this gap.

It is an interesting question to what extent the optimization of contrast is mediated directly by constraints on contrast, or arises by some other mechanism which optimizes global dependencies between constraint rankings. Flemming (1995) has proposed incorporating constraints on spacing and contrast directly into the grammar, implying that there are constraints driving optimization at a global system level. The issue is a profoundly interesting one, but I will adopt the null hypothesis that the information in which constraints trade is strictly local, and that global dependencies of this kind arise through the operation of some other extragrammatical mechanism.

By Stage III, the resulting system could still be understood in terms of an underlying singleton-geminate contrast with a contrast between Strong and Weak Grade overlaid.

6.2 The Historical Morphologization of Grade Alternation

The Strong (fortitioned) Grade appeared preceding an open syllable. Fortition was blocked when the foot-final syllable was closed, giving what is now the Weak Grade. This gave transparent alternations such as the following in (236).

(236) Triggering and blocking of fortition in Proto-Saamic

a. kuollee ‘fish, nom.sg’ knoleen gen.sg
b. menejem ‘go, 1sg.pret’ menenm 1sg.pres
menenme past part

Subsequently, in the transition from Proto-West-Saami to Proto-North-Saami, some of the purely phonological conditioning environments were lost, but the historical distinction between open and closed syllable was retained *opaquely* in the form of the opposition between Strong and Weak Grade (Sammallahti
On the surface, the relation between morphological grade (Strong vs. Weak) and the prosody of the following syllable (open vs. closed) is frequently contrary to what one would expect on phonological grounds alone. Thus, in some instances, we find the Weak Grade of a consonant or consonant cluster preceding an open syllable, or the Strong Grade preceding a closed syllable. The historical developments responsible for these perturbations fall into three types:

1. Final Consonant Apocope lead to the morphologization of the Weak Grade preceding an open syllable,
2. Final Vowel Apocope lead to the morphologization of the Strong Grade preceding a closed syllable, and
3. Word-internal vowel deletion (Contraction) in response to vowel hiatus lead to the morphologization of the Strong Grade preceding a closed syllable.

Let us examine and exemplify each of these developments in turn. The deletion of certain consonants opacified or morphologized some occurrences of the Weak Grade. This happened, notably, in the accusative and genitive singular forms of underlyingly bisyllabic nominal stems due to the loss of accusative singular *-m and genitive singular *-n.

(237) Morphologization of Weak Grade through Apocope

Proto-West-Saami *koahtte  *koahteen
Proto-North-Saami
Final C-loss *koahtte  *koahtee
Modern North Saami [koahti] [koahti]

Similar opaque triggers of the Weak Grade include the indicative connective present suffix -C (<PS *-n), the conditional suffix -f (<PS *-htf) and the potential suffix -tf (<PS *-ntf).

Final vowel apocope occurred around the same period as the loss of the accusative and genitive markers, and opacified or morphologized some occurrences of the Strong Grade. The loss of the final vowel on the essive marker -n (Proto-West-Saami *-nee), for example, resulted in the essive opaque conditioning the Strong Grade.

(238) Morphologization of Strong Grade by Apocope

Proto-West-Saami *koahtte  *koahtteenee
Proto-North-Saami
Final V-loss *koahtte  *koahtteene
Modern North Saami [koahti] [koahti]

Similar opaque triggers of the Strong Grade include the infinitive suffix -h (<PS *-tee-k), and the past participle suffix -n (<PS *-nee).

The third and final source of opacity in Grade Alternation is Contraction, which affected trisyllabic forms. As in the case of Final Vowel Apocope, the
core operation is the deletion of a vowel, resulting in a closed syllable. The term ‘contraction’, as used traditionally, generally refers to a complex process consisting of three parts:

- elision of an intervocalic consonant /s/ or /j/ resulting in vowel hiatus,
- resolution of the vowel hiatus, usually through deletion of the second component of the ‘contracted’ vowel, but also by merger into a long vowel, and,
- in some instances, compensatory lengthening for the deleted vowel through lengthening of a stem-internal geminate to Q3.

Sammallahti (1998b: 45) reconstructs the ‘contracted vowels’ resulting from elision as overlong, although comments that their exact nature is not known. It is likely that they were in fact not single nuclei as suggested by Sammallahti, but that the immediate consequence of the elision of the intervocalic consonant was a situation of vowel hiatus. The vowels then assimilated in quality across the syllable boundary, and, at a subsequent stage, the ‘contracted’ vowels were shortened, i.e. the second nucleus was deleted, often with compensatory lengthening of a geminate in the consonant centre.

(239) represents my own adaptation of Sammallahti’s reconstruction of the development of contracted forms in North Saami. The example leads us through the development of the reconstructed Finno-Saamic form *saloi-j-i-t ‘islands’ to its current form sul:loh in modern West Finnmark Saami.

(239) Morphologization of Strong Grade by Contraction

<table>
<thead>
<tr>
<th>Finno-Saamic</th>
<th>*saloi-jt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proto-Saami</td>
<td></td>
</tr>
<tr>
<td>Fortition</td>
<td>*sool:lu.jek</td>
</tr>
<tr>
<td>Elision</td>
<td>*sool:lu.ek</td>
</tr>
<tr>
<td>V-Assimilation</td>
<td>*sool:lu.uk</td>
</tr>
<tr>
<td>Proto-West-Saami</td>
<td></td>
</tr>
<tr>
<td>Hiatus Resolution</td>
<td></td>
</tr>
<tr>
<td>Compensatory Lengthening</td>
<td>*sool:lu.uk</td>
</tr>
<tr>
<td>Proto-North-Saami</td>
<td></td>
</tr>
<tr>
<td>Monophthongization</td>
<td>*sool:lu.uk</td>
</tr>
<tr>
<td>Pre-Q3 Shortening</td>
<td></td>
</tr>
<tr>
<td>Latic Lowering</td>
<td></td>
</tr>
<tr>
<td>Debuccalization</td>
<td>sul:loh</td>
</tr>
<tr>
<td>Modern West Finnmark Saami</td>
<td>sul:loh</td>
</tr>
</tbody>
</table>

Contracted vowels thus blocked Consonant Gradation from applying for the reason that they spanned a syllable boundary in a vowel hiatus configuration. In post-contraction Proto-Saami, the syllable-closing suffixes which resulted from the contraction process thus failed to condition lenition on the consonant centre of the stem. This is explained if we analyze the form as trisyllabic. Thus, older Proto-Saamic *sool:lu.jek underwent Elision to *sool:lu.uk. Since the plural formative -k does not close the second syllable of the initial foot at this stage, the structural description of lenition is not met, and hence we do not get reduction to ungrammatical *sool:lu.uk.
Following the proto-Saamic period, contracted vowels underwent shortening to resolve the marked hiatus configuration. This development in turn precipitated compensatory lengthening on a preceding geminate under certain conditions. The final mora of the historically contracted vowel \( u.u \) has delinked and reassigned to the geminate occupying the consonant centre of the stem, yielding a Q3 geminate, along with attendant shortening of the vowel in the stressed syllable.

Summing up the discussion so far, I have adopted the fortition-based account of the origin of consonant gradation of Sammallahti (1998b), with some modifications of interpretation. There is one final aspect to Sammallahti's diachronic account which I will not be endorsing here. This is Sammallahti's claim, carried over from earlier historical work (Korhonen 1981), that the introduction of Consonant Gradation gave rise initially to a four-way length contrast.

### 6.3 Four degrees of quantity?

Sammallahti (1998b: 3191) reconstructs Proto-Saami with four degrees of length. The four-way system was a direct result of fortion, which, for Sammallahti involved the addition of a 'subglottal pulse', a notion which approximates to our minor syllable. According to this view, the original two-way opposition of geminate (CC) vs. singleton (C) underwent a mitosis on the introduction of Fortition. Post-Fortition Proto-Saami distinguished both long and short singletons, transcribed in FU notation as Ĉ and C respectively. (The lengthened variants are transcribed with a grave symbol). There was, according to Sammallahti, a similar distinction in geminates, which distinguished a short (CC) from a relatively long variant (ĈC). Importantly, at this stage, the Singleton Strong Grade (Ĉ) and the Geminate Weak Grade (CC) were distinct.

Putting the representational issues aside, this looks like a regular case of allomorphy. We could in principle adopt a temporary representational solution which distinguishes the features \([\pm\text{geminate}]\) and \([\pm\text{long}]\). Before the Fortition rule was adopted, both geminates and singletons were \([-\text{long}]\). Fortition could then be understood as involving a feature change to \([+\text{long}]\). (240) summarizes Sammallahti's conception of the four-way length contrast in terms of these two binary features.

\[
\begin{array}{c|c|c}
\text{[\pm\text{geminate}]} & \text{[\pm\text{long}]} \\
\hline
C & - & - \\
Ĉ & - & + \\
CC & + & - \\
ĈC & + & + \\
\end{array}
\]

It is not difficult to see how Grade Alternation might be modeled on these representational assumptions as a case of classical allomorphy, although we won't go into the details of a full analysis here.

Appealing as this might seem, it makes unprecedented use of a novel feature \([\pm\text{long}]\). In the absence of cross-linguistic support of such a feature, this detracts considerably from any \textit{prima facie} appeal the analysis might have. Moreover, as I will show, there is no robust evidence for assuming a four-way contrast.
6.3. FOUR DEGREES OF QUANTITY?

Nevertheless, Sammallahti claims there is evidence for such a four-way length contrast historically, and that during the transition from Common Western Saami to Common North-Western Saami, the distinction between CC and Č was lost through merger, except, apparently, in the North Tärna dialect of Ume Saami, which retained the contrast until the beginning of the 20th Century (Sammallahti 1998b: 193). The four-way contrast is real enough: it is robustly attested in preaspirated stops and prestopped nasals. What I wish to contest here is that this four-way contrast should be taken as a contrast in duration, or reflective of a duration contrast having existed historically.

From what I have been able to gather from the limited material available to me, there are, in Tärna, phonological differences between the Geminate Weak Grade and the Singleton Strong Grade in the preaspirated stop series and the prestopped nasal series. However, these do not amount to evidence of a four-way phonological distinction in length either synchronically or diachronically. For example, the Strong Grade of singleton nasals failed to undergo prestopping, unlike the Weak Grade of geminate nasals. Consider the Tärna examples in (241) from Sammallahti (1998b: 195).

(241) a. liemme ~ liemen `broth, NOM.SG~GEN.SG'
    b. eaitnie ~ ietnien `mother, NOM.SG~GEN.SG'

Sammallahti takes the difference in behaviour between the Singleton Strong Grade of liemme and the Geminate Weak Grade of ietnien could be taken as a reflex of a phonological distinction in duration in the historical antecedents of these forms. However, there is really nothing which forces this interpretation. The difference is rather attributed to the relative chronology of two processes, Prestopping and Singleton Fortition. According historical precedence to Prestopping over Singleton Fortition is sufficient to account for the observed four-way contrast. (242) shows how this works. Note that (242) isn’t to be taken as a serious attempt at accurate reconstruction — the forms serving as the starting point may equally be taken as the inputs to a battery of synchronic rules. Nevertheless, the relative ordering of the processes is intended to represent historical truth.

(242)

<table>
<thead>
<tr>
<th></th>
<th>liemme</th>
<th>liemen</th>
<th>eannie</th>
<th>ietnien</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geminate Fortition</td>
<td>n/a</td>
<td>n/a</td>
<td>eannie</td>
<td>n/a</td>
</tr>
<tr>
<td>Prestopping</td>
<td>n/a</td>
<td>n/a</td>
<td>eatnie</td>
<td>eatnie</td>
</tr>
<tr>
<td>Singleton Fortition</td>
<td>liemme</td>
<td>n/a</td>
<td>n/a</td>
<td>ietnien</td>
</tr>
<tr>
<td>Other processes</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>ietnien</td>
</tr>
<tr>
<td>Outcome</td>
<td>liemme</td>
<td>liemen</td>
<td>eatnie</td>
<td>ietnien</td>
</tr>
</tbody>
</table>

The preaspirated stop series evinces something similar. Apparently, the four-way contrast is quite robustly attested in the material collected by Moosberg (1920), although I have only had access to this data through the limited discussion of it in Bergsland (1973: 51f.). A fragment of the relevant data is shown in

---

4 The access I have had to the relevant data is very slight. All of the data comes from an unpublished manuscript Moosberg (1920), which personally I have been unable to access. My understanding of the facts is thus restricted to what I have been able to glean from Sammallahti (1998b) and Bergsländ (1973).
The leftmost column is an IPA rendering of Moosberg's narrow phonetic transcription. The next column shows Bergsland's phonemicization.

(243)

\[
\begin{array}{l}
[ka\text{htu}\text{u}u] /gah\text{tuo}/ '\text{cat, nom.sg}' \\
[ka\text{htu}\text{u}\text{o}n] /gah\text{tuo}/ '\text{cat-gen.sg}' \\
[ka\text{htu}\text{t}\text{t}] /gah\text{tu}\text{ot}/ 'to be absent' \\
[ka\text{tu}\text{u}\text{op}] /\text{gatuub}/ 'I am absent'
\end{array}
\]

There are some questions surrounding the implementation of duration in /t/, but I venture to suggest here that it is more likely that the Tärna system represents a sandwiching effect involving two applications of a process of preaspiration separated from each other in time, and a single application of a delinking process. Again, (244) is not offered here as an accurate reconstruction. It merely serves to illustrate how the observed four-way length contrast might arise as the result of the interaction of independently attested and motivated processes.

(244)

<table>
<thead>
<tr>
<th></th>
<th>kaattuu</th>
<th>kaattuu+n</th>
<th>kaatuun+tV</th>
<th>kaatuun+p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geminate Fortition</td>
<td>kaattuu</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Preaspiration I</td>
<td>kaahtuu</td>
<td>kaahtuu</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Delinking</td>
<td>kaahtuu</td>
<td>kaahtuu</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Singleton Fortition</td>
<td>n/a</td>
<td>n/a</td>
<td>kaattuunV</td>
<td>n/a</td>
</tr>
<tr>
<td>Preaspiration II</td>
<td>n/a</td>
<td>n/a</td>
<td>kaahtuunV</td>
<td>n/a</td>
</tr>
<tr>
<td>Other processes</td>
<td>n/a</td>
<td>n/a</td>
<td>kaahtuu</td>
<td>n/a</td>
</tr>
<tr>
<td>Outcome</td>
<td>kaahhtuu</td>
<td>kaahtuu</td>
<td>kaahtuu</td>
<td>kaatuup</td>
</tr>
</tbody>
</table>

Although they approximate to Moosberg's narrow transcriptions, the outcomes in (244) fall short of getting us all the way. The reason for this, apart from ignoring irrelevant details of vowel quality implementation, is the unexplained duration on /t/ in Moosberg's examples. Unfortunately, the system is now lost, and Moosberg's data are the only source on the phonology of the dialect that has come down to us. Nevertheless, the proposed reconstruction in (244) probably comes closer to the right explanation of the observed four-way contrast than any assumption of a four-way contrast in length.

6.4 Conclusions

We have essentially endorsed Sammallahti's view that Saami gradation is an independent Saamic development, although it has the same basis as the parallel developments in the Balto-Finnic languages. Nevertheless, we have rejected two of Sammallahti's claims.

First, we rejected the notion that Fortition was a coherent development, but that it probably arose in two, possibly three distinct waves of Fortition. Although there is no historical or typological evidence for treating Singleton and Geminate Fortition as separate, the hypothesis is rendered unlikely in a constraints-based framework. The differential behaviour of clusters in North
Saami on the one hand and Pite and Ume Saami on the other, however, does argue for separate waves of Fortition in geminates and clusters.

Second, we rejected the claim that post-Fortition Proto-Saami had a four-way contrast in length. A genuine four-way contrast in length seems to be unattested. Although one dialect of Ume Saami has been found to evince a four-way contrast, it was argued that this needn’t be construed as a four-way contrast in length.
Chapter 7

A rule-based account of consonant gradation

Rule-based serial accounts are a useful way of getting a handle on complex interactions between phonological processes, and in this chapter a rule-based account of the alternation of grade and quantity is developed. Attendant segmental processes are also addressed.

The emphasis throughout is on the nominal morphology and phonology of West Finnmark Saami. The verbal morphology of Saami is by and large ignored here, owing to its greater complexity. The verb obligatorily inects for a greater number of morphosyntactic categories than the noun, and displays a higher degree of morphological conditioning. Preliminary examination of the verbal morphology suggests that the claims made here with respect to the nominal phonology ultimately carry over to the phonology of verbs as well, although considerations of space preclude a treatment of verbal phonology and morphology here.

Nominal stems fall into three prosodically-defined categories: parasyllabic, imparasyllabic, and ‘contracted’. The parasyllabic stems may be further subdivided into the bisyllabic stems, discussed in §7.1, and the quadrasyllabic stems (§7.2). Quadrasyllabic stems evince some additional phonological properties which warrant treating the two separately.

The main morphophonological difference between the parasyllabic and imparasyllabic (including the ‘contracted’) stems involves the relation between syllable parity and morphological grade. Roughly, wherever in the paradigm a parasyllabic stem surfaces in the Strong Grade, an imparasyllabic stem surfaces in the Weak Grade in the same cell of the paradigm. Conversely, wherever in the paradigm a parasyllabic stem surfaces in the Weak Grade, an imparasyllabic stem surfaces in the Strong Grade in the same paradigm cell.

The imparasyllabic and ‘contracted’ stems form a natural class with respect to some crucial phonological properties, although the phonology of contracted stems will not be discussed in any detail here. §7.3 addresses the phonology of trisyllabic stems.
## 7.1 The Phonology of Bisyllabic Stems

The Saami noun obligatorily inflects for two numbers (singular and plural), and six cases (nominative, accusative/genitive, illative, locative, comitative, and essive). The essive form is identical in the singular and plural.

Within the nominal paradigm, the distribution of morphological grade is largely transparent phonologically: of the eleven distinct forms in the paradigm, only three forms surface with a morphological grade counter to what one would expect on phonological grounds. In Kautokeino, additional phonological processes have further obscured the relationship in some cases.

The Strong Grade is found in the nominative singular form, and, unexpectedly, in the illative singular and the essive. The accusative/genitive singular evinces a morphologized Weak Grade consonant centre. The plural is in the Weak Grade throughout. This is to be expected, since the plural formatives (-h in the nominative, -j- elsewhere) is always syllable-closing. Only in Kautokeino, where a post-lexical rule mapping /ij/ → /ii/ renders the relevant syllable open in the illative plural and comitative plural, is there an opaque relation between grade and the surface phonology. Even this opaque relationship only holds for stems which end in -ii or -e.

The relationships are summarized in (245). Boldface entries mark a phonologically unexpected instance of a particular grade.

(245) Relationship between grade and morphosyntactic category in bisyllabic nouns

<table>
<thead>
<tr>
<th>Category</th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>Strong</td>
<td>Weak</td>
</tr>
<tr>
<td>ACC/GEN</td>
<td>Weak</td>
<td>Weak</td>
</tr>
<tr>
<td>ILL</td>
<td>Strong</td>
<td>Weak</td>
</tr>
<tr>
<td>LOC</td>
<td>Weak</td>
<td>Weak</td>
</tr>
<tr>
<td>COM</td>
<td>Weak</td>
<td>Weak</td>
</tr>
<tr>
<td>ESS</td>
<td>Strong</td>
<td></td>
</tr>
</tbody>
</table>

(246) to (248) exemplify the distribution of Strong and Weak in the nominal paradigm. Complete paradigms are shown here for *kuollii* ‘fish’, *maannaa* ‘child’, *viessuu* ‘house’, *tsielka* ‘corner’, *pase* ‘spoon’, and *resko* ‘stool’. All are adapted from Nickel (1994: 70ff.).

(246)

<table>
<thead>
<tr>
<th>Category</th>
<th>SG</th>
<th>PL</th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>kuollii</td>
<td>kuollih</td>
<td>maannaa</td>
<td>maanaah</td>
</tr>
<tr>
<td>ACC/GEN</td>
<td>kuollii</td>
<td>kuulijt</td>
<td>maannaa</td>
<td>maanaajt</td>
</tr>
<tr>
<td>ILL</td>
<td>kuollaaj</td>
<td>kuulijoe</td>
<td>maanaaj</td>
<td>maanaajoj</td>
</tr>
<tr>
<td>LOC</td>
<td>kuollis</td>
<td>kuulijn</td>
<td>maanaas</td>
<td>maanaajn</td>
</tr>
<tr>
<td>COM</td>
<td>kuulijn</td>
<td>kuulijkujn</td>
<td>maanaajn</td>
<td>maanaajkujn</td>
</tr>
<tr>
<td>ESS</td>
<td>kuollilin</td>
<td>maannaan</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(247)
7.1. BISYLLABIC STEMS

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>viessu</td>
<td>viessu</td>
<td>tšieka</td>
<td>tšieka</td>
</tr>
<tr>
<td>ACC/GEN</td>
<td>viessu</td>
<td>viessu</td>
<td>tšieka</td>
<td>tšieka</td>
</tr>
<tr>
<td>ILL</td>
<td>viessu</td>
<td>viessu</td>
<td>tšiekii</td>
<td>tšiekii</td>
</tr>
<tr>
<td>LOC</td>
<td>viessu</td>
<td>viessu</td>
<td>tšiekas</td>
<td>tšiekas</td>
</tr>
<tr>
<td>COM</td>
<td>viessu</td>
<td>viessu</td>
<td>tšiekajn</td>
<td>tšiekajn</td>
</tr>
<tr>
<td>ESS</td>
<td>viessu</td>
<td>viessu</td>
<td>tšiekan</td>
<td>tšiekan</td>
</tr>
</tbody>
</table>

(248)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>past</td>
<td>past</td>
<td>reŋko</td>
<td>reŋkoh</td>
</tr>
<tr>
<td>ACC/GEN</td>
<td>past</td>
<td>past</td>
<td>reŋko</td>
<td>reŋkkut</td>
</tr>
<tr>
<td>ILL</td>
<td>past</td>
<td>past</td>
<td>reŋkkut</td>
<td>reŋkkut</td>
</tr>
<tr>
<td>LOC</td>
<td>past</td>
<td>past</td>
<td>reŋkoss</td>
<td>reŋkkutn</td>
</tr>
<tr>
<td>COM</td>
<td>past</td>
<td>past</td>
<td>reŋkkutn</td>
<td>reŋkkukujn</td>
</tr>
<tr>
<td>ESS</td>
<td>past</td>
<td>past</td>
<td>reŋkon</td>
<td>reŋkon</td>
</tr>
</tbody>
</table>

(249) Consonant Gradation: possible analyses

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Mappings</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenition</td>
<td>Q3 → Q2, Q2 → Q1</td>
<td>Triggered preceding closed σ</td>
</tr>
</tbody>
</table>
| Fortition | Q2 → Q3, Q1 → Q2 | Blocked preceding closed σ  
or triggered preceding open σ |

Ultimately, we will come down in favour of the Fortition Analysis, and argue specifically for the context-sensitive (open-syllable-driven) variant. For the time being, though, we will entertain the possibility of both, and see how far each goes in describing and explaining the facts.

Lenition may be formulated as a rule delinking a single mora from the consonant centre preceding a tautopodal closed syllable. Note, that if we have to resort to a single rule, the rule must be specified as non-iterative: in order to derive the chain-shift effect, the rule must apply once only.

(250) Lenition (non-iterative)
Alternatively, we may set up two distinct rules, a rule of Geminate Lenition and a rule of Overlong Lenition, such that the latter counterfeeds the former on the focus as in (251).

(251) **Lenition as two rules in counterfeeding relationship**

a. **Geminate Lenition**
   \[CC \rightarrow C/\_V C]_F \]

b. **Overlong Lenition**
   \[C:C \rightarrow CC/\_V C]_F \]

By ordering Geminate Lenition before Overlong Lenition, the chain shift effect follows.

Fortition may be formulated as the association of a mora to the consonant centre preceding a tautopodial open syllable.

(252) **Fortition (non-iterative)**

\[\mu\]

\[\ldots\]

\[\ldots C(C)V]_F \]

Again, the same comments apply concerning the chain-shifting property of Fortition as were applied above to Lenition. This may also be derived by invoking counterfeeding on the focus of the rule as in (253).

(253) **Fortition as two rules in counterfeeding relationship**

a. **Geminate Fortition**
   \[CC \rightarrow C:C/\_V]_F \]

b. **Singleton Fortition**
   \[C \rightarrow CC/\_V]_F \]

Sample derivations of *kollii*~*kolliih*, ‘golds’, and *kuollii*~*kuollih*, ‘fish’, are shown in (254) and (255). (254) exemplifies the Lenition Analysis; (255) the Fortition Analysis.

(254)

<table>
<thead>
<tr>
<th>Input</th>
<th>/kollii/</th>
<th>/kollii+h/</th>
<th>/kuollii/</th>
<th>/kuollii+h/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenition</td>
<td>n/a</td>
<td>kolliih</td>
<td>n/a</td>
<td>kuollih</td>
</tr>
<tr>
<td>Output</td>
<td>kollii</td>
<td>kolliih</td>
<td>kuollii</td>
<td>kuollii</td>
</tr>
</tbody>
</table>
7.1. BISYLLABIC STEMS

The Ghost Hypothesis, (13), entails that Grade Alternation may be triggered by syllable-closing suffixes which may themselves have no surface reflex. One such suffix is the phonologically abstract accusative/genitive singular suffix -C.

Assume that such ‘ghost’ elements are deleted on the surface by a rule of Ghost Deletion. Ghost Deletion must, of course, be ordered after whichever rule we take to drive Consonant Gradation, since Ghost Deletion counterbleeds Lenition and counterfeeds Fortition. Either way, Consonant Gradation must take derivational precedence. (256) illustrates the rule ordering Lenition>Ghost Deletion, and (257) the ordering Fortition>Ghost Deletion.

<table>
<thead>
<tr>
<th>Input</th>
<th>/kollii/</th>
<th>/kollii+h/</th>
<th>/kuolii/</th>
<th>/kuolii+h/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>kollii</td>
<td>n/a</td>
<td>kuolii</td>
<td>n/a</td>
</tr>
<tr>
<td>Output</td>
<td>[kollii]</td>
<td>[kollii]</td>
<td>[kuolii]</td>
<td>[kuolii]</td>
</tr>
</tbody>
</table>

The Ghost Hypothesis, (13), entails that Grade Alternation may be triggered by syllable-closing suffixes which may themselves have no surface reflex. One such suffix is the phonologically abstract accusative/genitive singular suffix -C.

Assume that such ‘ghost’ elements are deleted on the surface by a rule of Ghost Deletion. Ghost Deletion must, of course, be ordered after whichever rule we take to drive Consonant Gradation, since Ghost Deletion counterbleeds Lenition and counterfeeds Fortition. Either way, Consonant Gradation must take derivational precedence. (256) illustrates the rule ordering Lenition>Ghost Deletion, and (257) the ordering Fortition>Ghost Deletion.

7.1.2 Epenthesis

In Strong Grade heterorganic Q3 clusters of the form C1:C2 or C1:C2C3, where C1 is a sonorant, there is epenthesis of a vowel after C1.

The epenthized vowel is frequently schwa-like, or else it agrees with the quality of the vowels in the neighbouring syllables (Sammallah 1977: 227–229).

As a concomitant to epenthesis, C1 surfaces as monomoraic, since the inserted vowel usurps the hyperchronous mora. This is shown in (258).

<table>
<thead>
<tr>
<th>Input</th>
<th>/kollii+C/</th>
<th>/kuolii+C/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenition</td>
<td>kolliiC</td>
<td>kuolliC</td>
</tr>
<tr>
<td>Ghost Deletion</td>
<td>kollii</td>
<td>kuolli</td>
</tr>
<tr>
<td>Output</td>
<td>[kollii]</td>
<td>[kuolli]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input</th>
<th>/kollii+C/</th>
<th>/kuolii+C/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Ghost Deletion</td>
<td>kollii</td>
<td>kuolli</td>
</tr>
<tr>
<td>Output</td>
<td>[kollii]</td>
<td>[kuolli]</td>
</tr>
</tbody>
</table>

7.1.2 Epenthesis

In Strong Grade heterorganic Q3 clusters of the form C1:C2 or C1:C2C3, where C1 is a sonorant, there is epenthesis of a vowel after C1.

The epenthized vowel is frequently schwa-like, or else it agrees with the quality of the vowels in the neighbouring syllables (Sammallah 1977: 227–229).

As a concomitant to epenthesis, C1 surfaces as monomoraic, since the inserted vowel usurps the hyperchronous mora. This is shown in (258).

<table>
<thead>
<tr>
<th>Heterorganic clusters (epenthesis applies in Q3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>skiol.b.lii</td>
</tr>
<tr>
<td>pär.ra.kan</td>
</tr>
<tr>
<td>kir.ri.jii</td>
</tr>
<tr>
<td>ol.lojju</td>
</tr>
</tbody>
</table>

(259) gives the representation of kuol.b.ka ‘hair-NOM.SG’, where C1 is realized as a geminate.

<table>
<thead>
<tr>
<th>Representation of kuol.b.ka ‘hair-NOM.SG’ (no delinking from minor σ)</th>
</tr>
</thead>
</table>
The first component of the cluster $C_1$ may also be realized as a non-geminate. In this case, the consonant surfaces not as an onset in the epenthized syllable, but as coda in the head syllable of the foot.\footnote{This is also in accordance with the syllabifications proposed by Sammallahti (1977: 61).} Thus, the forms in (258) may alternatively be realized as skäol:si, par:si kan, kir:i,ji, and ol:ojun. (260) illustrates the alternative representation of the same word as in (259), in which the consonant centre has optionally undergone degemination.

\begin{center}
(260) Representation of kuolka ‘hair-NOM.SG’ (with delinking from minor \(\sigma\))
\end{center}

This kind of epenthesis has all of the properties of vowel ex crescence as defined by Levin (1987) and Bagemihl (1991: 600). In contrast to ‘truly’ epenthetic vowels, ex crescent vowels are featurally unspecified on the surface, their quality depending on neighbours vowels, coarticulation, or tending to a schwa default.\footnote{For surface underspecification see also Choi (1995), Bergen (1994), and Myers (1998).}

Ex crescence is also non-structure-preserving, since the ex crescent vowel does not correspond to any underlying segment of the inventory. Insertion is not a response to the exigencies of syllabification, but rather by a need to mediate a transition between “adjacent articulations requiring some degree of constriction in the oral tract”.

Further, ex crescent vowels are not referred to by any of the phonological rules of the language.

Following Levin (1987: 5), I will assume that ex crescence consists in the insertion of a featurally unspecified vocalic root-node. This much, then, is phonological. The quality of the ex crescent vowel itself is not determined by the grammar, but arises out of phonetic coarticulation effects. See Sammallahti (1977: 229) for what these effects are in Saami.

Unsurprisingly, ex crescence is blocked in homorganic and obstruent-initial clusters, as shown in (261). Outputs such as *ka:nta and *hispuii (for ka:nta ‘boy’ and hispuii ‘outlet’) are completely ill-formed. Homorganic blocking is an instance of the geminate integrity resulting from shared place specifications (Hayes 1986). Ex crescence is absent in obstruent-initial clusters in general.

\begin{center}
(261) Ex crescence blocked in Q3
\end{center}

\begin{enumerate}
\item \textit{Homorganic clusters}
\begin{itemize}
\item ka:nta ‘boy, NOM.SG’
\item pum:paa ‘box, chest, NOM.SG’
\item aal:tuu ‘female reindeer, NOM.SG’
\end{itemize}
\item \textit{Obstruent-initial clusters}
\end{enumerate}
7.1. **BISYLLABIC STEMS**

<table>
<thead>
<tr>
<th>luspii</th>
<th>'outlet, NOM.SG'</th>
</tr>
</thead>
<tbody>
<tr>
<td>leasmii</td>
<td>'arthritis, NOM.SG'</td>
</tr>
<tr>
<td>maatškii</td>
<td>'journey, NOM.SG'</td>
</tr>
</tbody>
</table>

The Epenthesis rule is formulated in (262) as the delinking of $C_1$ from a mora, with reassociation of an empty vowel to the same mora.

(262) **Swarabhakti**

\[
\begin{array}{c}
\mu \\
[\text{son}]
\end{array}
\begin{array}{c}
C_1 \\
C_2 \ (C_3)
\end{array}
\begin{array}{c}
V
\end{array}
\]

Condition: $C_1$ and $C_2$ are not homorganic.

Epenthesis must be fed by Fortition, as shown in (263).

(263)

<table>
<thead>
<tr>
<th>Input</th>
<th>/kirjii/</th>
<th>/luspii/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>kirjii</td>
<td>luspii</td>
</tr>
<tr>
<td>Epenthesis</td>
<td>kir.ri.jii</td>
<td>n/a</td>
</tr>
<tr>
<td>Output</td>
<td>[kir.ri.jii]</td>
<td>[luspii]</td>
</tr>
</tbody>
</table>

Epenthesis may feed an *optional* rule of Degemination which applies preceding a syllable containing an excrecent vowel, as in (264).

(264)

<table>
<thead>
<tr>
<th>Input</th>
<th>/kirjii/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>kirjii</td>
</tr>
<tr>
<td>Epenthesis</td>
<td>kir.ri.jii</td>
</tr>
<tr>
<td>Degemination</td>
<td>ki.ri.jii</td>
</tr>
<tr>
<td>Output</td>
<td>[ki.ri.jii]</td>
</tr>
</tbody>
</table>

### 7.1.3 Balance I: Primary Lengthening

We have already come across Balance in §4.1. Sammallahti treats this as two rules: 'primary lengthening' and 'secondary lengthening'. They differ only in that *primary lengthening* applies in the Strong Grade to a monosegmental consonant centre, while *secondary lengthening* applies exclusively in the Weak Grade, usually to a cluster. There are also subtle differences in the environment in which each of these rules applies.

Balance lengthens a CC geminate to C:C preceding a bimoraic syllable nucleus. In primary lengthening, the vowel preceding the consonant centre must be short in both East Enontekiö and Kautokeino. Balance is blocked if the vowel in the preceding syllable is long. Thus, /paalli/ 'while, period of time' is blocked from undergoing lengthening to *paal:lii*. This is an instance of blocking, involving the interaction of **BALANCE** with a higher-ranked constraint which bans the resulting *VV C:C* configuration. In secondary lengthening, however, the two
dialekts differ. In East Enontekiö, secondary lengthening is subject to the requirement that the vowel in the first syllable be short (Sammallah 1998b: 49). In Kautokéino, however, there is no such restriction: Balance is a more general rule which applies whenever a geminate or cluster in Q2 precedes a long nucleus. In East Enontekiö, Balance also fails to apply after a syllable containing a non-high vowel or diphthong /æa/ (Sammallah 1998b: 49). Hence, in East Enontekiö, /tjuodí:fu:nt/ → tʃgatʃfuun, but /oadí:fu:nt/ → oatʃfuun. *oatʃfuun ‘I get’. In Kautokéino, on the other hand, secondary lengthening applies irrespective of the quantity of the vowel (Magga 1984: 42). Not only does secondary lengthening apply following a non-high rising diphthong (hence, oatʃfuun ‘I get’ is the grammatical form in Kautokéino), but it applies following a long /aa/ as well. Magga thus cites forms such as aapi:najt ‘mercy; acc/gen.pl, aallstaas ‘Alta-loc.sg’, maarfííih ‘sausage-nom.pl’, and so on.

There is an interesting functional parallel to Balance in Finnish (Lehtonen 1970; Bye 1998b). Lehtonen found that a long vowel conditioned a significant lengthening of the preceding consonant, geminates and singletons alike. For example, the intervocalic geminate velar stop in takka ‘fire place-part.sg’ is 25% longer in duration than the corresponding geminate in takka ‘fire place-nom.sg’. He reports similar enhancements in singletons, although the degree of lengthening here isn’t quite as striking. Still, the intervocalic dental stop in sataa ‘it rains’ is 16% longer than the corresponding singleton in sata ‘hundred’.

The alternations in (265) illustrate Balance under its ‘primary lengthening’ aspect.

(265) | jah:kii | jakii | ‘year~acc/gen.sg’
| pal:luu | paluu | ‘fear~acc/gen.sg’
| joh:taa | jokkaan | ‘he starts to travel~I travel’

(266) states the generalization in terms of a rule.

(266) Balance (rule)

\[
CC \rightarrow C:C/\_VV|FI
\]

Sample derivations illustrating ‘primary lengthening’ are given in (267) for palluu, ‘fear’, and jah:kii, ‘year’.

(267)

<table>
<thead>
<tr>
<th>Input</th>
<th>/pal:luu/</th>
<th>/jah:kii/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance</td>
<td>pal:luu</td>
<td>jah:kii</td>
</tr>
<tr>
<td>Output</td>
<td>[pal:luu]</td>
<td>[jah:kii]</td>
</tr>
</tbody>
</table>

A Fortition Analysis requires different assumptions about the input. Assuming underlying forms /pal:luu/ and /jah:kii/, Fortition and Balance would interact as in (268) below. Given closure of the syllable by a suffix, such as -h PLURAL, neither process applies.

(268)
As can be seen from (268), it is necessary for Fortition to feed Balance in order to get the correct output form.

The corresponding Weak Grade forms are in Q1, e.g. /paluu/ ‘fears’, and /jakii/ ‘years’. In order to get the correct quantity in the gradated cases, we have to assume that Lenition bleeds Balance. Consider the derivations of /paluu/ ‘fears’, and /jakii/ ‘years’ according to the Lenition Hypothesis in (269).

(269)

Ordering Lenition prior to Balance, however, predicts, counterfactually, that Balance applies in the Weak Grade of a lenited Q3 singleton or obstruent-initial cluster. In (270), Balance does not apply in the Weak Grade as expected, even though its structural description is satisfied at the relevant point of the derivation.

(270)

Lenition of Q3 to Q2 creates the environment for Balance to apply. Yet this is not what we find. Nothing we have said so far explains why.

One solution is to adopt a global constraint against backtracking in the derivation (cf. Chen 1999), whose effect is to prohibit reversing the effect of a previous rule (a Duke-of-York effect) — Balance must not undo the work of Lenition.

(271) **No Backtracking Principle**

If two rules, $P$ and $Q$, are ordered such that $P$ precedes $Q$, and $P$ describes a mapping $A \rightarrow B$ and $Q$ describes the mapping $B \rightarrow A$, then $Q$ is blocked from applying.

(256) exemplifies the blocking effect of (271) for /kollis/ ‘gold-LOC.SG’ and /ostuu/ ‘rowan bark-LOC.SG’.

(272)
Another nuance of Balance is that its sensitivity to the length of the vowel varies depending whether it applies to a monosegment or cluster. In the case of primary lengthening, Balance is blocked by a diphthong in the preceding syllable. Thus, /kuolli/ is not mapped to *kuol:lii as might be expected, but kuollii. This restriction is not observed in secondary lengthening, where the target of Balance is a cluster. Thus, /skuolfii+h/ becomes sküolf:fiih ‘owls’. We return to this point in §7.1.5.

7.1.4 Coda Maximization

Now let us turn to the behaviour of sonorant-initial clusters. Recall that biliteral sonorant-initial clusters surface with gemination of the second component of the cluster in the Weak Grade due to a rule of Coda Maximization. Given an analysis in which Grade Alternation is driven by Lenition, Lenition must feed Coda Maximization (273). Coda Maximization applies exclusively to the head syllable of the foot, and may be conceived of in terms of two disjunctively ordered rules. The first rule geminates C2 of a biliteral cluster, while the second rule ensures that C2 surfaces as a coda in a triliteral cluster.

(273) Coda Maximization (rule)

b. C1.C2C3 → C1C2.C3/V VX][Ft

Condition: C1 is [+sonorant, +voice].

In biliteral (C1.C2) clusters C2 undergoes gemination and ambisyllabication to maximize the coda of the stressed syllable, giving C1C2C2. Lengthening of the second member of a cluster will be conceptualized as linking to the coda mora of the preceding syllable, as in kuolkkä ‘hair-ACC/GEN.SG’.

(274) Coda Maximization in kuolkkä ‘hair-ACC/GEN.SG’

\[
\sigma
\mu
\mu
\mu
\k
\o
\l
\a
\]

Triliteral (C1:C2C3) clusters pattern differently under Coda Maximization. Strong Grade /C1:C2C3/ alternates with Weak Grade /C1C2.C3/. That is, neither C2 nor C3 undergo gemination as in the biliteral C1C2 case.

Note that Coda Maximization is precluded from occurring in Q3, i.e. it cannot apply to give surface clusters of the form C1:C2.C2. We will address the reasons for this gap when we address quantity from a constraints-based perspective in §8.2.2.
7.1. BISYLLABIC STEMS

So far we have established that Consonant Gradation, whether Lenition or Fortition, feeds Balance (266). Lenition must also feed Coda Maximization (273). Sample derivations for *arv'oh*, ‘spirited’, and *palt'tsah*, ‘useless implement (ski, knife, etc.)’ are shown in (275) and (276). In (275), Lenition feeds Coda Maximization. In (276), Fortition bleeds Coda Maximization.

(275)

<table>
<thead>
<tr>
<th>Input</th>
<th>/arv'o/</th>
<th>/arv'o+h/</th>
<th>/pal'ts'a/</th>
<th>/pal'ts'a+h/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenition</td>
<td>n/a</td>
<td>arv'oh</td>
<td>n/a</td>
<td>palt'tsah</td>
</tr>
<tr>
<td>Coda Max</td>
<td>n/a</td>
<td>arv'oh</td>
<td>n/a</td>
<td>palt'tsah</td>
</tr>
<tr>
<td>Other rules</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>palt'tsah</td>
</tr>
<tr>
<td>Output</td>
<td>[arv'o]</td>
<td>[arv'oh]</td>
<td>[pal'ts'a]</td>
<td>[palt'tsah]</td>
</tr>
</tbody>
</table>

The corresponding Fortition Analysis of the same alternations is given in (276).

(276)

<table>
<thead>
<tr>
<th>Input</th>
<th>/arv'o/</th>
<th>/arv'o+h/</th>
<th>/pal'ts'a/</th>
<th>/pal'ts'a+h/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>arv'o</td>
<td>n/a</td>
<td>pal'ts'a</td>
<td>n/a</td>
</tr>
<tr>
<td>Coda Max</td>
<td>n/a</td>
<td>arv'oh</td>
<td>n/a</td>
<td>palt'tsah</td>
</tr>
<tr>
<td>Other rules</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>palt'tsah</td>
</tr>
<tr>
<td>Output</td>
<td>[arv'o]</td>
<td>[arv'oh]</td>
<td>[pal'ts'a]</td>
<td>[palt'tsah]</td>
</tr>
</tbody>
</table>

7.1.5 Balance II: Secondary Lengthening

Coda Maximization in the Weak Grade of a cluster feeds Balance. Biliteral sonorant-initial clusters undergo Balance to give surface clusters with the shape C$_1$C$_2$:C$_3$. Triliteral clusters also undergo Balance to give surface clusters with the shape C$_1$C$_2$:C$_3$. Examples in (277) are drawn from the Kautokeino dialect (Magga 1984: 32–35). The Weak Grade column shows the effects of Balance.

(277)
CHAPTER 7. A RULE-BASED ACCOUNT

**Strong Grade** | **Weak Grade** | **Gloss**
---|---|---
a. mubpiii | nuppiiih | 'other NOM.SG~NOM.PL'
rud:tiis | rat:tiis | 'chest NOM.SG~LOC.SG'
rud:tiu | rut:tiujt | 'ravine NOM.SG~ACC.PL'
roq:kii | rok:kiis | 'hollow NOM.SG~LOC.SG'
roq:kii | rok:kaaf | 'hollow NOM.SG~DIM'
b. irkiii | irk:kaaf | 'suitor NOM.SG~DIM'
hil:puu | hil:puu | 'wild creature NOM.SG~NOM.PL'
arvu | arv:vuus | 'energy, will NOM.SG~LOC.SG'
pol:fi | pol:fiih | 'long-haired dog NOM.SG~NOM.PL'
kor:jii | kor:jiih | 'waterfall NOM.SG~NOM.PL'
c. pod:tiin | pod:tiis | 'bottom NOM.SG~LOC.SG'
kob:tiin | kob:maaj | 'ghost NOM.SG~DIM'
d. fier:rapii | fier:miis | 'fishing net NOM.SG~LOC.SG'
kir:rapii | kir:maaj | 'churn NOM.SG~ACC/GEN.SG'
e. mir:rikku | mirk:kuft | 'poison NOM.SG~ACC/GEN.PL'
imir:rikpu | mir:apaaf | 'S.O. who gets offended easily NOM.SG~DIM'
limpu | limpu | 'ump NOM.SG~ACC/GEN.PL'
rinjtjii | rinjtjih | 'person who goes half naked NOM.SG~NOM.PL'

If the vowel in $\sigma_2$ is short, however, Balance does not apply. Compare the words on the left with those on the right in (278).

(278)

<table>
<thead>
<tr>
<th>Input</th>
<th>/arvu/</th>
<th>/arvu+h/</th>
<th>/maajstii/</th>
<th>/maajstii+C/</th>
</tr>
</thead>
<tbody>
<tr>
<td>rikku</td>
<td>jikko</td>
<td>'hush, IMP'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>jakku</td>
<td>jappu</td>
<td>'stare, IMP'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kottaa</td>
<td>jutta</td>
<td>'grow, IMP'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nappii</td>
<td>toppu</td>
<td>'over there'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vilj:juu</td>
<td>hilip:eh</td>
<td>'closet'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kolkkuit</td>
<td>kolkkoh</td>
<td>'male reindeer exhausted from rutting'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arvu</td>
<td>arvoh</td>
<td>'spirited'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>julkunu</td>
<td>kalmme</td>
<td>'sensitive to cold, nesh'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample derivations are shown in (279) for /arvu+h/ 'will power+PL', and /maajstii+C/ 'by tasting'. These forms are contrasted with comparable Strong Grade forms, also in Q3, of arvu 'will (NOM.SG)' and maajstaa 'he tastes'.

(279)

<table>
<thead>
<tr>
<th>Input</th>
<th>/arvu/</th>
<th>/arvu+h/</th>
<th>/maajstii/</th>
<th>/maajstii+C/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lenition</td>
<td>n/a</td>
<td>arvuuh</td>
<td>n/a</td>
<td>maajstiiC</td>
</tr>
<tr>
<td>Coda Max</td>
<td>n/a</td>
<td>arvuuh</td>
<td>n/a</td>
<td>maajstiiC</td>
</tr>
<tr>
<td>Balance</td>
<td>n/a</td>
<td>arvuuh</td>
<td>n/a</td>
<td>maajstiiC</td>
</tr>
<tr>
<td>Other rules</td>
<td>n/a</td>
<td>n/a</td>
<td>maajstaa</td>
<td>maajstii</td>
</tr>
</tbody>
</table>

(280) establishes the same input-output mappings from the perspective of the Fortition-based analysis.
7.1. BISYLLABIC STEMS

So far we have not brought any considerations to bear on the relative merit of the Fortition and Lenition analyses. From the perspective of derivational economy, though, it is clear that the Fortition account needs fewer derivational steps, and avoids Duke-of-York style rule interaction. To see this, consider the derivation by lenition of the alternation *saltii ~ salttiis*, ‘bridge, nom.sg ~ loc.sg’.

(281) clearly has a Duke-of-York character. The crucial stage in the derivation in (281) is the point at which Balance applies, relinking or reinserting a mora which was delinked or deleted earlier at the point Lenition applies. Consider the same alternation from a Fortition perspective. The Duke-of-York effect vanishes, as a direct consequence of our assumptions about the input and the directional bias of Consonant Gradation.

(282) At no point in the derivation in (282) is a Duke-of-York style rule interaction invoked.

One consequence of rejecting the Lenition Hypothesis in favour of the Fortition Hypothesis involves the observed failure of secondary lengthening in the Weak Grade, where the result of applying the rule would give a surface form whose consonant centre is identical in form to the corresponding Strong Grade. On the Lenition Hypothesis, this generalization may be formulated in terms of a global constraint on the derivation, to the effect that they be monotonic. This was the *No Backtracking* constraint, given in (271)
In the Fortition account, on the other hand, this monotonicity cannot be appealed to, since the hypothesized input /kolli+s/ does stand in a cumulative relationship to the ill-formed output *[kolliis]. Balance would have to be blocked by some other as yet unknown derivational constraint.

Recall that Balance fails to apply in a monosegmental consonant centre following a diphthong or long vowel. So far, nothing we have said predicts this. One way of understanding the pattern is in terms of rule sandwiching. Suppose that Singleton Fortition is in fact properly split up into two rules: one rule applies following a short vowel (Singleton Fortition\(^1\)), and the other rule applies following a long vowel or diphthong (Singleton Fortition\(^2\)). Ordering the first prior to Balance, and the second rule subsequent to Balance would generate the correct surface distribution of Balance.

In this way, Balance is counterfed by Singleton Fortition\(^2\), but fed by Singleton Fortition\(^1\), which is precisely the schema for rule sandwiching (the faithfulness violations introduced by the two Singleton Fortition rules overlap).

Summing up, so far we have at least the following crucial orderings.

**Rule Ordering**

1. Lenition or Fortition, (250), (251), (252), (253)
2. Coda Maximization, (273)
3. Balance, (266)
7.1.6 Intervocalic Voicing and Spirantization

Grade or quantity affects segmental realization. An intervocalic dental stop /t/ or palatal stop /c/ in Q1 is mapped to the corresponding voiced approximants $\delta$ and $\jmath$.

(287) a. ht~$\delta$ koahti~koahii $\sim$ 'house~houses'
   b. cc~$\jmath$ vuocca~vuaja $\sim$ 'butter, Nom.sg~Acc.sg/Gen.sg'

Voicing and spirantization leave the Q1 stops at the corresponding labial and dorsal places of articulation unaffected. Also, the alveolar affricate /ts/ fails to undergo the rule.

Is the class of targets subject to the rules of Voicing and Spirantization a natural one? It would appear so: the class \{t c\} can be captured by the distinctive feature matrix \[ [+\text{coronal} \quad -\text{strident}] \]. The inclusion of \[-\text{strident}\] in the matrix excludes the affricate /ts/ from undergoing the process.

(288) Voicing and spirantization (rule)

\[
[+\text{coronal} \quad -\text{strident}] \rightarrow [+\text{voice} \quad +\text{cont}] /V\_V
\]

Voicing/Spirantization must be ordered after Fortition, since Fortition bleeds Voicing/Spirantization. Consider (289).

(289)

<table>
<thead>
<tr>
<th>Input</th>
<th>/koati/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>kootii</td>
</tr>
<tr>
<td>Voicing/Spir</td>
<td>n/a</td>
</tr>
<tr>
<td>Other rules</td>
<td>koati</td>
</tr>
<tr>
<td>Output</td>
<td>[koati]</td>
</tr>
</tbody>
</table>

Compare (289) with the derivation in (290), where Fortition is prevented from applying due to the following closed syllable. In this case, Voicing/Spirantization applies.

(290)

<table>
<thead>
<tr>
<th>Input</th>
<th>/koati+s/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>n/a</td>
</tr>
<tr>
<td>Voicing/Spir</td>
<td>koasis</td>
</tr>
<tr>
<td>Output</td>
<td>[koasis]</td>
</tr>
</tbody>
</table>

7.1.7 The phonology of voiceless stops: preaspiration and homorganic stop deletion

Preaspirated stops are found in the following positions:

- in marginal position, e.g. katsel$paf$ 'Siberian tit'.
- foot-medially in Q2 or Q3, e.g. kohpii 'hollow, depression', kohpi 'bay',
• foot-medially following a sonorant, e.g. kir.ro.lkuu ‘church’, taal.hkas
  ‘medicine’, haarthasfii ‘with legs wide apart’.

In a sonorant-initial cluster, the distinction between plain and preaspirated
stop is distinctive, as shown in (291).

(291)

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kaal.hpi</td>
<td>‘calf’</td>
</tr>
<tr>
<td>peal.hki</td>
<td>‘thumb’</td>
</tr>
<tr>
<td>taan:tsah</td>
<td>‘to tread’</td>
</tr>
<tr>
<td>heal.lo.hpuu</td>
<td>‘easily done’</td>
</tr>
<tr>
<td>peal.hkihih</td>
<td>‘to scold’</td>
</tr>
<tr>
<td>lean:tsii</td>
<td>‘stocky person’</td>
</tr>
</tbody>
</table>

Elsewhere, however, the occurrence of one or the other is predictable. Preaspirated stops are banned in the following positions:

• word-initially,
• foot-medially when in Q1,
• foot-medially following a voiced stop.

In foot-medial position between vowels, voiceless stops are always preaspirated when geminate. We may view the derivation of the surface clusters of
h+stop as a three-step process consisting of the following processes:

• Aspiration rule inserting [aspirated] feature into geminate.
• Fission rule which creates a separate root node to bear the feature [aspi-
  rated].
• Supralaryngeal Delinking rule.

First, a rule of Aspiration inserts the feature [+spread glottis] or [aspirated]
into an intervocalic geminate voiceless stop. The feature will be designated with
the symbol ‘h’.

The aspiration rule applies to an intervocalic geminate. Beyond the require-
ment of moraicity, it is indifferent to the number of moras (i.e. whether
the geminate is Q2 or Q3). This is rendered in the rule by the expression $\mu_n$ ($n$ is
1 or 2).

The analytical division between Fission and Supralaryngeal Delinking is sup-
ported by data on preaspiration in other dialects. The varieties of Saami spoken
in Varanger (Várjat), Spansdalen (Rungu) and Skåland (Skánit), for example,
pattern conservatively: they have Aspiration and Fission, but lack the Suprala-
ryngeal Delinking rule. The data from Skåland Saami in (292), drawn from
Jernsletten (n.d.: 66), illustrate the point.

(292) **Preaspiration in Skånland Saami**

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>oh:tsie</td>
<td>‘seeker’</td>
</tr>
<tr>
<td>achtifje</td>
<td>‘father’</td>
</tr>
<tr>
<td>cöhppat</td>
<td>‘to learn’</td>
</tr>
<tr>
<td>joh:tue</td>
<td>‘traveler’</td>
</tr>
<tr>
<td>jæhkkiet</td>
<td>‘insulated-pl’</td>
</tr>
<tr>
<td>oh:tsat</td>
<td>‘to look for’</td>
</tr>
<tr>
<td>achtifje</td>
<td>ACC/GEN.SG</td>
</tr>
<tr>
<td>oohppam</td>
<td>‘I learn’</td>
</tr>
<tr>
<td>joh:tiet</td>
<td>‘to move’</td>
</tr>
<tr>
<td>jæhkkæm</td>
<td>‘I am insulated’</td>
</tr>
</tbody>
</table>
7.1. BISYLLABIC STEMS

The Aspiration rule is given in (293).

(293)  
\[
\text{Aspiration (rule)} \quad \mu_n \\
\emptyset \rightarrow [\text{asp}] /V \\
\left[ \begin{array}{c} -\text{voi} \\ -\text{son} \end{array} \right]
\]

Fission, (294), involves the insertion of a root node onto which the feature [aspirated] spreads, effectively, the epenthesis of a laryngeal glide. I assume that, in the resulting configuration, the [aspiration] feature is shared by the laryngeal glide and the following voiceless stop, following (Ringen 1999), who builds on work by (Kingston 1990). The effect of the Fission rule is to render the aspiration a 'semi-independent' segment (cf. also Keer 1998: 106).\(^3\)

(294)  
\[
\text{Fission (rule)} \quad \mu_n \\
\mu_n \\
\left[ \begin{array}{c} -\text{voi} \\ -\text{son} \end{array} \right]
\]

Scanning Q3 and Q2 inputs, e.g. /kaat:tuu/ 'cat' and /koatt:ii/ 'Saami tent', (294) outputs intermediate \text{kaa}h:t:tuu and \text{koa}h:t:ii respectively.

Fission feeds a rule of Supralaryngeal Delinking. The supralaryngeal features of the stop delink from the mora (or moras, if Q3) leaving the [aspiration] component associated with the mora.

(295)  
\[
\text{Supralaryngeal Delinking (rule)} \quad \mu_n \\
\left[ \begin{array}{c} -\text{voi} \\ -\text{son} \end{array} \right]
\]

Supralaryngeal Delinking gives kaah:tuu and koahtii respectively.

Aspiration interacts opaquely with another segmental rule of Homorganic Voiced Stop Deletion. A voiced stop is deleted preceding a tautosyllabic homorganic stop. In the graphic representation in (296), ‘D’ stands for the voiced component of the cluster; ‘T’ stands for the homorganic voiceless component of the cluster.

(296)  
\[
\text{Homorganic Voiced Stop Deletion (rule)}
\]

\(^3\)Preaspiration in Saami has much in common with that of Icelandic, although further exploration of the parallels is beyond the scope of this dissertation. See Thránisson (1978), Hermans (1985), Selkirk (1990), Lombardi (1994), Iverson and Salmons (1995), and Ringen (1999) for discussion.
Crucially, since Homorganic Deletion creates the environment in which Preaspiration can apply (an intervocalic geminate voiceless stop), Preaspiration must precede Homorganic Deletion. Homorganic Deletion thus counterfeels Preaspiration. The interaction between the two is illustrated in (297) for *kattsah* 'claws'.

(297)

<table>
<thead>
<tr>
<th>Input</th>
<th>/kadtsa+h/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>n/a</td>
</tr>
<tr>
<td>Coda Maximization</td>
<td>kadtsaah</td>
</tr>
<tr>
<td>Aspiration</td>
<td>n/a</td>
</tr>
<tr>
<td>Homorganic Deletion</td>
<td>kattsah</td>
</tr>
<tr>
<td>Other rules</td>
<td>kattsah</td>
</tr>
<tr>
<td>Output</td>
<td>[kattsah]</td>
</tr>
</tbody>
</table>

7.1.8 The phonology of prestopped nasals: glottalization and voicing alternation

The prestopped nasals undoubtedly constitute the most complex chapter of West Finnmark Saami phonology.

The prestopping of nasals is not a particularly frequent process in the world’s languages. Nevertheless, it is attested in a variety of unrelated languages throughout the world. Within Europe, the phenomenon is found in Icelandic and some dialects of West Norwegian spoken in Hordaland (Chapman 1962; Sandøy 1995), as well as in Manx Gaelic (Broderick 1984).

Probably the most enlightening comparative data, though, come from the Australian Aboriginal languages, where prestopping of nasals is something of an areal phenomenon (Hercus 1972; Dixon 1970, 1980; Austin 1981; Maddieson and Ladefoged 1993; Ladefoged and Maddieson 1996).4

Diyari, there is optional prestopping of intervocalic apical alveolar and laminal dental nasals following a stressed nucleus. Interestingly, though, the process is blocked if the initial consonant is also a nasal, exactly as in North Saami.5

(298) The distribution of prestopping in Diyari

- a. /kani/ → [kání] ~ [kádni] ‘frillnecked lizard’
- b. /yina/ → [yíná] ~ [yídna] ‘you’
- c. /qama/ → [ńáma] ‘to be’

4Thanks to Brett Baker and Andy Butcher for bringing these cases to my attention.

5Identical factors were at work in the evolution of the prestopped vs. plain nasal contrast in languages such as Arrernte (Maddieson and Ladefoged 1993; Ladefoged and Maddieson 1996) and Olgoló (Dixon 1970, 1980). Thus, in Olgólo, *tama ‘man’ and *qama ‘mother’ became *ama and ama respectively, following a historical process which deleted the initial consonant of the word.
7.1. BISYLLABIC STEMS

The Arabana-Wajgaggeru languages (Hercus 1972) evince a similar pattern, extending the process optionally to bilabial nasals, but not, apparently, velars. In North Saami, the process is generalized to all places of articulation. Historically, an oral stop developed before a nasal (as in liepma ‘broth’, suotna ‘sinew’, jiega ‘ice’, etc.). The process was blocked when the preceding syllable onset was nasal (e.g. namma ‘name’, not *napma). Hence, there are no nasal-initial words with a consonant centre containing a prestopped (or glottalized) nasal.

The nasal series present an especially complicated picture in West Finnmark Saami. In Q1 we have plain nasals, but these alternate with one of three distinct series, the choice of which depends on sociolinguistic factors. In Q2 and Q3, the nasals surface in one of three ways:

- prestopped only,
- prestopped and glottalized, or
- glottalized only.

The prestopping-only system is favoured amongst older speakers of the dialect, while prestopping with glottalization system is prevalent amongst younger speakers. These systems are summarized in tabular form in (299) and (300) below.

(299) Nasals prestopped only

\[
\begin{array}{cccc}
Q3_s & b & m & d & n \\
Q2_w & p & m & t & n \\
Q3_w & p & m & t & n \\
Q1_s & m & n & n
\end{array}
\]

(300) Nasals prestopped and glottalized

\[
\begin{array}{cccc}
Q3_s & bP\tilde{m} & d\tilde{m} & n\tilde{m} \\
Q2_w & p\tilde{m} & t\tilde{m} & c\tilde{m} \\
Q3_w & p\tilde{m} & t\tilde{m} & c\tilde{m} \\
Q1_s & m & n & n
\end{array}
\]

The youngest system which evinces glottalization without prestopping is laid out in (301).

(301) Nasals glottalized only

\[
\begin{array}{cccc}
Q3_s & m\tilde{m} & n\tilde{m} & p\tilde{m} \\
Q2_w & ?m & ?m & ?m \\
Q3_w & ?m & ?m & ?m \\
Q1_s & m & n & n
\end{array}
\]
Historically, the prestopped nasals are probably derived from plain nasals in Proto-Saami by a rule of Prestopping. The rule applied to an intervocalic geminate nasal, provided that the preceding onset was not also [+nasal]. The synchronic status of the rule is less secure, however, since many words in the modern language evince plain nasals where prestopped nasals would be expected by rule, e.g. *tsamna 'kiss'. The result is a potential surface contrast between plain and prestopped nasals. Is this distribution underlying, or is prestoppping derived by rule, thwarted in some cases by lexical exceptionality? There are a couple of considerations which seem to point in the direction of a synchronic rule.

First, there is one environment in which prestoppping applies exceptionlessly: following a svarabhakti vowel (derived by epenthesis following an approximant /r l β ŋ/ in Q3), a following nasal is always prestopped. Indeed, the rule is entirely productive, and evidence of its productivity is seen in recent loanwords from Norwegian, such as *tea-n.ru.p'ma 'term' and telefuβJo.tNna 'telephone'.

Second, deriving prestoppping by rule is more consistent with the monosegmental behaviour exhibited by stop+nasal clusters under Consonant Gradation. That is, Q2 /p'm t'n č'?/ alternate with plain nasals /m n ŋ/ in Q1. True clusters do not alternate in this way.

Like preaspiration, prestoppping can be broken down into two distinct rules, (i) a rule inserting a voiced homorganic prestop preceding a geminate nasal in either Q2 or Q3, and (ii) a rule delinking the nasal component from any moras to which it is linked. The analysis is thus similar to that proposed for preaspiration, and the analytical split is similarly motivated by typological evidence. The Maattivuono dialect of Sea Saami (Ravila 1932) is an example of a dialect where prestoppping is absent on the surface.

(302) No prestoppping in Maattivuono Sea Saami

<table>
<thead>
<tr>
<th>Welsh</th>
<th>English</th>
<th>Norwegian</th>
</tr>
</thead>
<tbody>
<tr>
<td>timmə</td>
<td>'weak, ATTR'</td>
<td>timiis</td>
</tr>
<tr>
<td>jienə</td>
<td>'ice'</td>
<td>jieniill</td>
</tr>
<tr>
<td>cemnii</td>
<td>'mother'</td>
<td>cemniiin</td>
</tr>
<tr>
<td>kumjnuu</td>
<td>'vagina'</td>
<td>kumjnuu ACC/GEN.SG</td>
</tr>
<tr>
<td>tuoŋɡaanbik</td>
<td>'to patch up'</td>
<td>tuoŋɡaŋjam 'I could patch up'</td>
</tr>
</tbody>
</table>

In the Kaakkuri dialect of Sea Saami, however, prestoppping applies. Examples, adapted from Ravila (1932: 35) are shown in (303).

(303) Prestoppping in Kaakkuri Sea Saami

<table>
<thead>
<tr>
<th>Welsh</th>
<th>English</th>
<th>Norwegian</th>
</tr>
</thead>
<tbody>
<tr>
<td>jiešnua</td>
<td>'voice'</td>
<td>jiešnua ACC/GEN.SG</td>
</tr>
<tr>
<td>jiešnua</td>
<td>'ice'</td>
<td>jiešnua ESS</td>
</tr>
<tr>
<td>eəršnuui</td>
<td>'mother'</td>
<td>eəršnuui ACC/GEN.SG</td>
</tr>
<tr>
<td>piešnuunu</td>
<td>'food'</td>
<td>piešnuunu ACC/GEN.SG</td>
</tr>
<tr>
<td>tuoŋɡaŋaŋbik</td>
<td>'to patch up'</td>
<td>tuoŋɡaŋaŋ 'I patched up'</td>
</tr>
</tbody>
</table>

---

6Sammallahiti (1998b: 194) locates the development in the transition from Proto-Saami to Common Western Saami, although, puzzlingly, he reconstructs these with a plain glottal stop rather than a prestop, e.g. *eanne>ea?nne 'mother', and *pæmnu>pie?mnu 'food'. In the light of the blocking activity of a preceding nasal consonant, I deem this unlikely: positing a rule of prestoppping as opposed to glottalization only, brings Saami more into line with other known cases of this pattern, e.g. Australian languages.
7.1. BISYLLABIC STEMS

7.1. BISYLLABIC STEMS

\[\text{tuo} \text{g} \text{a}: \text{n} \text{ch} \text{up}\]

(304) gives the Prestopping rule.

(304)  \text{Prestopping \text{(rule)}}

\[
\begin{array}{c}
\mu_n \\
\text{[+nas]} \text{P1} \\
\end{array} \rightarrow \begin{array}{c}
\mu_n \\
\text{[-nas]} \text{P1} \text{[+nas]} \\
\end{array}
\]

Condition: Preceding onset (if any) is [-nasal].

Applied to an input with a geminate plain nasal in the consonant centre,

(304) describes intermediate mappings such as /eannii/ → edannii ‘mother’, or

/spoanii/ → spoanii ‘snow which is hard to drive on’.

Prestopping feeds a rule of Nasal Delinking , (305), the effect of which is

that the nasal now surfaces as the onset of the unstressed syllable of the foot.

(305)  \text{Nasal Delinking \text{(rule)}}

\[
\begin{array}{c}
\mu_n \\
\text{[-nas]} \text{P1} \text{[+nas]} \\
\end{array} \rightarrow \begin{array}{c}
\mu_n \\
\text{[-nas]} \text{P1} \text{[+nas]} \\
\end{array}
\]

By Nasal Delinking, intermediate [edannii] and [spoanii] map to eadnii and spoanii respectively.

Undoubtedly the trickiest question posed by prestopped nasals is the alternation in voicing in the prestop component. Descriptively, the value for voicing on the prestop is dependent on morphological grade. In the Strong Grade, the prestop is voiced, whereas in the Weak Grade, the prestop is voiceless. Thus, on the surface, eadnii ‘mother’, with a voiced prestop, but spoanii ‘snow which is hard to drive on’, with a voiceless prestop.

Neutralization of the voicing contrast in prestops is bidirectional, and neither neutralization context (Q3 or Q2) leaps out as being the Elsewhere context. For these reasons, it is difficult to see whether the preferred analysis should take the voiceless prestop as basic (and deploy a Voicing rule in the relevant context), or take the voiced stop as basic (and deploy a Devoicing rule in the complement context). Judging from the evidence of inchoate prestopping in Sea Saami dialects such as Kaakkuri (Ravila 1932: 35), preoclusion took the form of the insertion of a voiced stop historically. On these grounds it is reasonable to take the voiced stop as basic, and for the sake of simplicity, I will assume that there is a rule of Prestop Devoicing which operates in Q2.

(306)  \text{Prestop Devoicing \text{(rule)}}

\[
\begin{array}{c}
\mu_n \\
\text{[-son]} \text{[-cont]} \\
\end{array} \rightarrow \begin{array}{c}
\mu_n \\
\text{[-voice]} \text{[+nas]} \text{[-cont]} \\
\end{array}
\]

Condition: target is in Q2.
Consider the derivation of *skāappend* `Polar night-LOC.SG` in (307). The derivation of the corresponding nominative singular form is also supplied for comparison.

**Table (307)**

<table>
<thead>
<tr>
<th>Input</th>
<th>/skāappend+s/</th>
<th>/skāappend/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>n/a</td>
<td>skāappend</td>
</tr>
<tr>
<td>Prestopping</td>
<td>skāappmas</td>
<td>skāappma</td>
</tr>
<tr>
<td>Nasal Delinking</td>
<td>skāabmas</td>
<td>skāabma</td>
</tr>
<tr>
<td>Prestop Devoicing</td>
<td>skāappmas</td>
<td>n/a</td>
</tr>
<tr>
<td>Output</td>
<td>skāappmas</td>
<td>skāabma</td>
</tr>
</tbody>
</table>

Balance counterbleeds Devoicing in (308), resulting in the overapplication of Devoicing. This is a case of rule sandwiching again: Devoicing is bled by one rule (Fortition), and counterbled by a subsequent rule (Balance), but both Balance and Fortition introduce identical faithfulness violations. This is illustrated in (308) for *piepmumu* `food-LOC.SG`. The derivation of the corresponding nominative singular form is also supplied.

**Table (308)**

<table>
<thead>
<tr>
<th>Input</th>
<th>/piepmumu+s/</th>
<th>/piepmumu/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>n/a</td>
<td>piepmumu</td>
</tr>
<tr>
<td>Prestopping</td>
<td>piepmummas</td>
<td>piepmummu</td>
</tr>
<tr>
<td>Nasal Delinking</td>
<td>piepmumu</td>
<td>piepmumu</td>
</tr>
<tr>
<td>Devoicing</td>
<td>piepmumu</td>
<td>n/a</td>
</tr>
<tr>
<td>Balance</td>
<td>piepmumu</td>
<td>n/a</td>
</tr>
<tr>
<td>Output</td>
<td>piepmumu</td>
<td>piepmumu</td>
</tr>
</tbody>
</table>

The same overapplication of Devoicing is found in `primary lengthening` in the Strong Grade. Preceding an underlyingly long a Q1 nasal in the Weak Grade alternates with a nasal preceded by a voiceless stop in Q3 in the Strong Grade as in (309).

**Table (309)**

<table>
<thead>
<tr>
<th>Input</th>
<th>/imii/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>imii</td>
</tr>
<tr>
<td>Prestopping</td>
<td>i3mimii</td>
</tr>
<tr>
<td>Nasal Delinking</td>
<td>ilmii</td>
</tr>
<tr>
<td>Devoicing</td>
<td>ipmiii</td>
</tr>
<tr>
<td>Balance</td>
<td>ipmiii</td>
</tr>
<tr>
<td>Output</td>
<td>ipmiii</td>
</tr>
</tbody>
</table>

*(310)* gives the derivation of *ipmiii* `wife of uncle`.

**Table (310)**

<table>
<thead>
<tr>
<th>Input</th>
<th>/imii/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>imii</td>
</tr>
<tr>
<td>Prestopping</td>
<td>i3mimii</td>
</tr>
<tr>
<td>Nasal Delinking</td>
<td>ilmii</td>
</tr>
<tr>
<td>Devoicing</td>
<td>ipmiii</td>
</tr>
<tr>
<td>Balance</td>
<td>ipmiii</td>
</tr>
<tr>
<td>Output</td>
<td>ipmiii</td>
</tr>
</tbody>
</table>
An alternative is to relate the distribution of voicing to Coda Maximization. The observed distribution does have a precedent in the clusters of voiced+voiced stop. In these clusters, the voiced stop deletes as a consequence of Coda Maximization. Coda Maximization creates a tautosyllabic cluster of stops which differ in their laryngeal specification. The parallel is suggestive, and a tentative rule-based proposal is included here. The proposal turns on there being a second prestopping rule, which applies following Nasal Delinking. Prestopping\(_2\) applies to insert a voiceless prestop following a voiced stop. The rule is essentially one of partial devoicing, although it is interpreted here as fully phonological, consisting in the insertion of an actual root node. For the sake of concreteness, the epenthesized voiceless stop is shown in the derivation in (311) as parsed into the onset of the unstressed syllable, at least initially. Following Prestopping\(_2\), Coda Maximization applies. At the point at which it applies it scans a triliteral cluster, and its application results in the attraction of the medial consonant of the cluster over into the coda of the stressed syllable. Thus, by this process, intermediate [s'poad.t'na] is mapped to intermediate [s'poad.t'na]. This has the crucial effect of triggering an application of Homorganic Deletion, which deletes a voiced stop preceding a tautosyllabic homorganic stop with a conflicting laryngeal specification. In the corresponding Q3 form, [pieb\(\text{mn}\)uu], Coda Maximization cannot apply, and so the issue of Homorganic Deletion does not arise. The result: [pieb\(\text{mn}\)uu] surfaces with the voicing of the prestop intact, while [s'poot.na] surfaces with a voiceless prestop. (311) gives the derivations of pieb\(\text{mn}\)uu and s'poot.na based on this interpretation of the data.

(311)

<table>
<thead>
<tr>
<th>Input</th>
<th>/pieb(\text{mn})uu/</th>
<th>/s'poot.na/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>pieb(\text{mn})uu</td>
<td>s'poot.na</td>
</tr>
<tr>
<td>Prestopping(_1)</td>
<td>pieb(\text{mn})uu</td>
<td>s'poad.na</td>
</tr>
<tr>
<td>Nasal Delinking</td>
<td>pieb(\text{mn})uu</td>
<td>s'poad.na</td>
</tr>
<tr>
<td>Prestopping(_2)</td>
<td>pieb:(\text{mn})uu</td>
<td>s'poad.t'na</td>
</tr>
<tr>
<td>Coda Maximization</td>
<td>n/a</td>
<td>s'poad.t'na</td>
</tr>
<tr>
<td>Homorganic Deletion</td>
<td>n/a</td>
<td>s'poot.na</td>
</tr>
</tbody>
</table>

The claim is, then, that there derivational origin of the voiced stop and the voiceless stop are different, and so they do not stand in a relationship of correspondence: the voiced stop is introduced by Prestopping\(_1\), while the corresponding voiceless stop is introduced by Prestopping\(_2\). This has the additional advantage of being able to relate the the observed distribution to other robust givens of the language, to wit, Coda Maximization, and Homorganic Deletion, both of which are independently motivated. Suggestive as it is, though, the proposed analysis does leave a couple of questions unanswered. For example, it is not clear what the motivation for Prestopping\(_2\) might be.

Following a post-epenthesized sonorant in the Strong Grade, prestopping also takes the form of the insertion of a voiceless stop, as in (312), although it is not clear that this kind of prestopping is driven by the same conditions as in (311).

(312)  
lie\(\text{d.}\)o.pmi  
jal.la.qaat\(\text{h} \)  
‘leaf’  
‘tree stumps’
In the corresponding Weak Grade forms there is neither vowel nor consonant enepentheses, as illustrated by (313).

(313) lieönmniih
    aajmmnuus
    vuojmniah
    jälpmiis
    ifnmiih

‘leaves’
‘air, loc.sg’
‘spirits’
‘tree stump’
‘colours’

The behaviour of the Weak Grade forms suggests that it is the post-vocalic environment that is the trigger of Prestopping in (312). Since Epenthesis does not apply in the Weak Grade (because of Coda Maximization), we don’t get Prestopping either.

R-initial clusters evince epenthesis of a homorganic stop in the Weak Grade also, as in (314).

(314) fierpmiih
    paarntiin
    koarcmuun

‘nets’
‘sons’
‘I climb’

In the Strong Grade, the examples in (314) correspond to fier.ro.pmii ‘net’, paar.ro.tni ‘boy’, and koar.ro.cmua ‘to climb’ respectively. Compare the derivations for tfal.lb.pmii ‘eye’ and tfalm:miis ‘eye-loc.sg’ in (315) below.

(315)

<table>
<thead>
<tr>
<th>Input</th>
<th>/tfalmii/</th>
<th>tfalmii+s/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>tfalmii</td>
<td>n/a</td>
</tr>
<tr>
<td>Prestopping</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Coda Maximization</td>
<td>n/a</td>
<td>tfalmmiis</td>
</tr>
<tr>
<td>Balance</td>
<td>n/a</td>
<td>tfalmmiis</td>
</tr>
<tr>
<td>Epenthesis</td>
<td>tfal.lb.mii</td>
<td>n/a</td>
</tr>
<tr>
<td>Prestopping2</td>
<td>tfal.lb.pmii</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Sammallahiti (1977: 42) states that prestopping in sonorant-initial clusters is a secondary development. Indeed, there are dialects which still implement prestopping postvocalically, but not in a cluster. Jukkasjärvi is one such dialect (Collinder 1949: 145), e.g. kurpmaa ‘larva of gadfly Oestrus tarandi’, porpnaasah ‘part of knife which fits into handle (nom.pl)’, but perbammu ‘food’.

Let’s return to the issue of voicing in the prestops of underlyingly geminate nasals. Despite the intuitiveness of the analysis in (311), the issue remains unresolved, since it appears that a case may be made for the Devoicing analysis on dialectological grounds. The Skånland data in Jernsletten (n.d.: 67) are especially interesting in this regard. In Skånland, Coda Maximization is extended to
7.1. BISYLLABIC STEMS

all clusters, not just those which are sonorant-initial, giving alternations such as *jakʰtsa*—*jukttsa* ‘gruel, nom.sg—gen.sg and *puokʰtsa*—*puokʰtsama* ‘tongue, nom.sg—gen.sg. Significantly, we find the same pattern of devoicing occurring here in Q2. Consider the data in (316).

\[(316)\]  
Devoicing in Skånland

```
kabna  ‘fright’
napplntem ‘I name’
uoktmpn ‘I sell’
upplnu ‘bumble-bee’
upplnu ‘bumble-bee (gen.sg)’
upplnu ‘I sell’
upplnu ‘I sell’
upplnu ‘bumble-bee (gen.sg)’
upplnu ‘I sell’
upplnu ‘I sell’
upplnu ‘I sell’
upplnu ‘I sell’
upplnu ‘I sell’
upplnu ‘I sell’
upplnu ‘I sell’
``` 

In Skånland, we get devoicing of an underlyingly voiced stop in Q2. The motivation would seem to have nothing to do with the segmental environment: devoicing takes place irrespective whether the following consonant is voiceless or voiced, indicating that the quantity alone is responsible. I leave the resolution of these issues to future research.

Overlaid on this system is are two further rules: (i) a Prestop Glottalization rule inserting the specification [+constricted glottis] (or ?) on a (voiceless) prestop, and (ii) a Nasal Spreading rule spreading [+nasal] onto a preceding voiced prestop.

Appended to the derivation in (311) above, we get *pêkpp?nu* and *spøat? na*.

Finally, for most of today’s speakers, the prenasal voiced stop in the Strong Grade has become nasalized by a rule of Nasal Spreading (Sammallahti 1998b: 51f.) as in (317).

\[(317)\]  
```
komm?ni ‘ghost’
ponn?ni ‘bottom’
eann?ni ‘mother’
``` 

The relevant rule is formulated in (318).

\[(318)\]  
**Nasal Spread (rule)**

\[
\begin{array}{c}
\mu \mu \\
\sigma \sigma \\
\end{array} 
\rightarrow 
\begin{array}{c}
\mu \\
\mu \\
\end{array} 
\] 

\[
\begin{array}{c}
\mu \mu \\
\sigma \sigma \\
\end{array} 
\rightarrow 
\begin{array}{c}
\mu \\
\mu \\
\end{array} 
\] 

7.1.9 The opaque interaction of Balance and Lactic Lenthening

Balance interacts opaque with a rule which lengthens an underlyingly short /a/ in foot-final position following a consonant centre in Q1 or Q2. (Q3 blocks the lengthening.) A-Lengthening is given in (319).
A-Lengthening (rule)

\[
a \rightarrow aa\left/ \_r_1 \right.
\]

Condition: consonant centre is not in Q3.

A-Lengthening creates the environment which conditions Balance. Yet, Balance does not apply. A-Lengthening counterfeeds Balance, and must be ordered subsequent to it in the rule sequence. In the examples below, (320) illustrates the transparent application of Balance preceding an underlyingly long vowel. In (320a), Balance applies normally, preceding a surface long low vowel which is underlyingly long. In (320b), Balance applies preceding a surface front high vowel which is underlyingly long. (321) illustrates the corresponding opaque pattern. Balance is properly conditioned on the surface, yet fails to apply, since the vowel is underlyingly short.

(320) Transparent application of Balance

a. loniis~lotmaasah  `ransom~NOM.PL'
tal.la.pmii~tjalmaasaf  `eye~DIM'
lot:tiis~lot:taa$  `bird~DIM'
b. kob:miis~kop:miih  `ghost~NOM.PL'
sor.ro.pmii~sorp:miih  `accidental death ~NOM.PL'

(321) Underapplication of Balance

a. sonaas~sonaasah  `shrunken~NOM.PL'
tjalmaas  `reticum
pod:taa~pottaaf  `while~DIM'
b. t:jibp:ii~t:jippiif  `knee~ACC/GEN.PL'
sor.ro.pmii~sorp:miif  `accidental death ~ACC/GEN.PL'
kir.ro.tiif~kirttii  `to fly ~ he flew'

The underlying forms of the stems in (321) are, by hypothesis, /sonaas/, /tjalmaas/, and /podta/ respectively. The first two are underlyingly trisyllabic. For the phonology of these stems, see §7.3 below.

The underapplication of Balance gives rise to near minimal pairs such as tfjalmaa$ `eye~DIM' vs. tfjalmaas. In the latter form, the vowel forming the nucleus of the second syllable is underlyingly short, and the surface pattern of underapplication reflects this. In rule-based terms, the a-Lengthening rule counterfeeds Balance, rendering Balance opaque, as shown in the derivation in (322).

(322)

| Input         | /lodtii+/ | /podta+/
|---------------|-----------|-----------
| Coda Maximization | lodttaaf | podttaaaf |
| Balance       | lodt:taaf | n/a       |
| a-Lengthening | n/a       | podttaaaf |
| Homorganic Deletion | lot:taaf | pottaaf |
| Output        | [lot:taaf] | [pottaaf] |
Consider the derivation of *kirjjaah* ‘books’ in (323).

(323)

<table>
<thead>
<tr>
<th>Input</th>
<th>kirja+h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coda Maximization</td>
<td>kirjjaah</td>
</tr>
<tr>
<td>Balance</td>
<td>n/a</td>
</tr>
<tr>
<td>a-Lengthening</td>
<td>kirjjaah</td>
</tr>
<tr>
<td>Output</td>
<td>[kirjjaah]</td>
</tr>
</tbody>
</table>

As (323) makes clear, Balance is not properly conditioned at the point it applies, since the vowel in the unstressed syllable is not long.

Kautokéino has an additional rule, *j*-vocalization which interacts with Balance in a similar way.

(324) **J-Vocalization (rule)**

\[
ij \rightarrow ii/\!
\]

Consider the derivation of *mojvviit* ‘muddle-acc/gen.pl’ from *mojvi*. Note that underlyingly long /ii/ is subject to an additional rule of Pre-yotic shortening, which bleeds Balance of its environment.

(325) **Preyotic i-Shortening**

\[
ii \rightarrow i/\!
\]

(326)

<table>
<thead>
<tr>
<th>Input</th>
<th>mojvviit+jt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coda Maximization</td>
<td>mojvviit</td>
</tr>
<tr>
<td>Preyotic i-Shortening</td>
<td>mojvviit</td>
</tr>
<tr>
<td>Balance</td>
<td>n/a</td>
</tr>
<tr>
<td><em>j</em>-Vocalization</td>
<td>mojvviit</td>
</tr>
<tr>
<td>Output</td>
<td>[mojvviit]</td>
</tr>
</tbody>
</table>

Although *j*-Vocalization creates the conditions for Balance to apply, we get underapplication. Balance must therefore precede *j*-Vocalization. Along with the rule of *a*-Lengthening, the *j*-Vocalization rule has consolidated the phonologization of the Weak Grade opposition between Q2 and Q3. Thus we get (near) minimal pairs in Kautokeino such as *tsámänhii* ‘eyes’ vs. *tsämäniin* ‘in the eyes’, *koítiih* ‘wild reindeer -nom.pl’ vs. *koítii* ‘wild reindeer -acc/gen.pl’, *maajstii* ‘by tasting’ vs. *maaajstii* ‘he tasted’, and *luajhti* ‘by scolding’ vs. *luajhti* ‘he scolded’ (Sammallahi 1984: 139).

The East Enontekiö dialect lacks this rule, so the absence of Balance in these cases is as expected (the corresponding forms in East Enontekiö are *mojvviit, tsámänjii, koítii, maajstii, luajhti*).

### 7.1.10 Preyotic shortening and rule sandwiching

The two dialects also differ in another respect. While both dialects have the Preyotic i-Shortening rule, which bleeds Balance, Kautokeino has an additional rule of Preyotic u-Shortening. Interestingly, though, Preyotic u-Shortening is
ordered after the application of Balance, thereby counterbleeding it, and furnishing yet another instance of rule-sandwiching. As a consequence, there are, on the surface, two kinds of lactic uj: uj from an underlying combination of short /oj/ or /uj/, and uj from underlying /uuj/. Only the latter triggers Balance. (327) illustrates the overapplication of Balance preceding a short vowel which is underlingly long. (328) gives the normal application of Balance preceding a short vowel which is also underlingly short.

(327)  *Transparent application of Balance*

\[
\begin{align*}
\text{krol.kuu} & \rightarrow \text{kolk:kujt} & \text{‘sauntering~ACC/GEN.PL’} \\
\text{tjul.lo.kuu} & \rightarrow \text{tjul:kujt} & \text{‘clog on animal ~ACC/GEN.PL’} \\
\text{jodtuu} & \rightarrow \text{jot:kujt} & \text{‘sequence of nets ~ACC/GEN.PL’}
\end{align*}
\]

(328)  *Underapplication of Balance*

\[
\begin{align*}
\text{hel.lo.p?mo} & \rightarrow \text{heelm:ujujt} & \text{‘pearl~ACC/GEN.PL’} \\
\text{pel.lo.hko} & \rightarrow \text{peelh:kujujn} & \text{‘chopping-block~LOC.PL’}
\end{align*}
\]

Compare the derivation of *mojvviit* ‘muddle-ACC/GEN.PL’ above with that of *kolk:kujt* ‘sauntering-ACC/GEN.PL’ and *heelm:ujujt* ‘pearl-ACC/GEN.PL’ in (329).

(329)

\[
\begin{array}{cccc}
\text{Input} & /\text{mojvviit}+j\text{t}/ & /\text{kolk:kujt}+j\text{t}/ & /\text{heelm:ujujt}+j\text{t}/ \\
\hline
\text{Coda Maximization} & \text{mojvviit} & \text{kolk:kujt} & \text{heelm:ujujt} \\
\text{Preyotic i-Shortening} & \text{mojvviit} & n/a & n/a \\
\text{Balance} & n/a & \text{kolk:kujt} & n/a \\
\text{Preyotic u-Shortening} & n/a & \text{kolk:kujt} & n/a \\
\text{j-Vocalization} & \text{mojvviit} & n/a & n/a \\
\hline
\text{Output} & \text{[mojvviit]} & \text{[kolk:kujt]} & \text{[heelm:ujujt]}
\end{array}
\]

The configuration is one of rule-sandwiching. We have two rules, Preyotic i-Shortening and Preyotic u-Shortening, both of which introduce identical Faithfulness violations of WT-IDENT(V), yet Balance applies crucially in between the two rules.

### 7.1.11 Other rules

**Pre-Q3 Vowel Shortening**

Vowels are shortened preceding a Q3 consonant centre. Long monophthongs surface as short (330), while diphthongs surface as short rising diphthongs with the prominence on the second component of the diphthong (331).


(331)  *kiod:tiih*  ‘carry’

\[
\begin{align*}
\text{kiod:tiih} & \rightarrow \text{‘carry’} \\
\text{kied:tiih} & \rightarrow \text{‘field’} \\
\text{søab:piii} & \rightarrow \text{‘staff’} \\
\text{veæed:tiih} & \rightarrow \text{‘keep tethered’}
\end{align*}
\]
Note, however, that long /aa/ does not undergo the process. This is shown in (332).

(332) vaalstihih, *valstihih ‘to take’

Pre-Q3 Vowel Shortening is formulated here as a disjunction of near-identical rules. (333a) shortens a non-low vowel preceding a Q3 consonant centre. (333b) reduces a long rising diphthong to short, making the first component of a diphthong into a vocalic on-glide.

(333) **Pre-Q3 Vowel Shortening**

a. \( V_{-\text{low}}: \rightarrow \tilde{V}_{-\text{low}}/\_\_C:C \)

b. \( V_1V_2: \rightarrow \tilde{V}_1V_2/\_\_C:C \)

Pre-Q3 Vowel Shortening is a point of variation between dialects, since in the Girdzh variety of Torne Saami, also a North Saami dialect, the contrast between long /ii ee oo uu/ and short /i e o u/ preceding Q3 clusters is retained (Sammallahti 1998b: 20). Thus *oos:ten* ‘I bought’, *vuul:ken* ‘I left’. In West Finnmark Saami, these forms are rendered oosten and vulken respectively.

**Supralaryngeal Spreading**

In triliteral sonorant-initial clusters in which the second member is /h/, the laryngeal glide acquires the supralaryngeal specification of the sonorant through a rule of Supralaryngeal Spreading. In West Finnmark Saami, this only occurs in homorganic clusters such as /lht, mhsp, nhbt, . . . /, giving surface clusters [l̩t, m̩sp, n̩ht, . . . ]. In heterorganic clusters such as /lhk, rhp, . . . /, assimilation does not take place, although this is attested in Utsjoki (Pekka Sammallahi, p.c.). (334) illustrates the failure to spread in heterorganic clusters.

(334) **Heterorganic clusters with medial /h/ (no assimilation)**

| paalh.kas | ‘wage, LOC.SG’ | paal.ha.hkaa | ‘wage, NOM.SG’ |
| kart.pas | ‘frivolous’ | kar.ro.hpii | ‘frivolous person’ |

The homorganic case is given in (335).

(335) **Homorganic clusters with medial /h/ (assimilation)**

| pol:ltuun | ‘rummaging’ | polltuun | 1SG.PRES |
| kum:npe | ‘wolf, NOM.SG’ | kumn:peh | NOM.PL |
| par:ta | ‘living-room, NOM.SG’ | par.tas | LOC.SG |
| koan:juku | ‘stock for hanging’ | koanj.kuuh | NOM.PL |
| caulron, NOM.SG’ |

(335) shows that supralaryngeal spreading occurs in all grades and quantities when the cluster is homorganic. Spreading takes place from left to right, the features of the sonorant spreading onto the root-node occupied by /h/. In Sammallahi’s phonemization, the words in (335) are represented as polltuun, polltuun, kum:npe, and so on.
7.2 Quadrisyllabic stems

The phonology of quadrisyllabic stems have no more to tell us about the phonology of Grade Alternation, but they do evince some extra phonological alternations in which bisyllabic stems do not participate. We review these here briefly.

In the nominative singular, Apocope applies to delete the final vowel. This has ramifications for the preceding consonant as well, since geminates and some clusters are precluded from occurring in domain-final position.

Complete nominal paradigms are given here for *poarraaseabpō*—*poarraaseh* 'the older (one)' in (336), *mujhtalus* 'story' in (337), and *saapmelajf* 'Saami (person)' in (338). Note that the distribution of morphological grade within the paradigm is the same as in (245), abstracting away from the apocope operating in the nominative singular.

\[(336)\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>poarraaseabpō</td>
<td>poarraaseappoh</td>
</tr>
<tr>
<td></td>
<td>=poarraaseh</td>
<td></td>
</tr>
<tr>
<td>ACC/GEN</td>
<td>poarraaseappoh</td>
<td>poarraaseappujt</td>
</tr>
<tr>
<td>ILL</td>
<td>poarraaseappuj</td>
<td>poarraaseappujō</td>
</tr>
<tr>
<td>LOC</td>
<td>poarraaseappos</td>
<td>poarraaseappujn</td>
</tr>
<tr>
<td>COM</td>
<td>poarraaseappujn</td>
<td>poarraaseappujkujn</td>
</tr>
<tr>
<td>ESS</td>
<td></td>
<td>poarraaseabpō</td>
</tr>
</tbody>
</table>

\[(337)\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>mujhtalus</td>
<td>mujhtalusah</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC/GEN</td>
<td>mujhtalus</td>
<td>mujhtalusajt</td>
</tr>
<tr>
<td>ILL</td>
<td>mujhtalusīi</td>
<td>mujhtalusajō</td>
</tr>
<tr>
<td>LOC</td>
<td>mujhtalusas</td>
<td>mujhtalusajn</td>
</tr>
<tr>
<td>COM</td>
<td>mujhtalusajn</td>
<td>mujhtalusajkujn</td>
</tr>
<tr>
<td>ESS</td>
<td></td>
<td>mujhtalusān</td>
</tr>
</tbody>
</table>

\[(338)\]

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>saapmelajf</td>
<td>saapmelattfah</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC/GEN</td>
<td>saapmelattfaj</td>
<td>saapmelattfajt</td>
</tr>
<tr>
<td>ILL</td>
<td>saapmelattfajii</td>
<td>saapmelattfaiō</td>
</tr>
<tr>
<td>LOC</td>
<td>saapmelattfas</td>
<td>saapmelattfajn</td>
</tr>
<tr>
<td>COM</td>
<td>saapmelattfajn</td>
<td>saapmelattfajkujn</td>
</tr>
<tr>
<td>ESS</td>
<td></td>
<td>saapmelattfjan</td>
</tr>
</tbody>
</table>

The rule of Apocope is given in (339). Apocope is conspicuously lacking in the corresponding genitive/accusative singular form. The question arises whether Apocope is morphologically or phonologically conditioned. If morphologically conditioned, we must directly encode the fact that only the nominative singular form undergoes the rule. Since we have argued for the adoption of the Ghost Hypothesis in (13), which requires us to posit a syllable-closing
underlying consonant in the lexical representations of (possibly null) suffixes which unexpectedly subcategorize for the Weak Grade, we can take advantage of this here in the following way. Apocope underapplies in the vowel-final accusative/genitive singular on the assumption that it is ordered prior to Ghost Deletion. On this interpretation, Apocope is not properly conditioned at the point in the derivation it applies because the form is consonant-final.

Apocope is blocked from applying to the latus vowel of a bisyllabic, since this would violate the Minimality condition on word-size that words consist minimally of two syllables.

\begin{align*}
(339) & \text{ Apocope (rule)} \\
& V \rightarrow \emptyset \_{PrWd} \\
& \text{Condition: V is not in the head foot of PrWd.}
\end{align*}

\( (339) \) feeds another rule, \( (340) \), which degeminites a word-final geminate.

\begin{align*}
(340) & \text{Final Degemination (rule)} \\
& C_{\mu} \rightarrow C \_{PrWd}
\end{align*}

The derivation for the nominative singular form of \(/\text{mujhtalus}\) is given in \( (341) \), showing the interaction of \( (339) \) and \( (340) \).

\begin{align*}
\text{Input} /\text{mujhtalus}/ & \quad \text{Apocope} \rightarrow \text{mujhtuss} \\
& \quad \text{Word-final Degemination} \rightarrow \text{mujhtalus} \\
\text{Output} & \rightarrow [\text{mujhtalus}]
\end{align*}

Compare this with the derivation of the corresponding genitive/accusative singular form, which, by hypothesis, is C-final at the crucial stage at which Apocope applies.

\begin{align*}
\text{Input} /\text{mujhtalus}C/ & \quad \text{Apocope} \rightarrow n/a \\
& \quad \text{Ghost Deletion} \rightarrow \text{mujhtalus} \\
\text{Output} & \rightarrow [\text{mujhtalus}]
\end{align*}

Ghost Deletion creates the environment in which Apocope can apply, yet Apocope does not apply, giving a surface form which is vowel final. Ghost Deletion counterfeeds Apocope.

### 7.3 Trisyllabic stems

Trisyllabic stems come in three main types which we can classify according to the final segmentism of the nominative singular and accusative/genitive singular forms.
1. CVCV~CVCCVCV, nom.sg is V-final, e.g. 
   peana~pet?naka 'dog: nom~acc/gen'
2. CVCCVC~CVCCVCV, nom.sg is C-final, e.g. 
   tsielus~tsiilos 'abuse: nom~acc/gen'
3. Non-grade-alternating CVCCVC~CVCCVC(V), e.g. 
   kahpiir~kahpiir(a) 'hat: nom~acc/gen'

The first two types both evince a V~∅ alternation: the vowel is preserved in the accusative/genitive singular, but apocopated in the corresponding nominative singular form. The rule responsible is Apocope, in (339) above.

Nouns of the peana-type evince an additional C~∅ alternation: the nominative singular apocopates not only the final vowel, but also the next consonant in. Consonant Apocope is apparently an unpredictable property of the lexical item. (343) furnishes some more examples of nouns of this type.

(343) Trisyllabic stems with nom.sg in V

<table>
<thead>
<tr>
<th>Stem</th>
<th>Nom.sg</th>
<th>Acc/Gen.sg</th>
</tr>
</thead>
<tbody>
<tr>
<td>pet?naka</td>
<td>peana</td>
<td>pet?naka</td>
</tr>
<tr>
<td>luo&gt;pmiina</td>
<td>luo&gt;pmiina</td>
<td></td>
</tr>
<tr>
<td>kieluumaa</td>
<td>kieluu</td>
<td>kielooma</td>
</tr>
</tbody>
</table>

Further examples of nouns with a consonant-final nominative singular form are given in (344).

(344) Trisyllabic stems with nom.sg in C

<table>
<thead>
<tr>
<th>Stem</th>
<th>Nom.sg</th>
<th>Acc/Gen.sg</th>
</tr>
</thead>
<tbody>
<tr>
<td>oahpiisa</td>
<td>oahpis</td>
<td>oahpaasa</td>
</tr>
<tr>
<td>vuolluSa</td>
<td>vuoluu</td>
<td>vulloSa</td>
</tr>
<tr>
<td>nuerapu</td>
<td>nuerah</td>
<td>nuerapu</td>
</tr>
</tbody>
</table>

A striking feature of the trisyllabic paradigm is the prevalence of the Strong Grade. Also striking is the low incidence of morphological conditioning. The nominative singular is in the Weak Grade for reasons which are surface-apparent (except for stems which specify for Final Consonant Deletion, such as peana).

Only the essev form surfaces unexpectedly in the Weak Grade (the essev suffix closes the syllable, predicting Strong Grade). Otherwise, the Strong Grade recurs throughout the paradigm in a way which is phonologically entirely transparent. These relationships are summarized in (345). Compare (245).

(345) Relationship between grade and morphosyntactic category in trisyllabic nouns

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>nom</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>acc/gen</td>
<td>Strong</td>
<td>Strong</td>
</tr>
<tr>
<td>ill</td>
<td>Strong</td>
<td>Strong</td>
</tr>
<tr>
<td>loc</td>
<td>Strong</td>
<td>Strong</td>
</tr>
<tr>
<td>com</td>
<td>Strong</td>
<td>Strong</td>
</tr>
<tr>
<td>essev</td>
<td>Weak</td>
<td></td>
</tr>
</tbody>
</table>
The declension of three consonant-deleting trisyllabic stems are given in full below. These are *peana* 'dog' in (346), *luomii* 'cloudberry' in (347), and *kielu* 'bloodclot' in (348).

(346)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>peana</td>
<td>peatnakah</td>
</tr>
<tr>
<td>GEN/ACC</td>
<td>peatnaka</td>
<td>peatnakijt</td>
</tr>
<tr>
<td>ILL</td>
<td>peatnakij</td>
<td>peatnakijtta</td>
</tr>
<tr>
<td>LOC</td>
<td>peatnakis</td>
<td>peatnakijn</td>
</tr>
<tr>
<td>COM</td>
<td>peatnakijn</td>
<td>peatnakijkujn</td>
</tr>
<tr>
<td>ESS</td>
<td>peanian</td>
<td></td>
</tr>
</tbody>
</table>

(347)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>luomii</td>
<td>luopmaanah</td>
</tr>
<tr>
<td>GEN/ACC</td>
<td>luopmaana</td>
<td>luopmaanijt</td>
</tr>
<tr>
<td>ILL</td>
<td>luopmaanij</td>
<td>luopmaanijtta</td>
</tr>
<tr>
<td>LOC</td>
<td>luopmaanis</td>
<td>luopmaanijn</td>
</tr>
<tr>
<td>COM</td>
<td>luopmaanijn</td>
<td>luopmaanijkujn</td>
</tr>
<tr>
<td>ESS</td>
<td>luomiin</td>
<td></td>
</tr>
</tbody>
</table>

(348)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>kieluu</td>
<td>kiillomaah</td>
</tr>
<tr>
<td>GEN/ACC</td>
<td>kiilloma</td>
<td>kiillomijt</td>
</tr>
<tr>
<td>ILL</td>
<td>kiillomij</td>
<td>kiillomijtta</td>
</tr>
<tr>
<td>LOC</td>
<td>kiillomis</td>
<td>kiillomijn</td>
</tr>
<tr>
<td>COM</td>
<td>kiillomijn</td>
<td>kiillomijkujn</td>
</tr>
<tr>
<td>ESS</td>
<td>kieluun</td>
<td></td>
</tr>
</tbody>
</table>

The second class of trisyllabic stem is identical to the first except in one respect: the nominative singular does not implement consonant deletion. Examples are *tsieluus* 'abuse' in (349), *oaliis* 'groove running along underside of ski' in (350), and *einan* 'country, land' in (351).

(349)

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM</td>
<td>tsieluus</td>
<td>tsiillosah</td>
</tr>
<tr>
<td>GEN/ACC</td>
<td>tsiillosa</td>
<td>tsiillosijt</td>
</tr>
<tr>
<td>ILL</td>
<td>tsiillosij</td>
<td>tsiillosijtta</td>
</tr>
<tr>
<td>LOC</td>
<td>tsiillosis</td>
<td>tsiillosijn</td>
</tr>
<tr>
<td>COM</td>
<td>tsiillosijn</td>
<td>tsiillosijkujn</td>
</tr>
<tr>
<td>ESS</td>
<td>tsieluusin</td>
<td></td>
</tr>
</tbody>
</table>

(350)
Consonant deletion is lexically determined. In rule-based terms, we can account for this by positing a rule feature \ [+C del] in the lexical specifications of forms such as *peana*.

\[(351)\]

\[
\begin{array}{l|l|l}
\text{SG} & \text{PL} \\
\hline
\text{NOM} & \text{callis} & \text{callaasah} \\
\text{GEN/ACC} & \text{callaasa} & \text{callaasijt} \\
\text{ILL} & \text{callaasij} & \text{callaasijtta} \\
\text{LOC} & \text{callaasis} & \text{callaasijn} \\
\text{COM} & \text{callaasijn} & \text{callaasijkujn} \\
\text{ESS} & \text{calliisn} & \\
\end{array}
\]

Consonant Apocope (rule)

\[C \rightarrow \emptyset\]  \_pTw\_d\_ for lexical items marked \ [+C del].

The derivation of the nominative singular and genitive/accusative singular forms of *peana* ‘dog’ is shown in (353).

\[(353)\]

<table>
<thead>
<tr>
<th>Input</th>
<th>/peanaka/</th>
<th>/peanaka+C/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apocope</td>
<td>peanak</td>
<td>n/a</td>
</tr>
<tr>
<td>Fortition</td>
<td>n/a</td>
<td>peanaka</td>
</tr>
<tr>
<td>Other rules</td>
<td>n/a</td>
<td>peanaka</td>
</tr>
<tr>
<td>Consonant Deletion</td>
<td>peana</td>
<td>n/a</td>
</tr>
<tr>
<td>Output</td>
<td>[peanak]</td>
<td>[peanaka]</td>
</tr>
</tbody>
</table>

In order to account for the alternation *tsielus*–*tsillosa*, we have to posit two extra rules. First, a rule of *u*-shortening, in (354), applies to /uu/ in the environment (\(\sigma\)) to give \(\ddot{u}\). Since short \(\ddot{u}\) is disallowed in latus position, it is ultimately repaired to short \(\ddot{o}\) by (356). However, prior to repair, the short high vowel causes Monophthongization (357) of the diphthong in the preceding syllable. There is a similar alternation between *ii* and *aa*, e.g. *callis*–*callaasa*, for which we may posit a rule of *i*-Lowering as in (355).

\[(354)\]  \*U-Shortening (rule)\*

\[uu \rightarrow u/VC_0\_C_0V\]

\[(355)\]  \*I-Lowering (rule)\*

\[ii \rightarrow aa/VC_0\_C_0V\]
7.3. TRISYLLABIC STEMS

(356) Latic Lowering (rule)

\[ \V [+\text{high}] \rightarrow \V \text{-lowering} ] /VC_0 \]

(357) Monophthongization (rule)

\[ V_1 V_2 \rightarrow V_1 : /C_0 \{ \text{u} \} \]

(358) shows the derivations of tsieluus ‘abuse-nom.sg’ tsilloosa ‘abuse-gen.sg’ from underlying /tsieluusa/ and /tsieluusa+C/ respectively.

(358)

<table>
<thead>
<tr>
<th>Input</th>
<th>/tsieluusa/</th>
<th>/tsieluusa+C/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apocope</td>
<td>tsieluus</td>
<td>n/a</td>
</tr>
<tr>
<td>Fortition</td>
<td>n/a</td>
<td>tsieluusaC</td>
</tr>
<tr>
<td>U-Shortening</td>
<td>n/a</td>
<td>tsiellusaC</td>
</tr>
<tr>
<td>Monophthongization</td>
<td>n/a</td>
<td>tsiillosaC</td>
</tr>
<tr>
<td>Latic Lowering</td>
<td>n/a</td>
<td>tsiillosa</td>
</tr>
<tr>
<td>Ghost Deletion</td>
<td>n/a</td>
<td>tsiillosa</td>
</tr>
<tr>
<td>Output</td>
<td>tsieluus</td>
<td>tsiillosa</td>
</tr>
</tbody>
</table>

Interestingly, some idiolects have introduced a further apocope process, Apocope_2, which apocopates a final vowel in the accusative/genitive singular form. Crucially, though, Apocope_2 stands in a counter-bleeding relationship to Fortition, in contrast to Apocope_1, which bleeds Fortition. This is rule-sandwiching again: Fortition is sandwiched in between two rules with identical effects for faithfulness.

Vowel-final accusative/genitive singular forms vary facultatively as in (359).

(359) Facultative variation in the accusative/genitive singular of trisyllabic stems

| tsiolloosa~tsillos          | ‘abuse-gen’ |
| oallaasa~oallaas            | ‘groove on underside of ski-gen’ |
| peatnakal~peatnah           | ‘dog-gen’ |

(360) illustrates how this works, with Apocope_2 appended at the end of the derivation.

(360)

<table>
<thead>
<tr>
<th>Input</th>
<th>/tsieluusa/</th>
<th>/tsieluusa+C/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apocope_1</td>
<td>tsieluus</td>
<td>n/a</td>
</tr>
<tr>
<td>Fortition</td>
<td>n/a</td>
<td>tsieluusaC</td>
</tr>
<tr>
<td>U-Shortening</td>
<td>n/a</td>
<td>tsiellusaC</td>
</tr>
<tr>
<td>Monophthongization</td>
<td>n/a</td>
<td>tsiillosaC</td>
</tr>
<tr>
<td>Latic Lowering</td>
<td>n/a</td>
<td>tsiillosa</td>
</tr>
<tr>
<td>Ghost Deletion</td>
<td>n/a</td>
<td>tsiillosa</td>
</tr>
<tr>
<td>Apocope_2</td>
<td>n/a</td>
<td>tsiillosa</td>
</tr>
<tr>
<td>Output</td>
<td>tsieluus</td>
<td>tsiillosa</td>
</tr>
</tbody>
</table>
7.4 Morphological conditioning vs. abstract phonology

As we observed in §7.1.1, Grade Alternation may be triggered even in the absence of syllable closure on the surface, for example, in the genitive/accusative singular. If openness of the syllable is still taken to be the synchronic phonological trigger of Fortition in the modern language, this means that we are dealing with an underapplication of Fortition in the genitive/accusative singular: Fortition is blocked, but the phonological reason is not apparent on the surface.

In §2.3, we rejected the notion that phonological alternations could be morphologically triggered, and adopted the Ghost Hypothesis, (13), that morphologized alternations always implied the presence of a covert phonological entity. Still, what would an account along morphological lines look like? An account in terms of morphological conditioning might start out on these lines: the blocker is some conjunction of morphosyntactic features, such as [+genitive, −singular] and, since these blockers are overtly present in the representation, Fortition is transparently blocked. This account forces us to allow that constraints make free reference to conjunctions of phonological and morphosyntactic properties. For concreteness, the blocking of Fortition in the accusative/genitive singular form of bisyllabic nouns could be captured in terms of a constraint conjunction between *Genitive and *Strong Grade. Assuming the genitive has to be realized under compulsion from some highly ranked constraint, say, Max[genitive], the reduction from Strong to Weak Grade is inexorable.

The proposal quickly runs into dire problems for both theoretical and language-internal reasons. Crushingly, the Weak Grade is not reliably correlated with the accusative/genitive singular. The correlation only obtains when the stem is underlyingly bisyllabic. When the stem is trisyllabic in the input, the relationship between grade and case is reversed, making any attempt to drive Grade Alternation by morphosyntactic specification a non-starter. These relationships are shown in (361), (362) and (363) below.

(361) Bisyllabic stem (nom.sg=strong; acc/gen.sg=weak)

\[
\begin{align*}
\text{kɔllii} & \quad \text{‘gold, nom.sg’} \\
\text{puoDDuu} & \quad \text{‘dam, nom.sg’}
\end{align*}
\]

(362) Trisyllabic stem (nom.sg=weak; acc/gen.sg=strong)

\[
\begin{align*}
\text{peana} & \quad \text{‘dog, nom.sg’} \\
\text{kielu} & \quad \text{‘bloodclot, nom.sg’}
\end{align*}
\]

(363) Contracted stem (nom.sg=weak; acc/gen.sg=strong)

\[
\begin{align*}
\text{pɔatsuu} & \quad \text{‘reindeer, nom.sg’} \\
\text{suoluu} & \quad \text{‘island, nom.sg’}
\end{align*}
\]

This non-correlativity essentially fells the morphological conditioning account, and it is difficult to see how it could be developed to account for the data.
7.5. **CONCLUSIONS**

On the theoretical side, there is the more serious issue of phonological autonomy to consider. Phonological constraints can only make reference to phonological information. By this token, conjunctions of phonological and morphosyntactic constraints, such as [*Genitive & Strong Grade*], should be impermissible in the first place.

### 7.5 Conclusions

The phonology of West Finnmark Saami is exceptionally rich in opaque processes, many of which interact in crucial ways. The main interactions here between processes are summarized here.

We saw that Fortition is to be preferred to Lenition on grounds of derivational economy.

Sonorant-initial clusters are subject to an obligatory rule of Coda Maximization. Coda Maximization is bled by Fortition, and so does not apply in the Strong Grade of a sonorant-initial cluster. Coda Maximization feeds Balance ('secondary lengthening'), which lengthens the second component of a sonorant-initial cluster in the Weak Grade preceding a long vowel.

Balance interacts with the rules governing the distribution of voicing in prestopped nasals. While it is not entirely clear whether the distribution of voicing is manipulated directly by rules or emerges from other considerations, it is clear that this distribution is opacified by Balance, since, in the overlengthened Weak Grade, the prestop is unexpectedly voiceless.

Balance and Geminate Fortition introduce identical Faithfulness violations (**Wt-Ident**), yet the distribution of voicing must be established at a derivational stage between the application of Fortition and the application of Balance, a case of rule sandwiching.

Balance itself is opacified by 'later' processes which counterfeed the rule. The relevant processes result in a derived long vowel in the latus, creating the environment for Balance to apply.

Aspiration is opacified by Homorganic Deletion.

North Saami seems surprisingly rich in rule interactions of the sandwiching type. In these interactions, some rule is crucially sandwiched in between two other processes which introduce identical faithfulness violations. Let’s just sum up the ones we’ve encountered.

- Fortition (between Apocope\(_1\) and Apocope\(_2\)),
- Devoicing (between Fortition and Balance),
- Balance (between Singleton Fortition\(_1\) and Singleton Fortition\(_2\)),
- Balance (between Preyotic i-Shortening and Preyotic u-Shortening)

Grade Alternation in trisyllabic stems holds yet another case of rule sandwiching involving two distinct applications of an identical Apocope rule (both implying violation of **Max-V**), with Fortition sandwiched in between.
Chapter 8

Representing quantity in North Saami

Quantity contrast and Grade Alternation are inextricably linked: indeed, as we have seen, the development of the Grade Alternation system in Proto-Saamic is the very origin of the three-way length contrast. The purpose of the present chapter is to develop a theory of the representation of quantity in West Finnmark Saami. This chapter seeks to propose solutions to these questions:

- How is the three-way length contrast to be represented?
- How do we represent hypercharacterized syllables?
- How is hyperchrony treated in higher-level prosody, i.e. how is it integrated into syllable and foot structure?

The representation of quantity must be addressed on two distinct levels: (i) how moras map onto segments (segmental quantity), and (ii) how moras are parsed into higher-level prosodic structure (prosodic quantity).

In §8.1, we deal with the segmental aspect of quantity. §8.2 places hyperchrony and hypercharacterization in their prosodic context and argues:

- Hypercharacterized syllables are bimoraic.
- Hyperchrony is formally represented as a minor syllable without nuclear content.

---

1 Hyperchrony is characteristic of nearly all of the Saamic languages. Of the five Western Saami languages recognized by Sammallahti (1998b: 6), Q3 is found in the surface phonologies of Ume, Pite, Lule, and North Saami (excluding East Finnmark), although it is completely absent in South Saami. This is no accident: South Saami also lacks any form of Grade Alternation.

In the Eastern Saami languages, Q3 is generally neutralized to Q2 in all environments. Interestingly, though, these languages still preserve three phonologically distinct grades. This absolute neutralization of Q3 to Q2 is found in Inari, Skolt, Akkala, and Kildin. The easternmost dialect, Ter, apparently preserves surface Q3 in the same way as the West Saami languages, though. A good case can be made for the abstract presence of Q3 in all of these languages, since the Geminate Strong Grade is associated with very similar vowel shortening phenomena found in the Q3 environment in languages with overt Q3.

173
These hypotheses are amply illustrated with full prosodic structures of key examples.

§8.3 addresses the interaction of vowel and consonant quantity, and examines in particular the shortening effect of a Q3 consonant centre on the preceding vowel.

§8.4 presents the main conclusions.

8.1 Segmental quantity

The theory of length contrast used here is the Moraic theory of Hyman (1985), McCarthy and Prince (1986), Hock (1986), and Hayes (1989). These researchers introduced the now widespread use of the mora to distinguish between singleton and geminate consonants.

Singleton consonants lack a mora (364a), while geminates (364b) are linked to a mora. Extending this idea to deal with the opposition between long (Q3) and short (Q2) geminates, geminates in Q3 will be assumed to associate to two moras as in (364c).

(364) Distinctions in lexical quantity in consonants
   a. Q1 /C/ kaaruu ‘by consenting’
   b. Q2 /C/ kaaruu ‘consent, 3sg.pres’
   c. Q3 /C/ kaaruu ‘consent, pres.part’

These contrasts are rendered autosegmentally as in (365).

(365) Moraic representation of three-way length contrast in singletons
   a. /C/ b. /CC/ c. /C:C/

   \[\begin{array}{c}
   \text{C} \\
   \text{C} \\
   \mu
   \end{array}\]

   Obstruent-initial clusters support a two-way contrast in length, as exemplified in (366).

(366) Moraic representation of three-way length contrast in biliteral obstruent-initial clusters
   a. /C1:C2/ b. /C1C2,C2/

   \[\begin{array}{c}
   \text{C1} \\
   \mu
   \end{array}\]

   In sonorant-initial clusters, the possibilities for quantity contrast are richer. These are summarized in (367).

(367) Moraic representation of three-way length contrast in biliteral sonorant-initial clusters
8.2 Prosodic quantity

West Finnmark Saami allows both hypercharacterized syllables and overlength. This section tackles the questions: (a) how is excess segmental structure accommodated within the syllable, and (b) how is excess moraic structure accommodated within the syllable?

(Bye 1997a,b) argues that syllables are universally subject to a bimoraic maximum. Overlength warrants the same kind of treatment as sesquisyllabic structures in general, i.e. as formally bisyllabic (Shaw 1993, 1996).

(369) **Maximal Binarity**
A syllable may dominate at most two moras.

Maximal Binarity is assumed to be hardwired into Gen, i.e. trimoraic syllables are universally banned.

8.2.1 Hypercharacterization

‘Hypercharacterized’ is Sherer’s term for any syllable which either (a) combines both a complex nucleus (in the form of a diphthong or a long vowel) and coda material in addition, or (b) contains a complex coda (Sherer 1994).

West Finnmark Saami permits an impressive degree of hypercharacterization of syllable structure, allowing hypercharacterized syllables of not only the \([VVC]_σ\), and \([VCC]_σ\) types, but \([VVCC]_σ\) and even \([VVCC]_σ\) as well. In (370) we furnish a few examples, with the syllabication supplied.

(370) **Hypercharacterized syllables in West Finnmark Saami**

<table>
<thead>
<tr>
<th>Saami</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>kill.lij</td>
<td>‘language-ILL.SG’</td>
</tr>
<tr>
<td>park.kaan</td>
<td>‘work-1SG.PRES’</td>
</tr>
<tr>
<td>knoqr.mj</td>
<td>‘become thin; give up -3SG.PRET’</td>
</tr>
<tr>
<td>faarf.fuuh</td>
<td>‘loop, noose - NOM.PL’</td>
</tr>
<tr>
<td>koans.ta</td>
<td>‘art-ACC/GEN.SG’</td>
</tr>
<tr>
<td>faajh.po.ðih</td>
<td>‘keep waving one’s hands -FREQ.-CONT’</td>
</tr>
<tr>
<td>spaajl.liih</td>
<td>‘untamed, castrated reindeer’</td>
</tr>
<tr>
<td>maajjli.liih</td>
<td>‘juice, sap - NOM.PL’</td>
</tr>
<tr>
<td>skoavhl.liih</td>
<td>‘blister - NOM.PL’</td>
</tr>
<tr>
<td>tjaajhn.nihih</td>
<td>‘woodpecker; freckle - NOM.PL’</td>
</tr>
</tbody>
</table>
Maximal Binarity has immediate consequences for how we deal with structures of this type. Hypercharacterization has standardly been treated in terms of trimoraic syllable structure (Hayes 1989; Sherer 1994), but Maximal Binarity forces a different interpretation: hypercharacterized syllables can only be bimoraic, and the difference in markedness between a normal heavy [VC]$_n$ and a hypercharacterized [VVC]$_n$ or [VCC]$_n$ syllable has to be differently construed. With only two moras to go around, it seems reasonable to assign the head mora, or strong mora ($\mu_s$), to the first component of the nucleus, and force sharing of the weak mora ($\mu_w$) between the remaining segments as proposed by Broselow et al. (1995, 1997), Maddieson (1993), Maddieson and Ladefoged (1993), and Sprouse (1996).

Consider for a moment the alternative hypothesis, that there is no such bimoraic limit on the size of the syllable. If moras dominate segments uniquely, then the data in (370) can only be accounted for if we assume syllables with up to four, even five, moras. Even if we went as far as allowing trimoraic syllables, and using the trimoraic syllable as a representation of hypercharacterized kiilij or pärkään, this immediately invites the problem of how to represent overlength. If kiilij is trimoraic, then what is kolija? The two are patently not a natural class.

Hypercharacterization and hyperchrony may be combined (within some limits: overlength has a shortening effect on the vowel centre, eliminating some of the expansions with a long vowel). (371) illustrates the combination of the two.

(371) Hypercharacterization with overlength

\[
\begin{align*}
\text{kaanske} & \quad \text{‘maybe’} \\
\text{tielh:kuuh} & \quad \text{‘spot-nom.pl’} \\
\text{lieöp?:miüh} & \quad \text{‘antler with broad end -nom.pl’} \\
\text{purs:kaan} & \quad \text{‘burst out laughing -1sg.pres’} \\
\text{kürp?:maah} & \quad \text{‘larna which parasitizes reindeer -nom.pl’}
\end{align*}
\]

Hypercharacterized syllables are bimoraic. This is consistent with recent work on hypercharacterized syllables in languages such as Levantine Arabic and Malayalam (Broselow et al. 1997) and Sukuma (Maddieson 1993; Maddieson and Ladefoged 1993).

(372) and (373) illustrate what I take to be the wrong analysis of hypercharacterization. (372) shows the first syllable of kiilij ‘language-ill.sg’ parsed as trimoraic. The long vowel ‘hogs’ the first two moras of the syllable, and the coda is parsed into a third mora, making a trimoraic syllable. (373) shows the first syllable of faarf.fuuh ‘loop, noose -nom.pl’ as quadrimoraic. Again, a long vowel claims the first two moras of the syllable for itself, and each of the consonants in the coda uniquely project a mora, bring the mora count to a total of four.

(372) Hypercharacterized syllable as trimoraic: ©

\[
\begin{array}{c}
\sigma \\
\mu - \mu - \mu - \mu
\end{array}
\]

\[
\begin{array}{c}
k - i \\
\mu - \mu - \mu - j
\end{array}
\]

(373) Hypercharacterized syllable as quadrimoraic: ©
Contrast (372) and (373) with (374) and (375), where the initial syllables of both forms are parsed as bimoraic. The difference in complexity between the two turns not on the number of moras in the first syllable, but on the number of segments dominated by the weak mora ($\mu_w$) of the first syllable. In (374), the coda does not project its own mora, but the weak mora of the first syllable dominates both the vowel nucleus and the simplex coda. In (375), the weak mora of the first syllable dominates both the vowel nucleus and the complex coda.

(374)  
Hypercharacterized syllable as bimoraic: ☉

(375)  
Hypercharacterized syllable as bimoraic: ☉

In conclusion, hypercharacterized syllables are prosodically bimoraic. Their relative structural complexity must be characterized in segmental, not moraic, terms.

8.2.2 Hyperchrony

How is hyperchrony’s excess moraicity accommodated within higher prosodic structure? There are two representational options.

One possible analysis sees overlength as a property of the syllable, and takes the extra length to be parsed into a trimoraic syllable. Maximal Binarity, however, forces a different interpretation: there is no such thing as an ‘overlong’ or trimoraic syllable. Rather, the extra mora which characterizes overlength is properly seen as parsed into a minor syllable of its own. The correct representation for Q3 is as a formal disyllable, in which the second syllable lacks a nucleus as in (376).\(^2\)

(376)  
Q3 as a syllable dyad (Minor Syllable Hypothesis)

\(^2\)A nucleus is understood here as a full vowel, i.e. a vocalic root node specified for features. Consonants and schwa are not nuclei in this sense and are unable properly to head the syllable.
We will ground the reason for this choice in a moment. As an example of
the contrast between a Q2 and a Q3 geminate, consider the representations of
pallaah ‘to fear’ and pal:laah ‘to get scared (and run off)’ in (377) and (378)
respectively.

(377) Representation of pallaah ‘to fear’

```
  \sigma \mu \mu \mu p \ a \ l \ a \ h
```

(378) Representation of pal:laah ‘to get scared’

```
  \sigma \mu \mu \mu \mu p \ a \ s \ a \ a \ h
```

(379) and (379) illustrate the same Q3/Q2 opposition between clusters.

(379) Representation of paste ‘spoon/ACC/GEN.SG’

```
  \sigma \mu \mu \mu p \ a \ s \ t \ e
```

(380) Representation of pas:te ‘spoon/NOM.SG’

```
  \sigma \mu \mu \mu \mu p \ a \ s \ t \ e
```

(381) illustrates the representation of kaas:si ‘earwax; wax; colostrum’,
showing how hypercharacterization and overlength may coexist within the rep-
resentation.

(381) Representation of kaas:si

```
  \sigma \mu \mu \mu \mu k \ a \ s \ s \ i
```

Coda Maximization may coexist with Balance (or ‘secondary lengthening’).
When this occurs, we get a hypercharacterized syllable (with a complex coda),
followed by a minor syllable instantiating Q3. Consider the structure of ku:jk:ku:uh
‘heifer-NOM.PL’.

(382) Representation of ku:jk:ku:uh ‘heifer-NOM.PL’

```
  \sigma \mu \mu \mu \mu k \ u \ j k \ u \ h
```
Evidence for (376) being the correct representation of overlength comes from four domains:

- Metrical behaviour of 'superheavies' in other languages
- Coda Maximization
- Epenthesis
- Balance

In Estonian (Hint 1973, 1997; Hayes 1995; Bye 1997a, b), the analysis of overlength as a disyllable is motivated by its metrical behaviour. Abstracting away from the many intricacies of Estonian stress assignment (see Kager 1994, Hayes 1995 and Bye 1997b for three recent analyses), Estonian parses bisyllabic quantity-insensitive stress feet from left to right. Crucially, though, secondary stress has the option of resuming immediately following an overlong domain. Thus, a word such as *jõukête 'courageous-gen.pl', with initial overlength, is metrical equivalent to quadrисyllabic *sõojemätte 'warmer-adess.pl'. Similar metrical behaviour is attested in a variety of other languages with sesquisyllabic structures in their inventory. Examples include Eastern Hindi (Pandey 1989; Hayes 1995), and Arabic (McCarthy and Prince 1990).

(376) further predicts the observed failure of Coda Maximization to apply following overlength. Assuming that Coda Maximization applies to *enhance the head syllable of the foot, Coda Maximization will be universally blocked, since its application would generate a configuration involving crossing autosegmental association lines. This is shown in (383).

\[
\begin{array}{c}
\sigma \\
\mu \\
a \\
\end{array} \quad \begin{array}{c}
\sigma \\
\mu \\
\mu \\
\end{array} \quad \begin{array}{c}
\sigma \\
\mu \\
\mu \\
\end{array}
\]

(383) has no sensible phonetic interpretation. Logically, though, Coda Maximization could of course apply to the minor syllable to give *aalittuu. However, given our assumptions about the CodaMax constraint, applying the process here would be gratuitous: it does nothing to enhance the head syllable of the foot. This is shown in (384).

\[
\begin{array}{c}
\sigma \\
\mu \\
a \\
\end{array} \quad \begin{array}{c}
\sigma \\
\mu \\
\mu \\
\end{array} \quad \begin{array}{c}
\sigma \\
\mu \\
\mu \\
\end{array}
\]

Thus, a combination of representational and functional considerations converge on the prediction that there is no constraint ranking under which (384) will be optimal, given the constraints.

(376) also suggests a functional motivation for Balance. Balance applies before a syllable containing a long vowel. Bye (1997b) is an attempt to interpret
the responsible constraint as a ban on consecutive heavy syllables, *σσ drawing on ideas by Zoll (1992) and Alderete (1997). The process may then be seen as a repair for the offending configuration: by inserting a minor syllable between the two consecutive heavy syllables, the violation is removed.\(^3\)

(376) renders the relationship between Q3 and epenthesis (§7.1.2) more transparent. On the assumption of (376), epenthesis consists merely in the delinking of the consonant from the minor syllable's mora, and the insertion of a vocalic root node in its place. Syllable structure remains constant throughout.

In sum, we have argued that overlength is formally a sesquisyllabic structure consisting of two syllables. The second of these is a degenerate or minor syllable which lacks a proper head, i.e. a vocalic nucleus.

How is this minor syllable incorporated into higher level prosodic structure? I will assume that the minor syllable is adjoined to the foot node, along with the syllables which flank it on either side to give a formally trisyllabic foot as in (385).

(385) \[\text{Representation of kaassii, showing foot structure} \]

\[\text{Ft}\]

\[\sigma\]

\[\sigma\]

\[\mu\]

\[\mu\]

\[\mu\]

\[\text{k}\]

\[\text{a}\]

\[\text{s}\]

\[\text{i}\]

Where does the assumption of the representation of Q3 as disyllabic leave Foot Binarity?

I will assume, following Hewitt (1994), that Foot Binarity (Prince and Smolensky 1993; McCarthy and Prince 1993a) must be resolved into separate constraints which assess binarity on the moraic, nucleic, and syllabic levels of prosodic structure.\(^4\) The appropriate level at which to characterize the binarity of the foot in West Finnmark Saami is thus the nucleus. In lexical words, the foot consists maximally of two nucleic syllables.\(^5\)

(386) \[\text{FtBin-ν} \]

The foot must be binary at the nuclear level.

The corresponding constraint which evaluates the syllable level, FtBin-σ, must be violated in feet with Q3.

(387) \[\text{FtBin-σ} \]

The foot must be binary at the syllabic level.

\(^3\)The idea needs an additional assumption in order to fly. The process is only conditioned by a long-vowelled syllable, i.e. short-vowelled syllables which are closed by a coda do not trigger the rule. Codas would have to be non-moraic foot-finally. Bye (1997b: 88) assumes that foot-finally, codas are parsed directly into the syllable node, by-passing moraic structure, under the exigencies of a constraint Weak Edge, for which see Spaelti (1994).

\(^4\)Hewitt argues in addition that we must also assess violations of minimality and maximality differently. Ft Bin resolves itself into two constraints, FtMin, 'A foot must not be less than two units (moras, nuclei, syllables)', and FtMax, 'A foot must not be more than two units'. See also Green (1995), Everett (1995), and Green and Kenstowicz (1995) for similar views. I will abstract away from the minimality/maximality aspect of binarity violations here.

\(^5\)For the status of the nucleus as an authentic unit of prosody, see also Shaw (1993).
Minor syllables are thus doubly marked: they lack a proper head and they entail a trisyllabic foot. Minor syllables are permitted only at the cost of violating both $\text{FtBin}\neg\sigma$ and $\sigma\text{-Nuc}$ in (388).

(388) \[\sigma\text{-Nuc}\]

Syllables must be properly headed (contain a nucleus).

The inventories of the language are such that a violation of $\sigma\text{-Nuc}$ always implies a violation of $\text{FtBin}\neg\sigma$. Q3 only occurs foot-medially, and so the foot-medial position is the only place where syllables lacking a proper head may surface. Because feet in West Finnmark Saami are in addition obligatorily binucleic, this forces a violation of $\text{FtBin}\neg\sigma$.

8.3 Interaction of consonant quantity and vowel length

Pre-Q3 Vowel Shortening (§7.1.11) is attributed to one of a family of constraints which enhance the duration contrast between a vowel and the following consonant. In this case, shortening takes place provided that the target vowel or diphthong is $[\neg\text{low}]$.

(389) \*$\tilde{V}[\neg\text{low}]C:C$

A long $[\neg\text{low}]$ vowel followed by a Q3 consonant or consonant cluster is disallowed.

(389) is a specific instantiation of the more general constraint in (390).

(390) \*$\tilde{VC}:C$

A long vowel followed by a Q3 consonant or consonant cluster is disallowed.

Repair takes place by shortening of the vowel rather than the consonant, showing that $*\tilde{V}[\neg\text{low}]C:C$ must outrank \texttt{Wt-Ident(V)} and that \texttt{Wt-Ident(C)} must also dominate \texttt{Wt-Ident(V)}.

(391) \texttt{Wt-Ident}

Let $\alpha$ and $\beta$ be segments, $\alpha \in \text{input}$, $\beta \in \text{output}$ and $\alpha \Re \beta$. If $\alpha$ is $n$-moraic, then $\beta$ is $n$-moraic.

Evaluation: Assess one mark for each decrease or increase in $n$ in the output.

Finally, the more general constraint, $*\tilde{VC}:C$, must be ranked low, below both of the \texttt{Wt-Ident} constraints, in order to get out the fact that, following long $\text{aa}$, there is no quantitative adjustment at all.

(392) \textit{Vowel Shortening}
CHAPTER 8. QUANTITY

For Girjjis, the reverse ranking holds, i.e. \( \text{Wt-Ident}(V) \) dominates \( \text{C:C} \).

(393) No quantitative adjustments after long /aa/

For the shortening of input bimoraic rising diphthongs to monomoraic rising diphthongs, the ranking \( \text{C:C} \gg \text{Wt-Ident}(V) \) does the job. However, there is more to be said here because Richness of the Base has to be taken into account. We have to assume that monomoraic rising diphthongs can be present in the input, and in this case we want to be able to prevent them from surfacing in positions where they are disallowed, i.e. anywhere except preceding a Q3 consonant. To prevent this occurring we have to invoke the markedness constraint \( \text{*Short-Diphthong} \) in (394).

(394) \( \text{*Short-Diphthong} \)

Monomoraic diphthongs are disallowed.

Ranking \( \text{*Short-Diphthong} \) below \( \text{C:C} \) and above \( \text{Wt-Ident}(V) \) gets the right results for input monomoraic diphthongs. (395) and (396) show that monomoraic rising diphthongs are optimized before a Q3 consonant on either assumption about the input. (397) and (398) show that bimoraic rising diphthongs are optimized elsewhere, again under either assumption about the input.

(395) \( /V_1V_2/ \rightarrow \tilde{V}_1V_2 \)
8.3. CONSONANT QUANTITY AND VOWEL LENGTH

<table>
<thead>
<tr>
<th>/kuõssii/</th>
<th>*( V_{1-\text{low}} )</th>
<th>C-CHIN-W-IDENT(V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ꜑ kiõssii</td>
<td>:</td>
<td></td>
</tr>
<tr>
<td>b. knõssii</td>
<td>!_</td>
<td></td>
</tr>
<tr>
<td>c. knõssii</td>
<td>: !_</td>
<td></td>
</tr>
</tbody>
</table>

(396) /\( \tilde{V}_{1} V_{2} \)/ → /\( \tilde{V}_{1} V_{2} \)/

<table>
<thead>
<tr>
<th>/kuõssii/</th>
<th>*( V_{1-\text{low}} )</th>
<th>C-CHIN-W-IDENT(V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ꜑ kiõssii</td>
<td>:</td>
<td></td>
</tr>
<tr>
<td>b. knõssii</td>
<td>!_</td>
<td></td>
</tr>
<tr>
<td>c. knõssii</td>
<td>: !_</td>
<td></td>
</tr>
</tbody>
</table>

(397) /\( V_{1} V_{2} \)/ → /\( V_{1} V_{2} \)/

<table>
<thead>
<tr>
<th>/kuollii/</th>
<th>*( V_{1-\text{low}} )</th>
<th>C-CHIN-W-IDENT(V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. kiollii</td>
<td>: !_</td>
<td></td>
</tr>
<tr>
<td>b. ꜑ kuollii</td>
<td>:</td>
<td></td>
</tr>
</tbody>
</table>

(398) /\( \tilde{V}_{1} V_{2} \)/ → /\( \tilde{V}_{1} V_{2} \)/
8.4 Conclusions

In addition to overlength, West Finnmark Saami also allows hypercharacterized syllables.

Syllables are, however, maximally bimoraic, so trimoraic syllables are, by hypothesis, universally absent from the inventory of syllable types. Adopting this assumption as a hard constraint on syllable structure forces two conclusions concerning the representation of hypercharacterization and overlength.

The complexity of hypercharacterized syllables relative to CVV or CVC heavy syllables is a matter of the number of segmental associations to the weak mora of the syllable.

Hyperchrony involves a trimoraic complex. The third mora of this complex is parsed into a distinct minor syllable, lacking a proper head (nucleus). The degenerate head of a minor syllable contains either the segment parsed as coda in the preceding syllable, else an excrescent vowel.

Given that hypercharacterized syllables seem to be most judiciously analyzed as bimoraic, there would seem to be little reason to allow trimoraic syllables as part of the inventory of any language.

It was also argued that a number of empirical considerations lend further support to the Minor Syllable Hypothesis. Language-internally, Coda Maximization, Balance, and Epenthesis all have greater compatibility with the Minor Syllable Hypothesis than the alternative analysis as a trimoraic syllable.
Chapter 9

Syllabication of Sonorant-Initial Clusters

In the previous chapter, we argued that the essential representational distinction between Q2 and Q3 is that Q3 is augmented by an extra mora. This tertial mora, which defines overlength is, moreover, licensed by association to distinct minor syllable node. We have thus already established one of the central characteristics of syllabication in West Finnmark Saami: minor syllables are permitted.

In this chapter, we will examine the correlation between morphological grade (Strong vs. Weak) and certain syllabication processes. The generalizations to be explicated in this chapter are these:

1. In the Strong Grade, heterorganic sonoran-initial clusters undergo epenthesis.

2. In the Weak Grade, sonoran-initial clusters undergo Coda Maximization.

Obviously, notions such as Strong and Weak Grade are entirely epiphenomenal and language-specific constructs. Here we idealize by reconstruing these morphological notions in terms of primitive quantity, equating the Weak Grade with Q2, and the Strong Grade with Q3.

Although Epenthesis only takes place in heterorganic sonoran-initial clusters, Epenthesis and Coda Maximization are otherwise in complementary distribution: Epenthesis is restricted to the Strong Grade, while Coda Maximization takes place exclusively in the Weak Grade.

The challenge is to discern the default pattern underlying this surface distribution. The account proposed here capitalizes on the idea, developed in §8.2.1, that Coda Maximization is blocked in the Strong Grade due to hard constraints on representations. In the Weak Grade, what we find is not epenthesis, but maximization of the coda. This is taken to be reflective of the default defined by the grammar: because Epenthesis and Coda Maximization are in conflict, both cannot be implemented simultaneously. The pattern evinced by sonoran-initial clusters shows that this conflict is resolved in favour of Coda Maximization. Other things being equal, then, Coda Maximization is what we expect to find. In the Strong Grade, though, other things are not equal: Coda Maximization cannot apply in principle, for the reasons just indicated. This leaves the road
open for Epenthesis to apply in the Strong Grade, essentially in opportunistic fashion, allowing satisfaction of a constraint whose requirements are in the normal course of things subordinated to the demands of whatever constraint drives Coda Maximization.

The structure of this chapter is as follows.

In §9.1, the basic tools for dealing with syllabication in OT are laid on the table, and the place of West Finnmark Saami within a general typology of the syllable is discussed.

§9.2 deals with the syllabication of clusters in the Strong Grade in greater detail, addressing epenthesis first (§9.2.1) before turning to the syllabication of the medial consonant in triliteral clusters (§9.2.2).

§9.3 examines the syllabication of sonorant-initial clusters in the Weak Grade according to the requirements of Coda Maximization.

§9.4 presents the main conclusions.

9.1 North Saami within a general typology of syllable structure

This section aims to place West Finnmark Saami within a general typology of syllable structure. (For an early application of OT to the theory of syllabication, see Prince and Smolensky 1993: 85-97.)

Onsets are optional (there are vowel-initial words), and codas are permitted. In addition, both onsets and codas may be complex. These generalizations may be summarized in the rankings in (406) of the constraints (399) to (405).

(399) **Onset**
Syllables must have an onset.

(400) **NoCoda**
Codas are disallowed.

(401) ***CxOns**
Complex onsets are disallowed.

(402) ***CxCoda**
Complex codas are disallowed.

(403) **Dep-C**
Every consonant in the output must have a correspondent in the input.

(404) **Max-C**
Every consonant in the input must have a correspondent in the output.

(405) **Dep-V**
Every vowel in the output must have a correspondent in the input.

(406) **Basic syllabification in North Saami**

a. **Dep-C ≫ Onset**
Onsets are optional.

b. **Max-C, Dep-V ≫ NoCoda**
Codas are permitted
9.2 Syllabification in the Strong Grade

9.2.1 Epentheses

Recall from §7.1.2 that vowel epenthesis applies following a sonorant /βðl r j/ in a heterorganic cluster.¹

By hypothesis, the epenthized vowel is featurally unspecified on the surface, its actual quality determined by phonetic implementation.

The markedness constraint driving epenthesis is given in (407).

\[(407) \quad *\text{SonC} \quad \text{A liquid must not be immediately followed by a consonant.}\]

Ranking (407) above Dep-V returns the epenthized candidate as optimal.²

(408) shows this for kuol.la.ka ‘hair’ (Sammallah 1977: 59).

\[
\begin{array}{c|c|c|c}
\text{Word} & /\text{kuol.la.ka}/ & *\text{SonC} & \text{Dep-V} \\
\hline
a. & \text{kuol.la.ka} & \checkmark & \checkmark \\
b. & \text{kuolka} & \checkmark & \times \\
\end{array}
\]

9.2.2 Complex Onsets and the Syllabification of Strong Grade Triliteral Clusters

This section addresses the syllabification of Q3 triliteral clusters in the Strong Grade. As indicated in §§5.4.3, the medial consonant of the triliteral cluster must be either s or h. These medial consonants are always syllabified into a following onset as shown in (409). This is exactly the syllabification proposed by Sammallahti (1998b: 52).

\[
(409) \quad \text{Syllabification in triconsonantal clusters}
\]

\begin{align*}
a. \quad \text{paal.la.hkaa} & \quad \text{‘salary, NOM.SG’} \\
& \quad \text{tear.ro.hpah} \quad \text{‘forge’} \\
b. \quad \text{kob.b.pmii} & \quad \text{‘ghost’} \\
& \quad \text{pon.n.n?nii} \quad \text{‘bottom’} \\
c. \quad \text{roam.m.ske} & \quad \text{‘a real mess’} \\
& \quad \text{koam.n.sta} \quad \text{‘means’}
\end{align*}

¹ A similar phenomenon is found in Finnish (Harrikari 1999a). In Finnish, epenthesis takes place following a sonorant, but is blocked following an obstruent.
² Harrikari (1999a) argues for a different analysis for very similar phenomena in Finnish. She sees epenthesis as driven by NoCod. In both Finnish and Saami, epenthesis fails to occur following an obstruent. In the approach advocated here, this failure is simply a consequence of the fact that the epenthesis-driving constraint does not mark obstruent-initial sequences. For Harrikari, though, this failure becomes a matter of blocking.
Medial consonants cannot be syllabified into a minor syllable, i.e. any syllable lacking a featurally-specified vocalic root-node. One possibility is that this represents the activity of a local conjunction of markedness constraints: $σ$-Nuc and NoCoda within the domain of the syllable. Both of these are violated simultaneously in forms such as ungrammatical *[paal.h.kaa] and *[roam.ms.ke]. These constraints are satisfied by syllabifying /s h/ in onset, in violation of *CxOns. The conjoined constraint is stated in (410) below.

\[(410) \ [σ$-$Nuc$\&$ NoCoda] \]

Codas are not permitted in minor syllables.

Tableaux are shown for tearrahpah and koan:sta in (411) and (412) respectively.

\[(411) \ /tearr:hpah/ \ [σ$-$Nuc$\&$ NoCoda] \ *CxOns \]
\[\begin{array}{ccc}
a. & \checkmark & tear.rohpah & *! \\
b. & & tear.roh.pah & *!
\end{array}\]

\[(412) \ /koan:sta/ \ [σ$-$Nuc$\&$ NoCoda] \ *CxOns \]
\[\begin{array}{ccc}
a. & \checkmark & koan.n.sta & *! \\
b. & & koan.ns.ta & *!
\end{array}\]

Nevertheless, there is a more plausible explanation for this Onset Maximization pattern which at once obviates the need to postulate constraint conjunction and is more consonant with the reasoning in this chapter. We return to the issue in §9.3.3.

9.3 Coda Maximization in the Weak Grade

9.3.1 Coda maximization in biliteral clusters

Coda Maximization involves the enhancement of a particular prosodically prominent position, such as the root-initial syllable or head syllable of the foot, through the maximization of syllable content. This goes against the Jakobsonian principle that onsets are maximized in preference to codas. This violation of basic Jakobsonian principles of syllabification is restricted, however, to phonologically salient positions, in Saami, the stressed syllable of the foot. A bit of prospecting in the typological data reveals that Coda Maximization is not in fact particularly marked cross-linguistically (Beckman 1998): the constraint responsible for Coda Maximization in Saami is one and the same with that which drives ambisyllabicity in languages such as English. It simply happens that, in Saami, ambisyllabification is attended by gemination, unlike in English.

Coda Maximization in Saami is dominated by high-ranking constraints on syllabic well-formedness. Coda Maximization is restricted to those consonant clusters whose initial member is a sonorant. But there is an additional restriction. Recall that Saami allows voiceless sonorants. If the sonorant is [−voice], though, Coda Maximization is checked.
Examples of the application of Coda Maximization in the Weak Grade are shown in (413). (414) and (415) illustrate the blocking effects of the specifications \([−son]\) and \([−voice]\).

(413)  *Coda maximization applies: \(C_1 = [+son, +voice]\)*

- \([taaw:ta]\)  `illness, NOM.SG`
- \([taawt.ta]\)  `illness, ACC/GEN.SG`
- \([hil.l@.kuuh]\)  `reject, INF`
- \([hil.ko]\)  `reject, IMP`

(414)  *Coda maximization blocked: \(C_1 = [−son, −voice]\)*

- \([SuS:mii]\)  `heel, NOM.SG`
- \([SuSm.mii]\)  `heel, ACC/GEN.SG`
- \([faaT:mii]\)  `lap, NOM.SG`
- \([faaT.mii]\)  `lap, ACC/GEN.SG`

(415)  *Coda maximization blocked \(C_1 = [−voice]\)*

- \([pas:te]\)  `spoon, NOM.SG`
- \([past.te]\)  `spoon, ACC/GEN.SG`

Ambisyllabication results from the interaction of Onset and a constraint which requires maximization of the initial or head syllable of the foot. This constraint is designated Codamax, but it is clearly the same as the Max-σ proposed by Beckman (1998: 217). Beckman discusses a number of cases in which maximization is conditioned in initial or head position. Examples of the former type include ambisyllabication in Elko roots, and of the latter type, stressed syllable maximization in Scots Gaelic. Although a perceptually-grounded markedness constraint, Max-σ is formally an instance of the Max family of constraints. Beckman’s formulation of the constraint is given in (416).

(416)  \(\text{Max-σ (}= \text{CODAMAX})\)

\(\forall x, x \in S_1, y \text{ such that } y \in S_2, x \Re y \text{ and } y \text{ appears in the head syllable of the foot,}
\)

`Every element of the input has a correspondent in the head syllable of the foot in the output.'

(416) favours maximal syllabification of all head syllables in feet and penalizes candidates by assessing a mark against every segment which is not a member of the head syllable of the foot. Codamax must dominate \(*\text{CxCODA},\)

\(^3\text{In fact, the variant of this constraint discussed by Beckman is relativized to the initial syllable of the root, rather than the head syllable of the foot. Nevertheless, in all other respects the two constraints are identical.}\)
since other things being equal, the grammar of West Finnmark Saami accords
greater relative harmony to \(C_1C_2C_2\) than \(C_1C_2\). However, \textsc{CodaMax} itself
must be dominated by another constraint given the failure of Coda Maxi-

mization in (414) and (415) above. For (414) at least, the responsible constraint is
the well known \textit{Sonority Sequencing Condition} (SSC), given in (417).\footnote{This formulation ignores the grammaticality of \textit{s+stop} and \textit{h+stop} onset clusters. Either some onset-specific faithfulness constraint outranks SSC, or this formulation should be replaced with one specifically relativized to the coda.}

\begin{align}
\text{(417) } & \textit{Sonority Sequencing Condition (SSC)} \text{ (e.g. Clements 1990)} \\
& \text{Sonority must increase monotonically towards the syllable peak and de-
crease monotonically towards the coda.}
\end{align}

For the data in (413) above, Coda Maximization takes place without viola-
tion of the SSC. This is shown in (418) for \textit{aajpaa} 'absolutely, quite'.

\begin{align}
\text{(418) } & \textsc{CodaMax} \gg \text{\*CxCoda} \\
& \text{\begin{tabular}{|l|l|l|}
\hline
 & \text{SSC} & \text{CodaMax} & \text{\*CxCoda} \\
\hline
a & \& & a & \text{a} \\
b & aajpaa & p!, a & \text{p!, a} \\
\hline
\end{tabular}}
\end{align}

Still, (418) neglects to account for the ungrammaticality of a third possible
candidate, \textit{\text{*aajp.aa}}, in which Coda Maximization takes place at the expense of
violating \textit{Onset}. In order to ensure optimization of the real winner, \textit{aajp.paa},
over ungrammatical \textit{\text{*aajp.aa}}, we have to rank \textit{Onset} over \textit{Unique-}\sigma. This
constraint is defined in (419). For the \textit{Unique} family of constraints, see Ben

\begin{align}
\text{(419) } & \textit{Unique-}\sigma \\
& \forall x, x \text { a segment, } x \text { must have a unique syllable anchor } y.
\end{align}

The tableau in (420) incorporates this new ranking information and presents
the full picture.

\begin{align}
\text{(420) } & \textit{Onset} \gg \textit{Unique-}\sigma \\
& \text{\begin{tabular}{|l|l|l|l|}
\hline
 & \text{SSC} & \text{Onset} & \text{CodaMax} & \text{Unique-}\sigma \\
\hline
a & \& & a & \text{a} \\
b & aajp.paa & p!, a & \text{p!, a} \\
c & aajp.aa & \text{*!} & \text{*!} \\
\hline
\end{tabular}}
\end{align}

Next, let us consider the data in (414). Coda Maximization is blocked in
these cases, since application of the process would lead to a fatal violation of
the undominated SSC. This is shown in (421).

\begin{align}
\text{(421) } & \text{SSC} \gg \textsc{CodaMax} \\
& \text{\begin{tabular}{|l|l|l|l|}
\hline
 & \text{SSC} & \text{Onset} & \text{CodaMax} & \text{Unique-}\sigma \\
\hline
a & \& & \text{*!} & \text{*!} \\
b & \& & \text{leasm.mii} & \text{leasm.mii} \\
c & \& & \text{leasm.mii} & \text{leasm.mii} \\
\hline
\end{tabular}}
\end{align}
9.3. CODA MAXIMIZATION IN THE WEAK GRADE

The coda-maximized candidate (421a) scores a fatal mark on the SSC, leaving (421b) as the winner. As shown in (415), Coda Maximization of $C_1C_2$ to $C_1C_2C_2$ is also disallowed if $C_1$ is voiceless, e.g. /sθ hI/. I identify the relevant constraint as (422). The constraint is a specific manifestation of the preference for codas to be relatively high sonority (Clements 1990), $\text{CxCodaCond}$ in (422).

\[(422) \quad \text{CxCodaCond} \]
Do not have a tautosyllabic coda sequence of two consecutive $[\text{−voice}]$ consonants.

Ranking $\text{CxCodaCond}$ above $\text{CodaMax}$, as in the tableau in (423) gets the correct result.

\[(423) \quad \text{CxCodaCond} \gg \text{CodaMax} \]

9.3.2 Coda maximization in triliteral clusters

So far we have examined biliteral clusters with respect to Coda Maximization, and we have seen that Strong Grade clusters in Q3 of the form $C_1C_2$ alternate with $C_1C_2C_2$ clusters in the Weak Grade when considerations of sonorancy and syllable structure permit. Coda Maximization in clusters of three segments manifests itself not as gemination, but as a shift in the location of the syllable boundary. Here, Q3 clusters of the form $C_1C_2C_3$ alternate with $C_1C_2C_3$ in the Weak Grade, i.e. $C_2$ is syllabified as onset in the Strong Grade, but its Weak Grade correspondent surfaces as a coda, e.g.

\[(424) \]

\[
\begin{array}{llll}
\text{pěal.la.lkiih} & \text{‘to scold’} & \text{peelh.kii} & \text{‘scold, 3SG.PRET’} \\
\text{kiör.rap?.mi?} & \text{‘load’} & \text{kuurp?.miín} & \text{‘load, LOC.PL’} \\
\text{pööl.la.hpií} & \text{‘s.th. difficult} & \text{pöohl.pií} & \text{id., ACC/GEN.PL} \\
\text{hór.roスタ} & \text{‘sacking’} & \text{hors.taah} & \text{‘sacking NOM.PL} \\
\text{luj:ste} & \text{‘skate’} & \text{lujs.teh} & \text{‘skate NOM.PL} \\
\text{nålита} & \text{‘stubb; dwarf’} & \text{nálitaas} & \text{‘stunted’} \\
\end{array}
\]

Significantly, all the machinery needed to account for the behaviour of triliteral clusters under Coda Maximization is already in place, and no additional assumptions have to be deployed.
The obvious question is: why is there no gemination in triliteral clusters? The answer falls out from the ranking we’ve already established. Gemination fails because it always leads either to a fatal violation of undominated \( \text{CxCo-daCond} \), or a fatal violation of low-ranked \( \text{UNIQUE-\sigma} \). The optimization of `peelh.kii` ‘he scolded’ in (425) illustrates the point.

(425) Coda maximization in triliteral cluster

<table>
<thead>
<tr>
<th>Candidate</th>
<th>SSC</th>
<th>CxCo-da</th>
<th>CodaMax</th>
<th>UNIQUE-\sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. <code>peelh.kii</code></td>
<td>:</td>
<td>k, i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. <code>peelh.kii</code></td>
<td>:</td>
<td>k, i</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c. <code>peelh.kii</code></td>
<td>:</td>
<td>h!, k, i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. <code>peelh.kii</code></td>
<td>: *!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. <code>peelh.kii</code></td>
<td>: *!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Of the two candidates with gemination, (425b) fails because it has a gratuitous violation of \( \text{UNIQUE-\sigma} \), thereby losing the contest to optimal (425a). The other relevant rival, (425d) violates highly ranked \( \text{CxCo-daCond} \), eliminating it from the contest. (425c) loses to (425a) because it incurs an additional and fatal violation of CodaMax.

9.3.3 Blocking of epenthesis in Q2 weak clusters

We are now in a position to address the fact that epenthesis only occurs in the Q3 Strong Grade. Intuitively, what is going on is this: in Q3, CodaMax must of necessity be violated. This follows from our representational assumptions — hypermorality can only be accommodated within an extra minor syllable. In Q2, however, where neither the universal ban on crossing association lines, nor gratuitous association are at stake, the effect of CodaMax emerges, assuming CodaMax dominates \( ^*\text{SonC} \). If it is of greater importance to maximize the coda of the head syllable of the foot than to break up a sonorant+C cluster by epenthizing, then Coda Maximization will result. Formally, though, the analysis doesn’t quite fly, as shown in (426) for `ar:v a` ‘will power’. The frownie (\( \odot \)) marks the actual, but undesired winner.

(426) Epenthesis in Q3: desired winner beaten \( \odot \)

<table>
<thead>
<tr>
<th>Candidate</th>
<th>CodaMax</th>
<th>( ^*\text{SonC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( \odot ) <code>ar:v a</code></td>
<td>v, a</td>
<td></td>
</tr>
<tr>
<td>b. <code>ar:v a</code></td>
<td>p!a, v, a</td>
<td>( \odot )</td>
</tr>
</tbody>
</table>

The problem is that epenthesis of a vowel returns a candidate which fares worse on CodaMax, eliminating it at the first hurdle. The desired winner,
9.3. CODA MAXIMIZATION IN THE WEAK GRADE

(426b), incorporates less material into its coda than the winning candidate (426a). The reason for this, though, is trivial: (426b) implements epenthesis, and so there are more segments to offend CODAMAX. The problem would seem to lie with the formulation of the constraint CODAMAX. What CODAMAX does is require that every segment in the domain be syllabified in \( \sigma_1 \). However, this results in penalizing segments which can in principle never be syllabified in this way: it is universally the case that vowels uniquely project a syllable node.

It is not unreasonable to suggest that vowels are therefore exempt from evaluation by CODAMAX, since it is the consonants which matter. Reconstructed in this way, the offending marks are eliminated. I therefore propose a redefinition of CODAMAX as in (427). Substituting 'consonant' in for 'element' in the definition in (416) gets the desired result.

\[
\text{(427) } \text{Max-} \sigma (= \text{CODAMAX}, \text{revised from (416)})
\]

\[
\forall x, x \in S_1, y \text{ such that } y \in S_2, x \not\exists y \text{ and } y \text{ appears in the head syllable of the foot iff } y \text{ is a consonant.}
\]

'Every element of the input has a correspondent in the head syllable of the foot in the output,'

This reformulation changes the violation profile of the candidates. This is shown in (428).

\[
\text{(428) } \text{Epenthesis in Q3: correct winner optimized } \odot
\]

<table>
<thead>
<tr>
<th></th>
<th>CODAMAX</th>
<th>*SonC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ar:va</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>ar:.va</td>
<td>*</td>
</tr>
</tbody>
</table>

(428) gives the correct result. What is important for CODAMAX is whether the onset of the dependent syllable of the foot is parsed as a coda in the head syllable or not. As (428) shows, all of the candidates violate of CODAMAX equally, since /v/ is not parsed into the head syllable in any of the candidates. Coda Maximization in Q3 cannot possibly succeed, for the reasons already explored in §8.2. With CODAMAX thus muzzled, the task of arbitrating the relative harmony of the relevant candidates is passed down to the lower ranked constraint *SonC. Both candidates (428a) and (428b) violate *SonC, so (428b), the epenthesis candidate emerges as the winner.

In Q2, however, the prosodic conditions are crucially different. In contrast to the Strong Grade, there is actually a hope of satisfying CODAMAX. Ranking CODAMAX above *SonC derives the observed pattern.

\[
\text{(429) } \text{No epenthesis in Q2}
\]

<table>
<thead>
<tr>
<th></th>
<th>CODAMAX</th>
<th>*SonC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ar.va</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>ar:va</td>
<td>*!</td>
</tr>
<tr>
<td>c.</td>
<td>ar:va</td>
<td><em>!</em></td>
</tr>
</tbody>
</table>

Candidate (429c) emerges victorious, since (429a) and (429b) both violate higher-ranked CODAMAX.
We can use the same logic of opportunism to explain the pattern of Onset Maximization in the Strong Grade of triliteral clusters. Onset Maximization diagnoses the ranking $\text{NoCoda} \gg \text{*CxCOns}$, according to which a consonant cluster /CC/ is preferentially parsed as .CC rather than C.C. NoCoda stands in direct conflict with Codamax, since Codamax wants there to be as much material as possible in the coda of the stressed syllable. However, in the Strong Grade, Codamax cannot be satisfied without violating undominated constraints (eliminating overlength is not an option). Syllabification therefore defaults to an onset-maximizing mode in the Strong Grade.

(430) **Onset Maximization in Strong Grade triliteral cluster**

<table>
<thead>
<tr>
<th></th>
<th>Codamax</th>
<th>NoCoda</th>
<th>*CxCOns</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tear.rah.pah</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. tear.rah.pah</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, Coda Maximization can only apply following a consonant, never following a vowel. There is nothing in what we have said up till now which would prevent it applying to intervocalic consonants in Q1. With Codamax in the system, therefore, we would expect input Q1 to surface as Q2 in all circumstances! This does not happen, and so there must be a constraint which blocks the mapping. This constraint cannot, of course, be Wt-Ident. Wt-Ident is violated on applying Coda Maximization, and so Codamax must dominate Wt-Ident. I propose that the relevant constraint is (431).

(431) **Heavy Consonant Centre (*HCC)**

Moraic consonant centres are disallowed.

(431) militates against any foot-medial consonantism with moraic associations, penalizing geminates and clusters alike. With *HCC dominating Codamax, we predict one situation where Coda Maximization can apply without making matters any worse. This is when the consonant centre already comes bearing moraic associations because the consonant centre is a cluster. Coda Maximization itself thus applies in opportunistic fashion, when foot-medial consonant monorality is not at stake. Consider the difference in behaviour between rumaa $\text{`body'}$ and rimssaas $\text{`ragged'}$.

(432)

<table>
<thead>
<tr>
<th></th>
<th>*HCC</th>
<th>Codamax</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. runaa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. runmaa</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For rimssaas, the consonant centre is already moraic. Coda Maximization cannot make things any worse, and so this is what surfaces. (Obviously, other possible repairs, such as deletion of one of the components of the cluster must be ruled out by other constraints such as $\text{\textbar-}\text{-Max}$.)

(433)

<table>
<thead>
<tr>
<th></th>
<th>*HCC</th>
<th>Codamax</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. rimmaas</td>
<td>*</td>
<td>s, $|$</td>
</tr>
<tr>
<td>b. rimssaas</td>
<td>*</td>
<td>s</td>
</tr>
</tbody>
</table>
9.4 CONCLUSIONS

*HCC can’t be permitted to harry unchecked, however. Underlying geminates do not reduce as a consequence of *HCC. But we’ve already established that Wt-Ident is dominated by CodaMax, and, since *HCC in turn dominates CodaMax, it cannot be the case that Wt-Ident dominates *HCC: this would constitute a flagrant violation of the transitivity of constraint ranking. I will assume that there is a more highly ranked constraint, GemIdent, which dominates *HCC.

(434) GemIdent

If geminate in the input, then geminate in the output.

9.4 Conclusions

This chapter has focused chiefly on two processes affecting sonorant-initial clusters, Epenthesis and Coda Maximization. These processes are driven, respectively, by a constraint (*SonC) militating against a sequence of a sonorant followed by a heterorganic consonant, and a constraint (CodaMax) requiring maximization of the coda of the stressed syllable of the foot. Where the two constraints conflict, CodaMax takes precedence to *SonC, explaining the fact that, in the Weak Grade, Coda Maximization is implemented at the expense of violating *SonC. In the Strong Grade, the conflict between the two constraints effectively ceases to exist: Coda Maximization cannot in principle be implemented in the Strong Grade, leaving the way open for satisfaction of *SonC.

Similar reasoning was brought to bear in explaining the distribution of word-internal complex onsets. Complex onsets only occur following a minor syllable, again, the very position Coda Maximization cannot apply. The forced quiescence of CodaMax allows the activity of lower-ranked constraint interactions to emerge, revealing NoCoda ≫ *CxCoda.

In sum, two phonological processes restricted to the Strong Grade, Epenthesis and Onset Maximization, apply opportunistically, in situations where CodaMax is muzzled due to structural considerations.

The constraint rankings established in this chapter are summarized in (435) below.
(435) Ranking summary

\[
\begin{array}{c}
\text{GEMIDENT} \\
\text{*HCC} & \text{SSC} & \text{CXCODA} \\
\text{CODAMAX} \\
\text{ONSET} & \text{NOCODA} & \text{*SONC} & \text{*CXCODA} \\
\text{UNIQUE-σ} & \text{*CXONS} & \text{DET-V}
\end{array}
\]
Chapter 10

Segmental processes

This chapter returns to address the impact of quantity and morphological grade on segmental realization. The layout is as follows.

§10.1 addresses the phonology of clusters of voiced+voiceless stop. Voiced stops in homorganic clusters of voiced+voiceless stop are deleted in the Weak Grade.

§10.2 deals with preaspiration. Voiceless stops are preaspirated in Q2 and Q3.

§10.3 turns to the prestopping of nasals in Q2 and Q3. Prestopped nasals are also (optionally) glottalized.

§10.4 treats the phonology of supralaryngeal spreading: the medial /h/ of a triliteral sonorant-initial cluster surfaces with the supralaryngeal specification of the preceding sonorant.

Finally, §10.5 deals briefly with the intervocalic voicing and spirantization of /t/ and /c/ in Q1.

10.1 Homorganic voiced stop deletion

The most salient aspect of the phonology of clusters of voiced stop+voiceless stop is the deletion of the voiced stop in the Weak Grade. This is understood here as a consequence of the application of Coda Maximization. Since voiced stops in West Finmark Saami are, by hypothesis, [+sonorant], they fulfill the necessary criteria to undergo Coda Maximization. As a consequence of Coda Maximization, though, a tautosyllabic cluster of a voiced stop+voiceless stop is created. The deletion of the voiced stop thus plausibly represents a repair strategy to avoid this configuration: tautosyllabic clusters cannot differ in laryngeal specification.

(436) \text{Laryngeal Condition (LarCond)}

Tautosyllabic clusters must not differ in laryngeal specification.

Consider the derivation of aakkas "objection-loc.sg" from /aagk a+s/ in (437). Recall from §9.3.3 that Coda Maximization is under normal circumstances restricted to post-consonantal position. The reason was \text{*HCC} in (431) — moraic consonant centres are disallowed. \text{*HCC} is in turn dominated by \text{GemIdent}, which militates against the reduction of geminates to singletons.
However, in this case, the consonant undergoing Coda Maximization is not underlyingly geminate, and so there is no reason to expect it to surface as geminate. Coda Maximization in the Weak Grade of a voiced + voiceless stop cluster must be an instance of phonological opacity, because the segment which facilitates Coda Maximization in the first place is not itself present on the surface. The virtual form must preserve the voiced stop due to highly-ranked $\text{IV-Max-C}$. The overapplication effect is enforced by ranking $\text{VS-Wt-Ident(C)}$ above $\text::*HCC}$.

(437) **Voiced stops delete preceding tautosyllabic voiceless stop**

<table>
<thead>
<tr>
<th>Input</th>
<th>$\text{IV-Max-C}$</th>
<th>$\text{VS-Wt-Ident(C)}$</th>
<th>$\text{CodaxMax-C}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>/aagka+s/</td>
<td>***</td>
<td>**s, s, k!</td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td>[aagkkas]~aagkkas</td>
<td>...</td>
<td>**!</td>
</tr>
<tr>
<td>b.</td>
<td>[aagkkas]~aagkas</td>
<td>...</td>
<td>**s, s, k!</td>
</tr>
<tr>
<td>c.</td>
<td>[aagkkas]~aakkas</td>
<td>...</td>
<td>**s, s</td>
</tr>
<tr>
<td>d.</td>
<td>[aagkkas]~akkas</td>
<td>...</td>
<td>*!</td>
</tr>
<tr>
<td>e.</td>
<td>[aagkas]~aakkas</td>
<td>...</td>
<td>*!</td>
</tr>
<tr>
<td>f.</td>
<td>[aakkas]~aakkas</td>
<td>...</td>
<td>*!</td>
</tr>
</tbody>
</table>

(437a) is eliminated because it fails to delete the voiced stop, bringing in its wake an additional and fatal violation of LarCond. (437b) scores a fatal violation on CodaxMax, which is relatively low-ranked. This is sufficient to ensure that it is bested by the actual winner, (437c). (437d) is the transparent competitor, in which the effects of $\text{*HCC}$ are manifest. (437d) is eliminated by scoring a fatal mark on $\text{VS-Wt-Ident(C)}$.

10.2 Preaspiration

All geminate voiceless stops in the input are mapped to preaspirated stops (alternatively, clusters of h+stop) in the output. This is apparently productive, since preaspirates occur in recent loanwords such as konjakka ‘cognac’ and universitet ‘university’. The stop is variably geminate in the source language (Finnish konjakki or Norwegian konjak; Norwegian universitet).

As we established in §7.1.7, preaspirates are clusters with monosegmental status. This monosegmental status is diagnosed by the fact that preaspirated stops have an alternant in Q1. Keer adopts the representation in (438).

(438) **Aspirated Segment**
10.2. PREASPIRATION

\[
\begin{array}{c|c}
\text{rt}_i & \text{rt}_{i,j} \\
\text{PI} & \text{asp}
\end{array}
\]

I will assume, following Keer, that the preaspiration is coindexed with the following stop, to give \([h_{i,j}, p, h_{i,j} h_{i,j} k]\), and so on. This renders the cluster monosegmental from the perspective of certain constraints. For example, we have to be able to capture the fact that, in terms of Faithfulness, Fortition from plain stop /\text{pt} k/ to preaspirated /\text{hp} \text{ht} \text{hk}/ counts 'the same' as Fortition from sonorant of fricative /\text{fs} \text{fl r} \ldots / to geminate /\text{ff} \text{ss} \text{ss} \text{ll} \text{rr} \ldots /. That is, preaspirated stops have to be counted as geminates, and the mapping has to result in a violation of Wt-IDENT, and not merely \text{Dep-C}.

The alternative is to construe the surface form of preaspiration as an instance of the overapplication of the process: on the surface the stop itself is actually short, not geminate, and so, strictly speaking, there is no violation of Wt-IDENT. This interpretation would involve adding representations to the output array (see §11.4.2).

As argued in §7.1.7, Preaspiration in West Finnmark Saami is a composite of three sub-processes: Aspiration, Fission, and Supralaryngeal Delinking, each of which must be driven by independent constraints. As we established earlier, the effect of Aspiration is limited to geminates; singleton voiceless stops are not targeted by the process. Accordingly, the constraint responsible for Aspiration only penalizes the absence of aspiration in geminates (whether in Q2 or Q3). The failure of Q1 voiceless stops to undergo aspiration is therefore not an instance of blocking. The constraint responsible is formulated as specifically targeting geminates, in (439).

(439) \text{GemAsp}

Geminates voiceless stops must be aspirated.

Ranked above \text{Ident}\text{[asp]}, \text{GemAsp} will result in aspiration of a geminate.

Let's turn to examine the Fission process. It is apparently a point of cross-linguistic variation whether aspiration is implemented as \text{pre-} or \text{post-}aspiration. In Inari Saami, the corresponding geminates are \text{postaspirated} (e.g. \text{kat}^{thu} \text{roof} (Äimä 1918a,b; Itkonen 1946, 1971, 1973, 1986–1991); cf. North Saami \text{kah}^{thu}), diagnosing the activity of other constraints. Postaspiration is never found in North Saami, banned by an undominated constraint \*\text{PostAsp}.

(440) \*\text{PostAsp}

Postaspirated stops are disallowed.

We've established that short voiceless unaspirated stops are possible in North Saami. Yet, foot-medially, their occurrence is circumscribed: they are only permitted as the second member of a consonant cluster, never immediately following a vowel. In this environment, they lose their aspiration, presumably

---

1I assume coindexation as given by the representational theory. Another possibility is that coindexation is generated by the grammar. By assuming that the constraint driving preaspiration dominates \text{Integrity}, rather than \text{Dep}, we would get coindexation by reconstruing preaspiration as splitting rather than epenthesis. However, this only works assuming that all surface preaspirates are underlyingly geminate stops.

2John Kingston (p.c.) has suggested that preaspiration of singleton stops is universally unattested and quite conceivably excluded for principled reasons, although he cites the Iroquoian languages as a possible counterexample to this claim.
due to the activity of an undominated markedness constraint \(*[\ldots \text{VC}_{\text{asp}} \ldots]\)\textsubscript{Fe} which dominates \textit{Ident}[\text{asp}].

Let us return to the final aspect of Preaspiration, the process of Supralaryngeal Delinking. This process is behind the observed surface pattern of degemination. As we noted, degemination is concomitant to preaspiration in West Finnmark Saami, but not in other dialects of North Saami. As evidence for this claim, we adduced the Skånland Saami data in (292), in which the underlying weight of the stop component is actually preserved in the surface form. The constraint responsible for the degemination effect in West Finnmark Saami has already been introduced earlier as \textit{CxCodaCond} in (422), repeated here in (441).

\begin{equation}
\text{CxCodaCond}
\end{equation}

\begin{itemize}
  \item Do not have a tautosyllabic coda sequence of two consecutive \([-\text{voice}]\) consonants.
\end{itemize}

The difference between the two dialects consists in the ranking of \textit{CxCodaCond} relative to \textit{Codamax}. In West Finnmark Saami, the conditions on syllable structure are accorded the greater relative priority, and Supralaryngeal Delinking in the optimal candidate is the result. The tableaux in (442) and (443) for \textit{kohpii} 'depression' and \textit{koh:pii} 'bay' show how this falls out in West Finnmark Saami for Q2 and Q3 respectively.

\begin{table}[h]
\begin{tabular}{|c|c|c|}
\hline
\text{/koppii/} & \text{CxCodaCond} & \text{GemAsp} & \text{Codamax} \\
\hline
a. & \text{☞} & \text{...} & \\
\hline
b. & \text{koh.pii} & \text{*!} & \\
\hline
c. & \text{kop.pii} & \text{...} & \text{*!} \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\begin{tabular}{|c|c|c|}
\hline
\text{/koppii/} & \text{CxCodaCond} & \text{GemAsp} & \text{Codamax} \\
\hline
a. & \text{☞} & \text{...} & \\
\hline
b. & \text{koh:pii} & \text{*!} & \\
\hline
c. & \text{kop:pii} & \text{...} & \text{*!} \\
\hline
\end{tabular}
\end{table}

In Skånland, on the other hand, \textit{Codamax} must outrank \textit{CxCodaCond}, suppressing the application of Supralaryngeal Delinking in the winner.

### 10.2.1 The underapplication of preaspiration

Aspiration is countered by Homorganic Voiced Stop Deletion. Aspiration thus underapplies (along with \textit{preaspiration} and the degemination which follows on its heels).

Let us sketch the outlines of an analysis in the terms of Virtual Phonology.
The underapplication effect can be attained by encoding the voiced stop into the virtual form, i.e. in the virtual representation, the voiced stop fails to delete. In terms of ranking, this must mean that \( \text{V\text{-MAX-C}} \) dominates \( \text{LarCond} \), as we established above. In the surface representation, the voiced stop deletes, diagnosing that \( \text{LarCond} \) must in turn dominate both \( \text{V\text{-MAX-C}} \) and \( \text{IS\text{-MAX-C}} \).

The underapplication effect itself is a result of high-ranking constraints on virtual-surface identity. Specifically, Aspiration underapplies because \( \text{V\text{-IDENT[asp]}} \) dominates \( \text{GemAsp} \), which, as we established in (439), is the constraint driving aspiration.

The tableau in (444) summarizes the analysis, illustrating how the proposed grammar deals with the input /aagk\text{a+s}/ ‘objection-loc.sg’. For reasons of space, though, only those candidates which satisfy the constraints on Consonant Gradation and CodaMax are shown.

(444)

<table>
<thead>
<tr>
<th>/aagk\text{a+s}/</th>
<th>\text{V\text{-IDENT[asp]}}</th>
<th>\text{GemAsp}</th>
<th>\text{V\text{-MAX-C}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>: *!</td>
</tr>
<tr>
<td>b.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>: *</td>
</tr>
<tr>
<td>c.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>: *</td>
</tr>
<tr>
<td>d.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>: *</td>
</tr>
<tr>
<td>e.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>: **!</td>
</tr>
<tr>
<td>f.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>*! : *</td>
</tr>
<tr>
<td>g.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>*! : *</td>
</tr>
<tr>
<td>h.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>*! : *</td>
</tr>
<tr>
<td>i.</td>
<td>[aakk\text{a}s]</td>
<td></td>
<td>*! : *</td>
</tr>
</tbody>
</table>

The most interesting candidates in (444) are (b), (e) and (h). (444b) is the optimal candidate, evincing underapplication of Aspiration as required by highly ranked \( \text{V\text{-IDENT[asp]}} \). (444e) loses to (444b) by virtue of incurring an additional and fatal mark on the constraint \( \text{LarCond} \) — one mark each for the virtual and the surface candidate. (444h) is the transparent competitor, since it implements aspiration of the geminate which is expected to result from the deletion of the voiced stop under conditions of normal application. However, in doing so, it incurs a fatal mark on \( \text{V\text{-IDENT[asp]}} \), which requires the surface form to have the same value for [aspiration] as the virtual form.
10.3 Prestopping and glottalization

This section deals with the complex phonology of prestopping and glottalization in nasals. §10.3.1 addresses prestopping. §10.3.2 deals with glottalization.

10.3.1 Prestopping

As argued in §7.1.8, the surface pattern of prestopping in West Finnmark Saami is properly construed as the combined effect of two processes, a rule inserting a bare obstruent root node, and a rule delinking the nasal from the coda, displacing the nasal into the onset position of the following syllable.

As argued for preaspirated stops above, I will assume that the monosegmental status of prestopped nasals is captured by coindexing of the prestop with the following nasal, to give [bi, jmi, d_i, jni, jn]. In this way, Fortition from plain Q1 to prestopped nasal in Q2 counts as violating Wt-Ident, in the same way as preaspiration, and endows monosegmental status to the cluster.

Prestopping may be understood as a response to a context-sensitive markedness constraint *VoralN, given in (445), which bans oral vowels immediately preceding a nasal.

(445) *VoralN
An oral vowel immediately preceding a nasal is disallowed.

An important question to address at this stage is the issue of marking vs. blocking. Prestopping applies only to geminate nasals, not to intervocalic singleton nasals, and we must decide whether this restriction is appropriately written into the constraint itself, or an instance of blocking by a higher-ranked constraint. One way or another, we must ensure that the grammar does not produce prestopping in Q1.

Take an input such /joña/ 'lingonberry'. In the Strong Grade, the stem undergoes Fortition and prestopping to jocñaa. In the Weak Grade, for example the nominative plural, we get jogaah, without prestopping. Yet, given (445), we would expect, counterfactually, that prestopping should apply here too, to give *joc.paaah, or *jo.'paaah. There is apparently a parallel here with the phonology of preaspiration: postvocally in Q1 is precisely the environment in which preaspiration fails as well. It was suggested above, however, that the reason for this was marking: the constraint driving preaspiration simply doesn’t require preaspiration of voiceless stops.

We might consider invoking similar reasoning to explain the failure of nasal prestopping in the same prosodic environment and relativize (445) to geminate nasals. Other things being equal, this is probably what we should do. But other things are not equal. Crucially, singleton nasals do undergo prestopping in another environment: to wit, word-internally following an excrescent vowel. Here, prestopping applies exceptionlessly and productively, e.g. kur.ro.pmaa 'gorm (larval parasite)', paar.ro.tnaa 'boy', jall.ñaah 'tree stumps'. The productivity of the process is confirmed by the behaviour of recent loanswords from Norwegian, such as filho.pnaa 'film', telefuB.Btnaa 'telephone', tar.ro.pnaa 'term'.

Prestopping in this position would seem to be motivated by the same constraint as (445), yet the target of the process is not geminate. If the failure of prestopping in Q1 cannot be built into the constraint, this leaves blocking by a
higher-ranked constraint as the explanation. (445) must be dominated by some other constraint.

The answer would seem to lie with CODA MAX. Recall from §9.3.3, that complex onsets are only permitted following a minor syllable. The reason was one of opportunism: this is the only environment in which Coda Maximization is precluded from applying in principle, and so syllabification reverts to an onset-maximizing mode. By CODA MAX, the prestop would be syllabified in the stressed syllable: jocñaah would be deemed more harmonic than jo.çnaah. However, a foot-internal coda must be moraic, either universally or because of highly-ranked WEIGHT-BY-POSITION (Hayes 1989). This leads to a violation of *HCC, which bans moraic consonant centres. Complex onsets are only permitted in case Coda Maximization is not an option, and only when the first component of the onset (s or h) is underlying.

In many languages, satisfaction of (445) is achieved by nasalizing the vowel. In Saami, however, satisfaction of *VoralN is achieved by prestopping. The markedness constraint militating against nasalized vowels, *Vnas, must accordingly be highly ranked. The constraint is stated in (446).

\(446\) *Vnas

Nasal vowels are disallowed.

I will assume that prestopping violates a low-ranked structural well-formedness constraint *PRESTOP, stated in (447).

\(447\) *PRESTOP

Nasals must not be prestopped.

As can be seen from (299) and (300), the voicing specification of the epenthetic stop differs in Q3 and Q2. In Q3, the epenthetic stop is voiced, but in Q2 it is voiceless. While in Q3, the stop fills a dual syllabic role as both the coda of the stressed syllable and the head of the minor syllable, in Q2 the stop is associated exclusively to the coda position. I will assume that this configuration is marked by the ad hoc constraint *CODAVoi.

\(448\) *CODAVoi

Voiced stops are disallowed in coda position unless also associated to the head of a minor syllable.

But there is still a crucial part of the picture to fill in: nothing in what we have said so far is sufficient to rule out the epenthesis of a voiceless stop in Q3. That is, given an input with a Q3 nasal such as /ean:nii/ `mother’, we have to be able to capture the fact that the output candidate eadnii is more harmonic than the competing candidate eanmii. In order to capture this, let us also assume that *ObsVoi interacts with a context-sensitive markedness constraint which requires stops preceding nasals to be voiced. *C[−voi]N is given in (449).\(^4\)

\(449\) *C[−voi]N

The sequence of voiceless consonant followed by a nasal is disallowed.

\(^3\)Under the assumption that violation is minimal, the inserted consonant will always share the same place of articulation as the following nasal. Given an input /an/, the candidate abn will be harmonically bounded by adn, since the former adds a violation of Dep-Place.

\(^4\) *C[−voi]N is the inverse of the *NÇ constraint discussed by Pater (1998) and others.
The final aspect of the pattern to be accounted for is Nasal Delinking, i.e. along which dimension of well-formedness is \( tn \) more harmonic than \( tn' \)? The answer is the Sonority Sequencing Condition (SSC) in (417), which, as we have already seen in §9.3 is undominated in West Finnmark Saami.

The interaction is shown in the tableau in (450). Of the candidates which do not violate one of the top-ranked constraints, (450b) and (450d), (450b) bests (450d) by faring perfectly on \( \text{CODA}V\text{O}i \).

(450) Epenthesis of voiceless stop in Q2

<table>
<thead>
<tr>
<th>/suonna/</th>
<th>( \text{V}_{\text{nas}} )</th>
<th>( \text{V}_{\text{oral}} )</th>
<th>SSC</th>
<th>CODA</th>
<th>( \text{CODA}V\text{O}i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. suon.na</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. suot.na</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. suotn.na</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. suod.na</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. suodn.na</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. suon.na</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now let us turn to the epenthesis of a voiced stop in Q3. Since, by hypothesis, voiced stops are penalized specifically only in Q2, the pattern of voiced stop epenthesis in Q3 follows automatically from the ranking established above. This is shown in (451).

(451) Epenthesis of voiced stop in Q3

<table>
<thead>
<tr>
<th>/ean:nii/</th>
<th>( \text{V}_{\text{nas}} )</th>
<th>( \text{V}_{\text{oral}} )</th>
<th>SSC</th>
<th>CODA</th>
<th>( \text{CODA}V\text{O}i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ean:nii</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. eat:nii</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. eatn.ni</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ead:nii</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. eadn.ni</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. ean.ni</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Of the two frontrunning candidates, (451b) and (451d), \( \text{CODA}V\text{O}i \) does not assign a mark to the voiced prestop in the optimal candidate (451d), since the prestop is in Q3 and not Q2. The task of arbitrating between the two passes
10.3. PRESTOPPING AND GLOTTALIZATION

all the way down to the constraint $^{*}C_{\text{vol}}N$, which assigns the fatal mark to (451b).

Words with an initial nasal, such as nāmma 'name' are exempt from prestonning. Descriptively, it seems to be the internasal context that serves as the blocking domain for prestonning. On first appearances, it would seem we can encode the pattern into a highly ranked constraint $^{*}NV_{oral}N$, given in (452), which penalizes oral vowels between nasals.\footnote{I would like to thank David Odden for discussion of this point.}

(452) $^{*}NV_{oral}N$

An oral vowel between nasals is disallowed.

Considerations of the Richness of the Base, however, cast doubt on this being the correct interpretation. The reason is that the input must also contain forms such as /napma/ in which the foot-medial nasal is already prestonned. Such forms fail to incur the necessary mark from (452) since a stop intervenes between the vowel and the following nasal, eliminating the internasal context for which (452) scans. Forms of this type are nevertheless systematically absent in the North Saami data, and we need a way of ensuring that outputs of this form are excluded. If we assume, rather, that the relevant constraint is as in (453), then this would have the effect of ensuring that prestonning in the internasal context is gratuitous: the nasalization on the post-nasal vowel ensures the satisfaction of $^{*}V_{oral}N$.

(453) $^{*}NV_{oral}$ (McCarthy and Prince 1995)

An oral vowel immediately following a nasal is disallowed.

Again, one possible repair for a (453) violation is nasalization of the vowel. The other possible repair is post-stopping. I assume that post-stopping is ruled out by an undominated constraint $^{*}Poststop$, (454), which penalizes mappings such as /m/ $\mapsto m\beta$, /n/ $\mapsto nd$, and so on.

(454) $^{*}Poststop$

Nasals must not be post-stopped.

Nasalization, then, applies under the ranking $^{*}NV_{oral}\gg^{*}V_{nas}$, as shown in (455).

(455) No prestonning following a nasal onset
Post-nasal vowel nasalization is enforced here due to the high rank of *NVoral. Of the leading contenders for optimal status, both (504e) and (504f), both satisfy *NVoral. (504e), however, is harmonically bounded by (504f); the former adds a gratuitous and fatal violation of low-ranked *Prestop, ensuring the victory of (504f). Epenthesis is pointless here: *VoralN is satisfied by the nasalization of the vowel.

If the mini-grammar just outlined genuinely reflects the synchronic grammar of West Finmark Saami, then cases such as seammaa, ‘same’ have to be treated as lexical exceptions, subject to a top-ranked and lexically specific faithfulness constraint γ-Dep-C, which penalizes stop epenthesis in diacritically marked inputs.

(456) Failure of prestopping with diacritically marked inputs

<table>
<thead>
<tr>
<th></th>
<th>γ-Dep-C</th>
<th>Vnas</th>
<th>VoralN</th>
<th>Prestop</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>seammaa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>seapmaa</td>
<td></td>
<td></td>
<td>!</td>
</tr>
<tr>
<td>c.</td>
<td>seimmaa</td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

The prenasal vowel in the optimal form is apparently not nasalized, providing a ranking argument for *Vnas ≫ *VoralN.

10.3.2 Glottalization and nasal release

Now that we have established which constraints trigger and block prestopping, let us turn to glottalization. The question to be addressed is this: is the presence of glottalization functionally motivated by the presence of prestopping? We know that glottalization may appear independently of prestopping; this is indeed precisely what happens in the youngest system in (301), and so, to answer this question in the affirmative is to cast this particular system in opaque terms.

I propose that the functional reason for the insertion of glottal stop in the Stop-Nasal context is the avoidance of nasal release, which understood as the release of pulmonic air into the nasal cavity on opening the velopharyngeal port. Nasal release can be avoided if the onset of pulmonic airflow and the abduction of the velopharyngeal port are out of phase with each other. One way of achieving this is by implementing a glottal closure at the crucial oral-nasal transition so that glottal closure overlaps with the latter portion of the oral stop and the initial portion of the nasal. When airflow is turned on again by glottal abduction, the velopharyngeal port is already open, and hence there is no nasal release.

The typology of glottalization involves the interaction of two constraints, *NASRel (457) and *?\C, (458).

(457) *NASRel
Nasal release is disallowed.

(458) *?\C
Glottalized consonants are disallowed.
The prestopping-only system, in (299), which has prestopping but no glottalization ranks \( ^*C \succ ^*\text{NasRel} \). The prestopping-with-glottalization system, in (300), which has both prestopping and glottalization, has the ranking \( ^*\text{NasRel} \succ ^*C \). The tableaux in (459) and (460) illustrate the ranking for Q2 and Q3 inputs respectively.

(459) Prestopping with glottalization (Q2)

\[
\begin{array}{|c|c|}
\hline
\text{/suonna/} & ^*\text{NasRel} & ^*\text{?}/C \\
\hline
\text{a. } & *\# & \text{suot}\_\text{na} & \text{!} \\
\text{b. } & \text{suot}\_\text{na} & \text{!} \\
\hline
\end{array}
\]

(460) Prestopping with glottalization (Q3)

\[
\begin{array}{|c|c|}
\hline
\text{/ean:nii/} & ^*\text{NasRel} & ^*\text{?}/C \\
\hline
\text{a. } & *\# & \text{ead}\_\text{t}\_\text{nii} & \text{!} \\
\text{b. } & \text{ead}\_\text{nii} & \text{!} \\
\hline
\end{array}
\]

Recall that most speakers implement Nasal Spreading in the Strong Grade so that \textit{ead\_nii} ‘mother’, for example, is realized as \textit{ean\_nii}. On the assumption that Glottalization is triggered by the presence of the prestop, then Nasal Spreading counterbleeds Glottalization, and the fact of the surface presence of glottalization is an example of the overapplication of Glottalization. At least, this is the interpretation that is forced upon us if we assume that glottalization is \textit{derived}, and arises to avoid a violation of \( ^*\text{NasRel} \) which would normally attend a cluster of stop+nasal. However, there is an alternative stance which renders the assumption of any opacity here unnecessary, although I am not convinced that it is ultimately the right approach.

Suppose that the youngest generation of speakers has simply reanalyzed the prestopped nasals as being glottalized nasals underlingly. If this is right, there remains little more to be said, except that glottalized nasals in Q3 and Q2 alternate with plain nasals in Q1. The absence of glottalization in Q1 has to be explained, although this is hardly insuperable. All we need is to assume that an underlying \( [+\text{constricted glottis}] \) nasal maps to \( [-\text{constricted glottis}] \) foot-medially in Q1 to satisfy an undominated markedness constraint \( ^*\ldots \text{V}^\text{\text{N}}[+\text{cg}]\ldots \text{V}^\text{\text{Ft}} \ldots \text{V}^\text{\text{asp}}\ldots \text{V}^\text{\text{Ft}} \), in much the same way we can assume a constraint \( ^*\ldots \text{V}^\text{\text{asp}}\ldots \text{V}^\text{\text{Ft}} \) to eliminate aspiration in singleton stops, should the grammar be fed such an input.

Historically, all three systems are indeed derived from plain nasals, and in rule-based serialism it is certainly possible to adopt the synchronic assumption that the opaquely glottalized nasals of the glottalization-only system have plain nasals as their underlying form. Consider the derivation by rule of \( \text{ead\_nii} \). The derivation more or less exactly recapitulates the relevant aspects of the diachronic series of changes which gave rise to the modern surface form.

(461)
CHAPTER 10. SEGMENTAL PROCESSES

<table>
<thead>
<tr>
<th>Input</th>
<th>eannïi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortition</td>
<td>eannïi</td>
</tr>
<tr>
<td>Prestopping</td>
<td>eadh:nnïi</td>
</tr>
<tr>
<td>Nasal Delinking</td>
<td>eadh:nnïi</td>
</tr>
<tr>
<td>Glottalization</td>
<td>eadh:onnïi</td>
</tr>
<tr>
<td>Nasal Spread</td>
<td>eanhânnïi</td>
</tr>
<tr>
<td>Other rules</td>
<td>ëæenânnïi</td>
</tr>
<tr>
<td>Output</td>
<td>[ëæenânnïi]</td>
</tr>
</tbody>
</table>

Taken as a whole, the derivation contains a Duke-of-York Effect, involving the openthes and subsequent deletion at a later stage of an obstruent stop. In OT extended with Sympathy Theory, there is no way of implementing this mapping, given the constraints, and so (/eannïi/ | [ëæenânnïi]) is not a possible input-output pairing. The depth of the derivation renders the connection between the input and the output too tenuous, and we can identify the precise cut-off point at which the input goes beyond the pale of recoverability: the stage at which prestopping applies.

On the other hand, if the input is assumed already to contain a prestopped nasal, this Duke-of-York effect disappears, and the surface presence of glottalization in the output may still be analyzed as opaquely conditioned (by the underlying stop).

Fixing prestopping in the input is unsatisfactory, though. As pointed out already, prestopping applies entirely exceptionlessly and productively following an excrescent vowel, and, intuitively, both kinds of prestopping result from the activity of the same constraint. Furthermore, fixing prestopping in the input leaves an important phonological gap unexplained, to wit, the complete absence of prestopped nasals following a nasal onset. On these grounds, it seems preferable to see prestopping as introduced by the operation of a process, rather than as present in the input, and it is more likely we are looking at something like a restriction on the underlying form. We return to these issues in more detail in the next chapter.

10.3.3 Clusters of liquid+nasal

In §7.1.2 and §9.2.1, we saw that sonorant-initial clusters underwent vowel excrescence in Q3. Strong Grade Q3 clusters of sonorant+nasal are also broken up by vowel insertion. Epenthes of a vocalic root node is driven by *SonC, stated in (407). Interestingly, Epenthesis would appear to be feeding a second application of Prestopping: recall from §7.3 that, in rule based terms, this can only be described in terms of the application of two identical rules widely separated in the derivation as in (360). From an OT perspective, both applications are driven by the same constraint. The prenasal position is epenthesized with a homorganic voiceless (glottalized) stop in satisfaction of *VoralN and *Nas-Rtl. Consider the tableau for the mapping /aajmuu/—vaaj jor p'muu 'air'. Only candidates satisfying the constraints on Consonant Gradation are shown.

(462)
10.4. LARYNGEAL SPREADING

Where there is no Coda Maximization, there is no Epenthesis, and without a vowel to condition the application of Prestopping, the insertion of a prestop is just gratuitous. Epenthésising a prestop will also result in comparatively worse performance on CodaMax. (463) illustrates the pattern for /vuojn.pah ‘spirits’.

Where there is no Coda Maximization, there is no Epenthesis, and without a vowel to condition the application of Prestopping, the insertion of a prestop is just gratuitous. Epenthésising a prestop will also result in comparatively worse performance on CodaMax. (463) illustrates the pattern for /vuojn.pah ‘spirits’.

Clusters of r+nasal pattern slightly differently from the rest of the set, because of a highly ranked constraint *RNas, given in (464), which penalizes the adjacency of /r/ and nasal.

(464) *RNas

/r/ must not be adjacent to a nasal.

(464) must outrank CodaMax. Consider the tableau for /paart.nii ‘boy’ in (465).

10.4 Laryngeal Spreading

Recall from §7.1.11 that, in triliteral sonorant-initial clusters in which the second member is /h/, the laryngeal glide acquires the supralaryngeal specification of the preceding sonorant (Sammallahti 1977: 39f.). I take the spreading to be
motivated by the OCP relativized to the Place tier. (For the development of the OCP in phonological theory, see Leben 1973; Goldsmith 1976; McCarthy 1986; Yip 1988; Alderete 1997.)

\[(466)\] \(\text{OCP}(\text{Pl}_\alpha)\)

Tier-adjacent identical place nodes are disallowed.

The effect of \(466\) is to force multiple linking of the place between the sonorant and the following stop. Spreading of the sonorant features through the root node sponsor of /h/ is driven by the universally undominated constraint against gapped configurations (see Chiosáin and Padgett 1997 for motivation and discussion).

\[(467)\] \(\text{NoGap}\)

Examples \((468)\) to \((471)\) illustrate the relevant candidate parses of the cluster /lh.t/. \((468)\) illustrates the prosodic structure of [l.ht], where h is parsed as an onset, only without spreading of the lateral’s supralaryngeal features. The resulting structure violates \((466)\).

\[(468)\] \([l.ht]\)

\[(469)\] illustrates the parsing of /h/ into the onset position along with spreading of the supralaryngeal features of the preceding lateral onto /h/.

\[(469)\] \([l.lt]\)

\[(470)\] In \((470)\), /h/ is parsed as coda, and without spreading of the supralaryngeal features of the preceding sonorant.

\[(470)\] \([lh.t]\)
10.4. LARYNGEAL SPREADING

Finally, (471) illustrates the parsing of medial /h/ into the coda along with spreading of the supralaryngeal features of the preceding sonorant.

\[
\sigma \quad \sigma \\
| \mu \\
| rt_1 \quad rt_2 \quad rt_3 \\
| \tilde{h} \\
P_l_a
\]

Structures such as the one shown in (472) below, which formally obey the OCP are ruled out on the grounds that they violate universally undominated NoGap.

\[
\sigma \quad \sigma \\
| \mu \\
| rt_1 \quad rt_2 \quad rt_3 \\
| \tilde{h} \\
P_l_a
\]

CODAMAX plays a crucial role in Q2 in attracting the aspirated sonorant over into the initial syllable. This is shown in (473).

(473) *Optimizing supralaryngeal spread in homorganic cluster (Q2)*

<table>
<thead>
<tr>
<th></th>
<th>OCP(P_l_a)</th>
<th>CODAMAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[lht]</td>
<td>*!</td>
</tr>
<tr>
<td>b.</td>
<td>[lht]</td>
<td>1, t!</td>
</tr>
<tr>
<td>c.</td>
<td>[lht]</td>
<td>*!</td>
</tr>
<tr>
<td>d.</td>
<td>[llt]</td>
<td>t</td>
</tr>
</tbody>
</table>

In the Strong Grade, CODAMAX cannot be satisfied, with the result that the aspirated sonorant ends up in the onset. Since CODAMAX is bound to be violated, it is the OCP that plays the decisive role in adjudicating between the relevant candidates. This is shown in (474).

(474) *Optimizing supralaryngeal spread in homorganic cluster (Q3)*

<table>
<thead>
<tr>
<th></th>
<th>OCP(P_l_a)</th>
<th>CODAMAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[lht]</td>
<td>*!</td>
</tr>
<tr>
<td>b.</td>
<td>[lht]</td>
<td>*!</td>
</tr>
</tbody>
</table>

In (474), the optimal candidate, (474b), evinces spreading of the features of /l/ to the root node of /h/, avoiding a fatal violation of the OCP.

Let us return briefly to the issue of directionality of spreading. The direction of feature spreading is rightward. Leftward spreading would entail the spreading of the stop’s features onto /h/. The reason for this is that spreading of
the stop’s features rather than the sonorants would violate $\text{Ident}_{\text{son}}[\text{rt3}]$.\footnote{Thanks to Ed Keer for discussion of this point.} Any candidate with leftward spreading from the stop will thus be harmonically bounded by the candidate evincing rightward spreading from the sonorant.

Supralaryngeal Spreading does not apply where the sonorant and following stop are heterorganic, because the OCP remains unviolated.

### 10.5 Voicing and Spirantization

In Q1, the dental and palatal stops undergo qualitative reduction to a spirant or approximant (see §7.1.6). I take this to be driven by the constraint $^*\text{VC}[−\text{voi}]$, where C is a member of the set \{t, c\}.\footnote{An alternative strategy would be to assume that the relevant markedness constraint is $^*\text{VC}[−\text{voi}]$, where C is simply the set of obstruent stops. This would indeed be consistent with the pattern of spirantization in other dialects of North Saami, where, say, all Q1 stops undergo voicing and spirantization. However, we would still have to account for the failure of the corresponding labials and dorsals to undergo the process. One way we might consider doing this is to follow Kirchner (1998) by assuming the existence of a set of context-sensitive faithfulness constraints which enforce preservation of [continuant] and [voice] in specific places of articulation, e.g. Lab-$\text{Ident}[\text{voice}]$, Dors-$\text{Ident}[\text{cont}]$, and so on. This, of course, necessitates positing certain things about the input: the input must be a voiceless stop, as opposed to a voiced stop or fricative.} $^*\text{VC}[−\text{voi}]$ must then dominate $\text{Ident}[\text{voice}]$. That is, we are primarily dealing with Voicing. So, how do we get the spirantization to follow as well? In order to account for the spirantization we need simply take advantage of the fact that the segments $d$ and $j$ are impermissible in Q1. (Recall that voiced stops are permitted \textit{exclusively} as the first component in a Q3 cluster.) Thus, we can make the spirantization follow by ranking $^*d$ and $^*j$ above $\text{Ident}[\text{cont}]$. In this way, spirantization is secondary to voicing. We can exclude the possibility that the relation between the two is the reverse of that suggested here. Suppose that the relevant markedness constraint was $^*\text{VC}[−\text{cont}]$. In this case, we would expect the intervocalic spirantization of /t/ to $\theta$. However, $\theta$ is a licit Q1 consonant centre (e.g. ruoqaas ‘in Sweden’), and it would not be clear in this case what would motivate the additional violation of $\text{Ident}[\text{cont}]$ in the mapping /t/→$\delta$.\footnote{Thanks to Ed Keer for discussion of this point.}
Chapter 11

Consonant Gradation

The aim of this chapter is to develop in some detail a constraints-based account of Grade Alternation.

In §11.1, we will return to the crucial question whether there is a directional bias in the Grade Alternation system. That is, is Consonant Gradation essentially Lenition or Fortition?

§11.2 provides a constraint-based analysis of Grade Alternation and establishes the basic rankings of the constraints.

§11.3 reintroduces the ‘primary’ and ‘secondary lengthening’ processes presented in §7.1.3 and §7.1.5.

§11.4 addresses the problems posed by Balance as an opacifier for Virtual Phonology.

§11.5 examines a case of multiple non-interacting opacity, whose existence is predicted by the architecture of Virtual Phonology.

§11.6 addresses some of the concerns about the Richness of the Base which arise at several points, and indicates how these worries may ultimately be resolved.

Finally, §11.7 presents the conclusions.

11.1 Consonant Gradation: Lenition or Fortition?

Logically, we may entertain three alternative hypotheses concerning the directional bias of Consonant Gradation.

1. Consonant Gradation is essentially Fortition-driven, in which case surface Q3 is ‘underlyingly’ Q2, and surface Q2 is ‘underlyingly’ Q1; Fortition is the active process (triggers quantitative enhancement), and, either Lenition essentially only blocks the application of Fortition in the relevant environment, or there is no such constraint as Lenition at all.

2. Consonant Gradation is essentially Lenition-driven, in which case surface Q2 is ‘underlyingly’ Q3, and surface Q1 is ‘underlyingly’ Q2; Lenition is the active process (triggers quantitative reduction), and Fortition is redundant.

3. Consonant Gradation is essentially bidirectional, mixing both Fortion and Lenition, in which case it shouldn’t matter what the input is: the
grammar ensures convergence on the correct result. Both Fortition and Lenition are active in the sense that both trigger quantitative adjustments in the input.

On the face of it, the bidirectional story might appear to be the most felicitous choice, given Richness of the Base. This increases the burden on the analyst to develop the constraints necessary to neutralize any 'rogue' inputs. Implementing Grade Alternation as the push-and-pull interaction of antagonistic constraints, however, turns out to be unworkable: there is no way around having to make assumptions about the underlying form, apparently in contravention of the Richness of the Base. In order to make this explicit, the emphasis will be shifted in this chapter away from the input and onto the 'underlying form' (U), which is here a representation forming part of the output. Unlike the input, the underlying form is necessarily unique. The underlying form stands in a relationship of correspondence to the other terms of the output array, and for the remainder of this chapter, I will abstract away from the correspondence relations between input and output, and frame the discussion instead in terms of UV- and US-Faith.

As we saw in Chapter 7, Fortition is to be favoured over Lenition on grounds of derivational economy. The Fortition Analysis is also supported by diachronic evidence. According to Sammallahti’s view (see Chapter 6), Grade Alternation arose from a process of Fortition which applied preceding an open syllable. Nevertheless, it has been usual, at least, it seems, tacitly, in synchronic descriptions to take the Strong Grade as basic and the Weak Grade as derived from the Strong Grade by a Lenition rule. Of course, the fact that Grade Alternation arose from fortition rather than lenition, to the best of current understanding, doesn’t automatically imply that Fortition is the appropriate way to construe the synchronic facts. Diachronic and synchronic questions are, after all, logically independent. We cannot exclude the possibility that Lenition is the correct approach. Still, the case for Lenition has never explicitly been made, and should we ever wish to put the case, it is well to bear in mind that we place an additional burden on the historical account. It implies that, at some point in Saami’s development, the alternation must have undergone reanalysis from a Fortition- to a Lenition-driven system. At present I am aware of no evidence which might be used to argue in favour of such a reanalysis having taken place. The null hypothesis should be, then, that no such reanalysis has occurred, and that Grade Alternation is still driven by Fortition. This has ramifications for what counts as a possible underlying form in Saami. Underlyingly, there is only a contrast between singleton and plain geminate: overlength is derived exclusively by Fortition.

This is the perspective I will adopt. The Lenition possibility will not be taken into account here. This is both for the reasons outlined above, as well as the fact that taking account of it would increase the complexity of the analysis by orders of magnitude. To do it justice would be beyond the scope of the dissertation.

11.2 Consonant Gradation: An OT Analysis

Now that we have addressed the fundamental question of directional bias, we have cleared the way to providing a constraints-based analysis of the Grade
11.2. CONSONANT GRADATION: AN OT ANALYSIS

Alternation patterns of West Finnmark Saami.

The core of Consonant Gradation consists in the interaction of at least three constraints, the faithfulness constraint \textit{Wt-Ident}, and the markedness constraints \textit{SingFort} and \textit{GemFort}, which drive the fortition of singletons and geminates respectively. \textit{SingFort} is essentially a special instantiation of \textit{CodMax}, which, as we saw in Chapter 9, is responsible for Coda Maximization in clusters. \textit{GemFort} must also dominate \textit{σ-Nuc} in (388), which militates against Q3.

This claim finds additional support in the behaviour of clusters under Grade Alternation in other dialects of Saami. In Ume Saami (Sammallahti 1998b: 23), the fortition of clusters is achieved through Coda Maximization, and not by the insertion of a mora into \textit{C} of the cluster, lending weight to the idea that \textit{CodMax} is deeply implicated in the process. So, can \textit{SingFort} be equated with \textit{CodMax} entirely? The answer is no, and that the reason is attributable to the difference in context-sensitivity between the two constraints. While \textit{CodMax} is context-free (blind to the closed vs. open prosody of the following syllable), \textit{SingFort} is context-sensitive, restricted to the environment preceding an open syllable, although it is identical to \textit{CodMax} in all other respects.

An important question concerns the identity of the Faithfulness with which the Fortition-driving constraints interact. There are two possibilities: \textit{Wt-Ident(\text{C})} in (475), and \textit{Dep-µ} in (476).

(475) \textit{Wt-Ident}

\begin{itemize}
  \item Let \(\alpha\) and \(\beta\) be segments, \(\alpha \in \text{Input}\), \(\beta \in \text{Output}\) and \(\alpha \mathcal{R} \beta\). If \(\alpha\) is \(n\)-moraic, then \(\beta\) is \(n\)-moraic.
  \item Evaluation: Assess one mark for each decrease or increase in \(n\) in the derivatum.
\end{itemize}

(476) \textit{Dep-µ}

\begin{itemize}
  \item Every mora in the Output must have a correspondent in the Input.
\end{itemize}

Recalling the discussion in Chapter 8, it should be clear that Singleton Fortition does not necessarily imply a violation of \textit{Dep-µ}. This is a consequence of our assumptions concerning the representation of hypercharacterized syllables (see §8.2.1). Heavy syllables are maximally bimoraic irrespective of the complexity of their segmental structure. Thus, the Fortition of the consonant centre in a word whose first syllable contains a long vowel or diphthong, say, \textit{/kualii/} to \textit{kualli} `fish', preceding the open syllable environment, does \textit{not} entail a violation of \textit{Dep-µ}. In moraic terms, the first syllable of the underlying form and the first syllable of the output form are the same. The consonant centre undergoes Fortition, but does so by associating to the weak mora of the first syllable. This weak mora is already present at the input.

This contrasts with the situation in words whose initial syllable contains a short vowel underlyingly. For example, \textit{/jasa/} undergoes Fortition to \textit{jassaa} `patch of unmelted snow which reindeer use to keep cool during the summer'. In this case, \textit{Dep-µ} is clearly violated in the consonant centre.

If \textit{Dep-µ} is only variably violated under Singleton Fortition, it is invariably violated under Geminant or Cluster Fortition. \textit{Dep-µ} is similarly violated under the implementation of Balance, another process which results in an overlong
CHAPTER 11. CONSONANT GRADATION

consonant centre. Although Dep-µ must be dominated by both SingFort and GemFort, the unfaithful insertion of a mora is not what all Fortition processes have in common. All of them have in common, however, that the consonant centre is unfaithful to its underlying weight specification or moraicity. That is, Wt-Ident is invariably violated as a result of implementing Singleton or Geminate Fortition. In sum, Singleton Fortition may or may not involve the insertion of a mora. Geminate Fortition and Balance always involve the insertion of a mora.

Now let us address the crucial question of the context-sensitive versus context-free nature of the Fortition-driving constraints. How exactly do we encode the truth that Fortition does not apply preceding the closed syllable environment? Once again, we are faced with a choice between blocking and marking. Is this effect to be understood as an instance of blocking by a higher-ranked constraint Lenition? If so, then we can formulate the relevant Fortition-driving constraints as context-free. They would approximate to saying something like ‘the consonant centre must be in the Strong Grade’. However, there is an alternative, and that is to construe the Fortition-driving constraints as context-sensitive, i.e. that they approximate to saying something like ‘the consonant centre must be in the Strong Grade preceding the open syllable environment’. If this is right, then the Fortition-driving constraints simply do not mark an unfortitioned consonant centre preceding a closed syllable, in which case recognizing Lenition as a constraint at all is superfluous.

There would seem to be no typological evidence bearing directly on the issue, except of course that if we eliminate the need for a separate Lenition constraint, we get tighter typologies as a result. So, while the proposal does involve postulating that the relevant Fortition-driving constraints are relatively more complex, there are factorial benefits to be reaped from the move. Indeed, the judicious path is always to minimize the number of constraints in the system, and this is especially true given the scarcity of the typological data.

I will assume, then, that there is no such constraint as Lenition, and that the Fortition-driving constraints are formulated in context-sensitive fashion, as in (477) and (478).

(477) GemFort (Geminate Fortition)
The foot-medial consonant centre must be in at least Q3 preceding an open syllable.

(478) SingFort (Singleton Fortition)
The foot-medial consonant centre must be in at least Q2 preceding an open syllable.

There are thus, minimally, two fortition constraints. (477) is responsible for the fortition of geminates, while (478) is responsible for the fortition of singletons. I will assume that the violation of SingFort entails the violation of GemFort. The question is where clusters fit into this picture. In Chapter 6, we saw that, in Ume and Pite Saami, clusters underwent maximization of the coda preceding an open syllable. On the evidence of these dialects, there would seem to be a binary choice in how to implement Fortition in clusters: either the coda can be enhanced to Q3, giving the West Finmark Saami pattern, or the onset can be enhanced to Q2 (by maximizing the coda of the preceding syllable), giving the Ume/Pite pattern. GemFort may be interpreted as the
requirement to enhance the coda of the stressed syllable, whereas \textsc{SingFort} may be interpreted as the requirement to enhance the onset of the unstressed syllable. In clusters, the two constraints, \textsc{GemFort} and \textsc{SingFort}, are thrown into conflict, and thus, the behaviour of clusters provides us with a ranking argument for the relative prioritization of \textsc{GemFort} and \textsc{SingFort}. In Pite and Ume, clusters undergo the coda-maximizing type of Fortition, diagnosing the ranking \textsc{SingFort} \textless \textsc{GemFort}. In West Finnmark Saami, on the other hand, clusters pattern with the geminates, undergoing coda enhancement to Q3, diagnosing the opposite ranking \textsc{GemFort} \textless \textsc{SingFort}.

Obviously, \textsc{Wt-Ident} must be low-ranked. (479) illustrates the ranking \textsc{GemFort} \textless \textsc{SingFort}, which forces the unfaithful mapping of an underlying geminate to overlong.

\begin{tabular}{|c|c|c|}
\hline
\textit{/kollii/} & \textsc{GemFort} & \textsc{Wt-Ident} \\
\hline
(a) kollii & * & \\
(b) \textbullet kollii & \\
\hline
\end{tabular}

(480) illustrates the ranking \textsc{SingFort} \textless \textsc{Wt-Ident}, which forces the unfaithful mapping of an underlying singleton to geminate.

\begin{tabular}{|c|c|c|}
\hline
\textit{/kuolii/} & \textsc{SingFort} & \textsc{Wt-Ident} \\
\hline
(a) kuolii & * & \\
(b) \textbullet kuolii & \\
\hline
\end{tabular}

Significantly, a Q1 input does not undergo fortition all the way to Q3. However, given the current rankings, there is no way of capturing this generalization. (481) illustrates the pathological result: an input in Q1, \textit{/kuolii/} is erroneously mapped in a total shift to Q3 in the output, giving *\textit{kuolii}.

\begin{tabular}{|c|c|c|c|}
\hline
\textit{/frownie kuolii/} & \textsc{GemFort} & \textsc{SingFort} & \textsc{Wt-Ident} \\
\hline
(a) kuolii & * & * & \\
(b) \textbullet kuolii & * & \\
(c) \textbullet kuolii & \\
\hline
\end{tabular}

As (481) shows, the desired winner, (481b), has a superset of the violations of (481c), with the result that (481c) is incorrectly optimized. According to standard reasoning, the undesired winner (481c) must be eliminated by some constraint of which we have not yet taken account.

According to Kirchner (1996), the blocking effect which characterizes chain shifts arises due to a highly ranked local conjunction of two faithfulness constraints within the local domain of the segment. The present case can be interpreted in terms of the local self-conjunction of the faithfulness constraint \textsc{Wt-Ident}, referred to as `powers' \textsc{Wt-Ident} or \textsc{Wt-Ident}^2. \textsc{Wt-Ident}^2 penalizes any candidate in which the consonant centre shifts more than one unit of moraicity, i.e. violates \textsc{Wt-Ident} more than once within the domain of the segment. Ranking \textsc{Wt-Ident}^2 above \textsc{GemFort} derives the chain-shift effect.
(482) Chain-shift effect due to highly ranked Wt-Ident\(^2\)

<table>
<thead>
<tr>
<th></th>
<th>/kuolii/</th>
<th>Wt-Ident(^2)</th>
<th>GemFort</th>
<th>SingFort</th>
<th>Wt-Ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. *</td>
<td>kuollii</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The distribution of the Weak Grade is simply accounted for by these rankings. By hypothesis, there is no Lenition constraint, so the Weak Grade is simply the default, in the sense that it doesn’t incur violations on any of the relevant markedness or faithfulness constraints, and so harmonically bounds the corresponding Strong Grade over the set of constraints. This follows from our formulation of the Fortition-driving constraints as context-sensitive: they simply do not assign a mark to a Q2 or Q1 consonant centre if immediately followed by a closed syllable. Compare the tableau in (483) and (484). (483) shows the derivation of /kuollii+s/-→/kuollii ‘fish-LOC.SG’.

(483) Q1 remains unchanged preceding closed syllable environment

<table>
<thead>
<tr>
<th></th>
<th>/kuollii+s/</th>
<th>SingFort</th>
<th>Wt-Ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. *</td>
<td>kuollii</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>kuollii</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

This point carries over to the alternation between Q3 and Q2. Consider in (484) the derivation of /kollii+s/-→/kollii ‘gold-LOC.SG’.

(484) Q2 remains unchanged preceding closed syllable environment

<table>
<thead>
<tr>
<th></th>
<th>/kollii+s/</th>
<th>GemFort</th>
<th>Wt-Ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. *</td>
<td>kollii</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>kollii</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

In (483) and (484) both SingFort and GemFort remain silent on the choice between the relevant candidates, and optimality is adjudicated by the lower-ranked constraint, Wt-Ident. Clusters receive identical treatment to geminates, undergoing Fortition to Q3 in the Strong Grade, e.g. /paste/-→/paste ‘spoon, NOM.SG’, as in (485), and failing to undergo Fortition preceding a closed syllable, e.g. /paste+s/-→/pastes ‘spoon-LOC.SG’ in (486).\(^1\)

(485) Cluster undergoes Fortition to Q3

<table>
<thead>
<tr>
<th></th>
<th>/paste/</th>
<th>GemFort</th>
<th>Wt-Ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. *</td>
<td>paste</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>paste</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

(486) Cluster Fortition blocked preceding closed syllable

<table>
<thead>
<tr>
<th></th>
<th>/paste+s/</th>
<th>GemFort</th>
<th>Wt-Ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>pastes</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. *</td>
<td>pastes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)By hypothesis, the coda is moraic universally, at least word-medially. This precludes the possibility of leniting the coda to non-moraic: amoraic codes are universally banned as part of the representational theory.
11.2.1 Morphologically conditioned grade alternation

Let us briefly turn to morphologically conditioned gradation. By hypothesis, all morphologically conditioned gradation is to be dealt with as opaque phonology. For example, the lexical representation of the surface null accusative/genitive singular suffix is thus -C, since this is the minimal phonological representation which is consistent with the fact that it subcategorizes for the Weak Grade. Fortition underapplies in the accusative/genitive singular because the consonant which blocks the process is absent on the surface. The failure of Fortition is thus appropriately relegated to the virtual form, and the underapplication arises due to the activity of highly-ranked constraints on virtual-surface identity. Specifically, \( \text{UV-MAX}-C \) dominates the markedness constraint which militates against ghost consonants. Underapplication is captured by ranking \( \text{VS-WT-Ident} \) above \( \text{GemFort} \) and \( \text{SingFort} \).

\[(487) \quad \text{Morphologically conditioned lenition as phonological underapplication}\]

At this point, it is appropriate to summarize the rankings established so far in a Hasse diagram.
CHAPTER 11. CONSONANT GRADATION

11.3 Balance

The most salient aspect of the phonology of sonorant-initial clusters and prestopped nasals is that they undergo Balance in the Weak Grade (see sections §7.1.3 and §7.1.5). Where Balance is properly conditioned (i.e. in the environment VVV in East Enontekiö, VV in Kautokeino), this gives surface alternations of the form $C_1:C_2 \sim C_1C_2:C_2$. If the cluster is triliteral, the surface alternation is between $C_1:C_2C_3 \sim C_1C_2:C_3$.

(489) Balance
A consonant centre preceding a long vowel must be in Q3.

11.3.1 Balance as an opacifier

There are two circumstances in which Balance acts as an opacifier. First, we have the testimony of sonorant-initial clusters. The requirements of Balance can in principle be met simply by implementing Geminate Fortition. This would be less costly in terms of Faithfulness as well. Given the input /arvu+s/ 'will', mapping to *arvuus only violates WT-Ident and Dep-$\mu$ once each; mapping to arvuus, on the other hand, violates Dep-$\mu$ once and WT-Ident twice. Put differently, over the set of markedness and plain IO-Faithfulness constraints of Classical OT, arvuus will be harmonically bounded by the ungrammatical candidate *arvuus. Nevertheless, arvuus is what we actually get. Since Balance and GemFort can in principle both be satisfied simultaneously, there can...
be no purely surface motivation for the peculiar pattern of mora redistribution we actually find: at some abstract level, Balance must be inapplicable. It seems appropriate to say that this level is the virtual level, in which case we have evidence for the ranking statement $\text{UNIV-WT-IDENT} \gg \text{Balance}$.

Further support for this position comes from the distribution of voicing in the prestop component of prestopped nasals. Amongst the prestopped nasals, voicing is present exclusively in the Strong Grade. Prestops undergo Devoicing in Q2, but this Devoicing rule is counterfeited by Balance, as established in §7.1.8. Fortition, however, must bleed the Devoicing rule, since in the Strong Grade, the nasal stop is voiced. The configuration is one of rule-sandwiching: the Devoicing rule is sandwiched between the Fortition and Balance rules, both of which are associated with identical faithfulness violations, in this case $\text{WT-IDENT}$. As was argued in §3.3, the appropriate way to analyze cases like this was to sandwich the relevant $\text{UNIV}$-Faithfulness constraint between the relevant markedness constraints. In the case at hand, the relevant markedness constraints are of course $\text{GemFort}$ and Balance. We can generate the observed pattern by invoking $\text{UNIV-WT-IDENT}(C)$, and interleaving it between $\text{GemFort}$ and Balance, giving the ranking $\text{GemFort} \gg \text{UNIV-WT-IDENT}(C) \gg \text{Balance}$. Note that this has consequences for the ranking we established earlier. Balance dominates $\text{UNIV-WT-IDENT}^2$. However, this immediately leads to a ranking paradox, since we have independently established that $\text{UNIV-WT-IDENT}^2$ dominates $\text{GemFort}$. If $\text{GemFort}$ dominates Balance, this requires that $\text{UNIV-WT-IDENT}^2$ be ranked simultaneously above and below Balance, a logical impossibility. $\text{UNIV-WT-IDENT}^2$ was instrumental in deriving the chain-shift effect noted earlier. The paradox can be resolved, however, if we attribute the chain-shift effect to $\text{UNIV-WT-IDENT}^2$. This means we have the ranking in (490). The constraints $\text{UNIV-WT-IDENT}(C)$ and $\text{UNIV-WT-IDENT}(C)^2$ may be taken to be bottom-ranked, and are not shown here.
The effect of the ranking is to render Balance inapplicable in the virtual form, setting up the necessary phonological conditions for a transparent application of the Devoicing process. The surface form /piepmuu/ 'meat' thus has a virtual counterpart /piepmuu/, which is protected from the effects of Balance by \( \text{UV-Wt-Ident} \gg \text{Balance} \). The overapplication of Devoicing can then be modeled as the result of ranking \( \text{US-Ident} \) [voi] high relative to the markedness constraint whose activity is checked, namely \( *C[\sim\text{voi}]N \). This constraint, repeated in (491) from (449), militates against a voiceless obstruent stop preceding a nasal.

\[
(491) \quad *C[\sim\text{voi}]N
\]
A voiceless consonant must not be followed by a nasal.

Balance applies on the surface, though, indicating that Balance must dominate \( \text{US-Wt-Ident} \). For the sake of expository simplicity, candidates with glottalization are suppressed, and only those candidates with prestopping are shown.
In the Strong Grade, however, fortition applies in both the virtual and the surface form. Unlike in the Weak Grade, Voicing is properly conditioned in the virtual form, and so underapplication is not an issue. The pattern emerges from ranking GemFort above VS-Wt-Ident.

<table>
<thead>
<tr>
<th>/piepmuːs/</th>
<th>VS-Wt-Ident</th>
<th>GemFort</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ⚫</td>
<td>⚫</td>
<td>*</td>
</tr>
<tr>
<td>b. ⚫</td>
<td>⚫</td>
<td>**</td>
</tr>
<tr>
<td>c. ⚫</td>
<td>⚫</td>
<td>*</td>
</tr>
<tr>
<td>d. ⚫</td>
<td>⚫</td>
<td>*</td>
</tr>
<tr>
<td>e. ⚫</td>
<td>⚫</td>
<td>*</td>
</tr>
</tbody>
</table>

Balance applies similarly in sonorant-initial clusters, except of course that the misapplication of a process is no issue here. Note that CodAMax must also dominate VS-Wt-Ident, otherwise Coda Maximization will not apply in the virtual form. Ensuring the application of Coda Maximization in the virtual representation is crucial, because the distribution of moras to segments is a combined function of the effect of VS-Wt-Ident and Balance. The fact that Balance applies in the surface form implies that VS-Wt-Ident must be dominated by Balance. However, this is incompatible with the ranking established earlier: if Balance dominates VS-Wt-Ident, then it follows from the transitivity of the domination relation that GemFort must dominate VS-Wt-Ident, since GemFort dominates Balance. However, the underapplication of Fortition seems to require that VS-Wt-Ident dominate GemFort, as in (488). The paradox indicates the involvement of more than just a single virtual level. We return to the problem in §11.4.2.
### 11. Consonant Gradation

(494)

<table>
<thead>
<tr>
<th>/arvuu+s/</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| a.  | arvuu
| b.  | arvuu
| c.  | arvuu
| d.  | arvuu |

For the grammar of quantity, we have established the rankings in (495).
11.4 Balance opacified

Balance itself is opacified in two sorts of situation. At least one of these situations is a counterexample to the Uniqueness hypothesis that there is a maximum of one virtual form. The other is a kind of underapplication but it is not obviously a counterexample, although it is deeply problematic nonetheless, because it appears irreducibly morphological. We enumerate these misapplication effects below.

1. Balance underapplies preceding a derived long vowel (because of a late rule of $a$-Lengthening, and in Kautokeino, $j$-Vocalization).
2. Balance overapplies preceding a derived short vowel (Kautokeino only, due to Preyotic $u$-Shortening).
3. Balance is blocked in the Weak Grade if it should result in a consonant centre identical in form to the Strong Grade.

We shall deal with the last of these, the morphological blocking effect, first, in §11.4.1. We return to the first two in §11.4.2.
11.4.1 Morphological blocking of Balance

As Sammallahti (1998b: 49) observes, Balance, specifically its Weak Grade instantiation ('secondary lengthening'), is blocked if the resulting consonant centre is identical to the corresponding Strong Grade. Thus, we don't find *hspitih for grammatical hspitih 'outlets', or *kollis for grammatical kollis 'in gold', and so on, even though this is precisely what we would expect given the constraint rankings as we have established them. (496) exposes the infelicity.

\[(496)\] Balance applies with undesired results

\[
\begin{array}{|c|c|}
\hline
/kollis\+s/ & \text{UV-Wt-Ident} \\
\hline
a. \mathbf{\not=} [kollis]\sim kollis & ;^*! \\
\hline
b. \mathbf{\otimes} [kollis]\sim kollis & \\
\hline
\end{array}
\]

**Balance** must plainly be dominated by some other constraint which blocks the application of the process.

I will more or less simply reify Sammallahti's descriptive statement as an ad hoc OT constraint (cf. Bye 1997b: 90), although I take advantage here of the result that Balance does not apply in the virtual form. Since Balance is inapplicable in the virtual form (because **Balance** is dominated by UV-Wt-Ident(C)), the virtual form maintains a contrast between the Strong and the Weak Grade. The constraint may be seen as requiring the preservation of this contrast in the surface form.\(^2\)

\[(497)\] GradeContrast (GradCont)

If there is a contrast between Weak Grade and Strong Grade in the virtual forms of a pair of alternating expressions in a paradigm, then there must be a contrast between Weak Grade and Strong Grade in the surface form.

Consider the derivation of kollis 'gold-loc.sg' again, taking into consideration the domination of **Balance** by **GradCont**. The corresponding Strong Grade form is kollis 'gold', providing the necessary foil for evaluating the performance on **GradCont**.

\[(498)\] Balance blocked by anti-identity

\(^2\)See Alderete (1999b) for an approach to similar phenomena in terms of transderivational anti-faithfulness.
11.4. BALANCE OPA CIFIED

\[
\begin{array}{|c|c|}
\hline
\text{/kollis}^+\text{s/} & \text{Grad=No-}\text{Wt-Ident} \\
\hline
\text{a. } \text{⇔ } [\text{kollis}]-\text{kollis} & \text{Grad=No-}\text{Wt-Ident} \\
\text{b. } [\text{kollis}]-\text{kollis}^{+} & \text{Grad=No-}\text{Wt-Ident} \\
\hline
\end{array}
\]

11.4.2 The underapplication of Balance: Uniqueness counterexemplified

*Uniqueness* (§3.1.1) predicts that interacting multiple opaque processes should be unattested in natural language. Yawelmani exemplified such a case. However, in §3.2.5, an attempt was made to fit the Yawelmani data to the Procrustean bed of the *Uniqueness* by building in both of the interacting opaque processes into the virtual form. Here we are confronted with a case which irreducibly consists in two interacting opaque processes, thereby furnishing a counterinstance to *Uniqueness*.

*Uniqueness* specifically entails that, if some process \(\mathcal{Q}\) renders some other process \(\mathcal{P}\) opaque, there can be no third process \(\mathcal{R}\) which renders \(\mathcal{Q}\) opaque.

We have seen in §7.1.8 that Balance opacifies Devoicing in prestopped nasals, with Balance filling the role of \(\mathcal{Q}\) and Devoicing the role of \(\mathcal{P}\) in the above rule schema. *Uniqueness* thus precludes the possibility of Balance itself being opacified by some other process. Nevertheless, this is precisely what we find.

Balance underapplies preceding a long \(\text{aa}\) (and, in Kautokeino, a long \(\text{ii}\)) which is underlyingly short. In §7.1.9, this pattern was given a rule-based explanation in terms of the counterfeeding of Balance by *a*-Lengthening (and, in Kautokeino, *j*-Vocalization; §7.1.10). We'll return to *j*-Vocalization in §11.4.3 and focus on *a*-Lengthening here for the purposes of developing the theory.

I will assume that the lengthening reflects the activity of the constraint in (499) \(^3\)

(499) *\(\text{VC}_0\)\(\text{â}\)

Short \(\text{â}\) is disallowed following a syllable containing a short vowel.

\(^3\)There are two possible interpretations of the *a*-lengthening phenomenon. One possibility is that it is a specific instantiation of the tendency, observed in other Finno-Ugric languages such as Finnish and Estonian, for vowels to undergo lengthening in foot-final position (Gordon 1997). In Estonian, it applies to all vowels, again only following a Q1 or Q2 consonant centre. The equivalent vowel in Estonian is traditionally known as 'halfLong', since its phonetic duration is approximately intermediate between the values for short and long vowels in stressed syllables. Estonian lacks any length distinction in the lengthening environment, and so considerations of contrast do not preclude that 'half-lengthening' in fact involves the addition of a mora. In Saami, however, latic lengthening is restricted to /a/.

The other possibility is that *a*-Lengthening has its functional basis in the maximization of the duration contrast between /a/ and the vowel of the preceding syllable. The rule only applies following a syllable containing a short nucleus. A similar phenomenon is observed in Inari Saami, which has a high ranking constraint *\(\text{VC}_0\)\(\text{V}\), which militates against short vowels in consecutive syllables. In the Inari Saami case, however, the lengthening which attends the satisfaction of the constraint may take place in either the stressed or unstressed syllable of the foot, revealing that the foot-final lengthening analysis (as motivated for Estonian) is inappropriate.
 CHAPTER 11. CONSONANT GRADATION

\*VC₉₉ must dominate Wᵀ-IDENT(V), although, now the crucial question of representational levels arises. A-Lengthening is clearly not a property of either of the ‘virtual’ or ‘surface’ levels we have identified. This is because Balance counterfeeds Voicing at the same time that Balance itself is counterfed by a-Lengthening, motivating yet another virtual level of representation. The representational level that we have identified here as ‘surface’ must constitute another virtual level, which we may designate as \( \mathcal{V}' \).

Let’s ground the discussion in the \( \mathcal{V}' \), \( \mathcal{V} \), and \( \mathcal{S} \)-representations envisioned for piep\\!\\!\\!m ‹u ‹ ‘food-acc/ gen.sg’ and jocñaa ‘lingonberry’.

(500)

\[
\begin{array}{c}
\mathcal{U} & /piemm ‹u ‹+s/ & /jop ‹a/
\\
\mathcal{V} & [piep\\!\\!\\!m ‹u ‹] & [jocñaa]
\\
\mathcal{V}' & [piep\\!\\!\\!m ‹u ‹]' & [jocñaa']
\\
\mathcal{S} & piep\\!\\!\\!m ‹u ‹ & jocñaa
\end{array}
\]

As presented here, the virtual forms actually replicate the crucial steps in the corresponding derivational account, and they stand in a total ordering defined by *cumulativity* with respect to the underlying form, designated here as \( \mathcal{U} \). That is, the \( \mathcal{S} \)-form accumulates the unfaithful mappings of the \( \mathcal{V} \)-form, and the \( \mathcal{V}' \)-form in its turn accumulates the unfaithful mappings of the \( \mathcal{V} \)-form. The underlying form is by definition fully faithful to itself, although, given *Richness of the Base*, there is no reason why the underlying form itself should not be unfaithful to its input (compare the discussion of American English vowel nasalization, Abercosh Welsh lenition, and Japanese ga-gyō variation in sections 3.4.1, 3.4.2, and 3.4.3 respectively). Thus, while the path from underlying form to surface form is monotonic, there is no such restriction on the path from the input to the surface form. Remember, the universal set of inputs to the grammar is not to be equated with the language-specific set of underlying forms.

Recall from §3.2.4 that, in Sympathy Theory, the sympathetic candidates do not themselves stand in a cumulative relationship to each other. On the Sympathy-theoretic understanding of opacity, therefore, intercandidate correspondence relations are not in any way a reification of the steps in a derivation (cf. Chen 1999).

I will assume that virtual representations do stand in precisely such a cumulative relationship both with each other and with the input, although for the moment, I will simply stipulate this as a restriction on \( \text{GEN} \), deferring discussion of the issues to §11.4.5 below. The principle, *Monotonicity*, formulated as holding of \( \text{GEN} \) is given in (501). Assume also that the output array also contains the underlying representation \( \mathcal{U} \) as its first member, such that the output is a linear array \( \{ \mathcal{R}_0, \mathcal{R}_i, \mathcal{R}_{i+1}, \ldots, \mathcal{R}_n \} \). \( \mathcal{R}_0 \) is the underlying form \( \mathcal{U} \), \( \mathcal{R}_n \) the surface form \( \mathcal{S} \), and all intermediate forms intervening between \( \mathcal{U} \) and \( \mathcal{S} \) are virtual (\( \mathcal{V} \)) forms.

(501) \textbf{Monotonicity}

In an output array \( \{ \mathcal{R}_0, \mathcal{R}_i, \mathcal{R}_{i+1}, \ldots, \mathcal{R}_n \} \), any decrease in faithfulness from \( \mathcal{R}_0 \) to \( \mathcal{R}_n \) must be monotonic.

Let’s address the constraint rankings which must hold in order for these arrays to have the form they do.
11.4. BALANCE OPAFICIED

We address the V-form first. Balance fails to apply in the V-form, (here, [piepmus]) as inscribed in the ranking \( \text{UV-Wt-Ident}(C) \) already established. This is behind the overapplication of Devoicing. A-Lengthening is also absent in the V-form, reflecting the ranking \( \text{UV-Wt-Ident}(C) \gg \text{VC}_0 \).

Balance applies in the V′- and S-forms. However, a-Lengthening is absent in the V-form, reflecting the ranking \( \text{UV-Wt-Ident}(C) \gg \text{VC}_0 \). A-Lengthening is also absent in the V′-form, reflecting the ranking \( \text{VC}_0 \gg \text{VC}_0 \).

Balance applies in the V′- and S-forms. However, a-Lengthening is deferred until the S-form, and does not apply in the V′-form. This is important, since Balance itself underapplies preceding a derived long vowel \( \tilde{a} \). If a-Lengthening applied in the V′-form, then Balance would apply, counterfactually, preceding any long vowel, whether underlying or derived.

Assume, as a null hypothesis, that, for every pair in the output array, there is a relation of correspondence between the members of that pair, as well as a correspondence relation holding between the input and each member of the output array. This means that, in addition to constraints on \( \text{UV}, \text{US}, \) and \( \text{VS}-\text{Faith} \), there are also constraints on \( \text{UV}, \text{US}, \) and \( \text{VS}-\text{Faith} \), and so on. From a formal point of view, identity between the terms of the array is characterized by a global potential. However, we have adopted the assumption that the representations in the output are structured as a linear array, and so the terms of the array may be more or less proximate. It seems sensible enough to assume that the faithfulness constraints which mediate between proximate (adjacent) terms of the output array are universally higher-ranked than the corresponding constraints mediating between remote (non-adjacent) terms. For example, this implies that \( \text{US-Ident voi} \) is universally higher ranked than \( \text{US-Ident voi} \). The V′- and S-forms are contiguous terms in the linear output array, in contrast to \( \text{V} \) and \( \text{S} \), which are non-contiguous. I will ignore the effects of the more ‘global’ faithfulness constraints in what follows, assuming they are universally too low-ranked to merit closer attention. The alternative is to assume simply, that there is no such global reference and the constraints requiring identity between remote terms simply do not exist.

The arrays in (500) are described by the following constraint rankings:

\[
\begin{align*}
\text{(502) } & \quad \text{Consonant weight} \\
& \quad \text{UV-Wt-Ident}(C) \gg \text{Balance} \gg \text{V′-Wt-Ident}(C), \text{US-Wt-Ident}(C)
\end{align*}
\]

\[
\begin{align*}
\text{(503) } & \quad \text{Vowel weight} \\
& \quad \text{UV-Wt-Ident}(V), \text{V′-Wt-Ident}(V) \gg \text{VC}_0 \gg \text{US-Wt-Ident}(C)
\end{align*}
\]

Given that \( \text{UV-Wt-Ident}(C) \) dominates Balance, Balance will fail to apply in the V-form, but if Balance dominates \( \text{V′-Wt-Ident}(C) \) and \( \text{V′-Wt-Ident}(C) \), then it will apply in the V′-form. Now, Devoicing applies in the surface form as an overapplication effect attributable to the ranking \( \text{US-Ident voi} \gg \text{C}_0 \). Given Monotonicity, it also has to be the case that this overapplication effect is in force in the V′-form as well. That is, it has to be the case that \( \text{V′-Wt-Ident voi} \gg \text{C}_0 \).

This ranking is illustrated in (504) using the mapping /namma/ → nammama ‘name: NOM.SG’. Only those candidates which satisfy \( \text{UV-Wt-Ident}(V) \) are shown.

\[
\begin{align*}
\text{(504) } & \quad \text{A-Lengthening}
\end{align*}
\]
CHAPTER 11. CONSONANT GRADATION

A-Lengthening fails to apply following a Q3 consonant centre. We can block its application by invoking a constraint militating against the sequence of a Q3 consonant centre followed by a long vowel.

(505) *C:CV
A long vowel following a consonant centre in Q3 is disallowed.

Ranking (505) above (499) will generate the correct result in this case, but it will also predict, counterfactually, across-the-board shortening of input long vowels following a Q3 consonant centre. Our analysis therefore has to capture the fact that it only derived long vowels are subject to (505): vowels which are underlyingly long surface faithfully following foot-medial Q3. It seems we have to invoke a local conjunction of the markedness constraint with the faithfulness constraint Wt-Ident to account for the pattern. Consider the treatment of /tsum:ma/ ‘kiss’. We ignore the virtual dimension here.

(506) A-Lengthening blocked following Q3 consonant centre

Importantly, the conjunction does not target non-derived long vowels following Q3, since they do not violate both conjuncts of the conjoined constraint. Contrast the behaviour of /il:laa/ ‘hardly’ with that of tsum:ma above. The vowel in question has to violate both faithfulness and markedness in order for this pattern of blocking to arise.

(507) No vowel shortening following Q3 consonant centre
11.4. BALANCE OPAÇIFIED

Now that we have established a working constraints-based analysis of the $\alpha$-Lengthening rule, let us return to the its opaque interaction with Balance. Consider the optimization of $t\text{f}al\text{maas}$ 'reticulum' (<$/t\text{f}al\text{maas}/$). Balance under-applies, indicating that BALANCE must be dominated by $\mathcal{F}$-$\text{Wt-Ident}(\mathcal{C})$.

11.4.3 Preyotic shortening

Recall from §7.1.10 that Balance is sandwiched inbetween two preyotic shortening rules. One rule of Preyotic $\iota$-Shortening bleeds Balance of its conditioning environment, while another, Preyotic $\upsilon$-Shortening, which is attested in Kautokeino only, counterbleeds Balance. The relevant output arrays are given in (509) for inputs $/moj\text{vii}+jt/$ 'mudder-ACC/GEN.PL' and $/kol\text{kuu}+jt/$ 'sauntering-ACC/GEN.PL'.
CHAPTER 11. CONSONANT GRADATION

The picture is further complicated by an additional rule, which also applies in Kautokëino, vocalizing a /ŋ/ following a latic /i/, counterfeeding Balance.

The relevant markedness constraints driving the preyoctic reductions in quantity are given in .

(510) *ui
Long ĭ preceding j is disallowed.

(511) *ūj
Long ā preceding j is disallowed.

Both (510) and (511) bring about violations of the same faithfulness constraint, Wt-Ident(V). In order for Balance to be bled by the shortening of preyoctic ĭ, it is sufficient that the ˧³-form unfaithfully parses the underlying weight of the vowel. (There is no evidence as to whether the ˧³-form parses vowel length faithfully or not.) Thus *ui > ˧³-Wt-Ident(V). Yet, at the level of the ˧³-representation, latic ā fails to shorten, diagnosing the ranking ˧³-Wt-Ident(V) > *ūj. The result is the overapplication of Balance preceding a short ā in the latus. (512) shows how this works for /kolk:kujt/ 'sauntering-acc/gen.pl'. For expository simplicity, the virtual form [kolkkuujt] is suppressed.

(512)

```
\begin{tabular}{|l|c|}
\hline
/kolkum+jt/ & ˧³-Wt-Ident(V) > *ūj \hline
\hline
a. *ui & kolkkuujt' \sim kolkkuujt & \cellcolor{white} \\
\hline
b. [kolkkuujt'] \sim kolkkuujt & *! : * & \cellcolor{white} \\
\hline
c. [kolkkuujt'] \sim kolkkuujt & : *!* & \cellcolor{white} \\
\hline
d. [kolkkuujt'] \sim kolkkuujt & : *! & \cellcolor{white} \\
\hline
\end{tabular}
```

_J-Vocalization is suspended in the ˧³-form. This is attributed to the effect of ˧³-Max-C dominating the markedness constraint *ūj.

(513) *ui
The sequence of short high front vowel followed by palatal glide is disallowed.

(514) shows how this works for /mojvvijt/ 'muddle-acc/gen.pl'. For expository ease, the virtual form [mojvvijt] is suppressed.

(514)
11.4. BALANCE OPACIFIED

A very real concern involves the threat of proliferation in virtual forms in the output array once *Uniqueness* is relinquished. *Uniqueness* at least reigned in this number, since each output contained at most one virtual form. Sacrificing *Uniqueness* implies giving up any principled restriction on the number of virtual forms in the array: the cardinality of the output array may be arbitrarily large.

One way of dealing with this is to reformulate the virtual faithfulness constraint in the following way. The semantics of a classical faithfulness constraint are as in (515). Given some input property $\phi$,

\[
\begin{array}{c}
\text{Semantics of faithfulness constraint in Classical OT} \\
\text{For all } \phi, \phi \in I, \text{ there is a correspondent of } \phi, \phi' \in S.
\end{array}
\]

Here, $S_1$ is the input; $S_2$ is the output. Suppose that virtual faithfulness constraints are formulated as in (516).

\[
\begin{array}{c}
\text{Revised interpretation of semantics of faithfulness constraint} \\
\text{There must be a virtual form, } \mathcal{V}_i, \text{ such that for all } \phi, \phi \in I, \text{ there is a correspondent of } \phi, \phi' \in \mathcal{V}_i.
\end{array}
\]

In (516), the core of the faithfulness constraint is embedded in an existential construction. Rather than constituting an obligatory component of every output array, virtual forms are freely generated by GEN, and there is no principled limit on the number of virtual representations in the output array. The situation thus resembles epenthesis: in principle, GEN is free to posit any amount of epenthetic structure. However, candidates evincing gratuitous epenthesis will never surface due to the fact that violation in the OT grammar is minimal (cf. Prince and Smolensky 1993: 27).

On this alternative interpretation, the learner does not posit virtual structure unless there is surface evidence which requires him to do so. In the limiting case,
then, there is no virtual form, and the output array consists of a bare surface form.

11.4.5 The Fear of Overgeneration

In §3.5, it was suggested that the appropriate way to understand cumulativeness was in terms of learnability. Recall the hypothetical language Tiberian Hebrew', in which the input /deʃ?/ `tender grass' was mapped in the output to the array ⟨[deʃ?], deʃ⟩. The crucial point is that the virtual form in this case undergoes phonological processes for which there is no surface evidence: the surface form is transparent, and so there is no reason to posit the application of an eponthesis rule to break up the consonant cluster. In the event that the surface form fails to accumulate the faithfulness violations of the virtual form, the phonology of the virtual form is unlearnable. If this is correct, writing cumulativeness in to the formalism of the grammar itself is superfluous. In real Tiberian Hebrew, we have the array ⟨[deʃ?], deʃ⟩, and the surface form accumulates the virtual form’s violation of DEP-V. The reality of cumulativeness is as a global dependency between constraint rankings. That is, there is an implicational relationship between two distinct ordered constraint pairs, which in principle may be ranked differently. Nevertheless, the violation of Ψ-Dep-V (as a consequence of the ranking *? > Ψ-DEP-V) carries over to the surface form, implying that Ψ-Dep-V must also be violated (as a consequence of the ranking Ψ-MAX-Ψ > Ψ-Dep-V). This is necessary if the opaque pattern is to be acquired at all. Since permitting the constraints can generate Tiberian Hebrew', the model is open to the objection that it overgenerates. This criticism is misplaced, though, because the dubiousness of Tiberian Hebrew' owes itself to its unlearnability, not any formal restriction.

Extending this line of reasoning to Duke-of-York effects involving multiple opacity is rather more delicate, since we have to recognize the existence of more than one virtual form. By hypothesis, the output array is linear, and each of its terms stand in a relationship of correspondence to each of the others. These relationships may be more or less proximate. Also by hypothesis, the faithfulness constraints mediating identity between representations which are contiguous in the array universally outrank any faithfulness constraint which mediates between non-contiguous terms of the array. Consider an alternative scenario, though, in which the faithfulness constraints mediating remote terms of the output array may be ranked anywhere in the hierarchy relative to the corresponding proximate faithfulness constraints. Recall in this connection the interaction of Vowel Harmony and High Vowel Lowering in Yawelmani addressed in §3.2. A form such as ?oθun `steal-nonfut' is derived from underlying /?uθ+hin/ first by a rule of Vowel Harmony, to give intermediate /?uθun/. Subsequently, a rule applies to lower a long high vowel, giving /?oθun/, and finally, the stem vowel undergoes closed-syllable shortening to give the actual surface form /?oθun/. The decrease in faithfulness is monotonic as we progress through the array. The Ψ-form accumulates the violation of IDENT[Color] incurred by the Ψ-form. Vowel Harmony applies in the Ψ-form reflecting the exigencies of the ranking ALIGN-COLOR > Ψ-IDENT[Color], from (54) and (56). Inheritance of the IDENT[Color] violation is encoded into the ranking ΨΨ-IDENT[Color] > ΨΨ-IDENT[Color]: it is more highly valued in
11.4. BALANCE OPACIFIED

this language to carry the violation of faithfulness than it is to reinstate the underlying value of the feature.

The surface form evinces a misapplication of both Lowering and Vowel Harmony. The overapplied Lowering is inherited from the \( V' \)-form, in which the conditioning environment for Lowering is present. The surface form, \( \text{Po}\text{t}h\text{in} \), accumulates the violation of \text{Ident}[\text{high}] from the \( V' \)-form, which violates \( VV'\text{Ident}[\text{high}] \) because it is dominated by \text{Long}/\text{−High} from (65).

Where surface misapplication is at issue, the effects of ranking faithfulness constraints mediating remote terms high, giving undesired global effects, can be ignored on learnability grounds. Consider the overapplication of Lowering again. Inheritance of unfaithful lowering in the surface form is achieved by ranking \( VS\text{-Ident}[\text{high}] \) above \( US\text{-Ident}[\text{high}] \) (or \( VS\text{-Ident}[\text{high}] \) for that matter), but if the pattern is to be learnable, it could not be any other way. Restoring the underlying value here would be equivalent, in learnability terms, to not applying Lowering in the first place: on the surface evidence, there is no reason to posit it.

However, virtual misapplication is a different matter. In diagnosing the phonology of the virtual form, there is only the evidence of misapplication in the surface form to go by. There can be no surface evidence one way or another as to whether the virtual form actually accumulates the faithfulness violations of its predecessor or not. In this case, remote faithfulness constraints raise the spectre of pathology. Let’s consolidate this with a thought experiment based on the Yawelmani pattern.

If we allow remote faithfulness constraints to be ranked above their proximate congeners, then the surface form may inherit its suffixal colour from either the \( V \)- or the \( V' \)-form, it wouldn’t matter which. Suppose, though, that the \( V' \)-form is freed from the constraint of Cumulativity and that it doesn’t accumulate the suffixal rounding of the \( V \)-form. This gives us \( \text{Po}\text{t}h\text{in} \). This leaves the way open for the surface form to accumulate its suffixal colour from the non-cumulative \( V' \)-form, predicting transparent \( \text{Po}\text{t}h\text{in} \) as the surface form. However, given this surface form, the opaque generalizations are unlearnable anyway — the surface evidence for assuming an underlyingly long high vowel in the root and a process of Lowering applied to it vanishes.

More serious is the candidate \( \langle \text{Pu}\text{t}h\text{un}, \text{Po}\text{t}h\text{in}, \text{Pu}\text{t}h\text{in} \rangle \), where the surface form inherits its suffixal colour from the \( V' \)-form, and its root colour from the \( V \)-form. The surface form inherits the faithfulness violations of neither virtual candidate, rendering it doubly non-cumulative, and furthermore, it instantiates a spurious pattern which is not found elsewhere in the language. The corresponding short high vowel in the root would, of course, condition Vowel Harmony on the following suffix transparently, giving a bizarre possible contrast between \( \text{Pu}\text{t}h\text{in} (<\text{Pu}+\text{hin}) \) and \( \text{Pu}\text{t}h\text{un} (<\text{Pu}+\text{hin}) \). This kind of inheritance, then, is entirely dubious, yet it is what is predicted if we give free reign to remote \( VS\text{-Ident}[\text{color}] \). If remote faithfulness constraints are either not in \text{Con} or universally too low-ranked to worry about, then there is no way of generating this spurious pattern — the relevant property can only be inherited from the immediate predecessor. Eliminating remote faithfulness constraints or ranking them low frees the virtual form from the monotonicity requirement in (501), i.e. it doesn’t have to be stipulated. What does have to be stipulated, apparently, is the output array as an linearly ordered array. Unless the output array is linearly ordered, the distinction between proximate and remote of
course collapses.

OT's parallelism seems inherently at odds with the existence in natural language of the kind of derivational effects we have been considering. Once derivational stages are replaced by correspondences holding between static output representations, overgeneration looms, for the reason that we predict identity effects between terms which never arise in natural language. The onus is on the constraints-based theorist to come up with a plausible story why remote or global identity effects are unattested. There are two approaches to this problem: (i) the formal approach, and (ii) the acquisition-based approach. In the view of McCarthy (1999), overgeneration should be checked by incorporating additional formal constraints which eliminate the spurious patterns at source. I have argued that the correct explanation is to be sought elsewhere, in the theory of language acquisition. Indeed, cases of simple opacity involving the acquisition of a single opaque generalization would seem to indicate that something precisely like this was the case. More work remains to be done to extend this line of reasoning to cases of multiple interacting opacity. Future research on the algorithms learners bring to bear on the acquisition of opaque patterns are expected to cast further light on these issues.\(^4\)

11.5 The Phonology of Trisyllabic Stems

Sympathy fails to deal with multiple non-interacting opacities. These are interactions involving at least processes \(P, Q, R,\) and \(S\), such that \(P\) precedes and is opacified by \(Q,\) and \(R\) precedes and is opacified by \(S,\) but \(P\) and \(Q\) do not interact. West Finnmark Saami provides evidence that these exist. Trisyllabic nouns (§7.3) furnish an example of an interaction of this type. Trisyllabic nouns undergo Apocope in the nominative singular, creating the environment for Fortition to apply. Yet, Fortition does not apply, so Fortition is countered by Apocope. At the same time, Latic Lowering counterbleeds Monophthongization. Thus, the phonology of the virtual form has to be faithful to the input in two independent respects: (a) underlying consonants must be parsed (by \(\text{UV-Max-C}\)), and, (b) the underlying height of the latic vowel must be preserved (by \(\text{UV-Ident[high]}\)). Both are opaque, but they do not interact.

As we argued in §7.4 and §11.2, Fortition could be opaquely blocked by an underlying syllable-closing suffix. As an example we used the genitive/accusative singular suffix, by hypothesis underlyingly -C. Underapplication of Fortition is not the only effect of this suffix, but it is also implicated in the underapplication of Apocope (§7.2), lending further support to the abstract phonological approach.

We can reconstrue the rule-based account of Apocope given in §7.2 in terms of constraint interaction. Lexical words in Saami are subject to the requirement that they minimally contain one foot (syllabic trochee). Apocope, driven by (518), applies word-finally, except where this would lead to a violation of Word Minimality in (517).

\(^4\)We have been assuming that the distinction between the virtual and the surface form is one of designation: the surface form is designated as such. If Cumulativity is genuinely a global restriction on identity relations between terms of the output array, this stipulative designation may turn out to be reducible. The surface form can be found simply by inspecting which of the representations in the output array accumulates the unfaithful properties of the other representations in the array.
11.5. THE PHONOLOGY OF TRISYLLABIC STEMS

(517) \( WdMin \)
A lexical word must contain a foot.

(518) \(*Fina1-V\)
Vowels are disallowed final in \(PrWd\).

\( WdMin \) dominates \(*Fina1-V\), so, while the quadrisyllabic stem \(peajvvaädha\)ka-'sunshine' undergoes Apocope in the nominative singular to \(peajvvaädha\) (Sammallahti 1998b: 64), Apocope is blocked in bisyllabic stems, because of the high rank of \(WdMin\): the mapping /kihta/ → *kieh is sub-optimal, because the output violates the more highly-ranked constraint. Tableaux (519) and (520) show the interaction between the two constraints.

(519) **Apocope applies**

<table>
<thead>
<tr>
<th>/peajvvaädhaka/</th>
<th>(WdMin)</th>
<th>(*Fina1-V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\varepsilon) peajvvaädh</td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>b. peajvvaädhaka</td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

(520) **Apocope blocked in disyllable**

<table>
<thead>
<tr>
<th>/kihta/</th>
<th>(WdMin)</th>
<th>(*Fina1-V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\varepsilon) kieh</td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td>b. kieh</td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

Apocope also applies in stems which are underlyingly trisyllabic (§7.3). The failure to apocopate in the accusative/genitive singular of a trisyllabic noun is an opacity effect. The virtual form is consonant-final: the C of the accusative/genitive singular suffix is preserved because of high-ranking UV-Max-C, and this prevents the environment for apocope from being realized. The virtual form of the accusative/genitive singular of 'dog' is thus \([peanaka]C\). Although this environment for Apocope is present on the surface, Apocope fails to apply because of high-ranking VS-Max-V.

(521)
The form *tsiillosa reflects the activity of three further structural well-formedness constraints, which correspond to the rules of *§3.1. Monophthongization (357), and Latic Lowering (356) adduced in §7.3.

(522) *CVCuCV
Disallow *u in a non-final latus.

Since the repair for a latic *u is shortening, (522) must dominate UV-Wr-Ident(V).

(523) Diphthong Condition (DiphCond)
Diphthongs are disallowed preceding a latic syllable contain a short high vowel *i or *u.

(523) is the constraint driving Monophthongization. On the assumption that monophthongization results in the deletion of a vocalic root node, then (523) must dominate UV-Max-V.

(524) *{u \{i \u\}}
Disallow a [+high] short vowel in the latus.

Since Latic Lowering applies, (524) must dominate Ident[high].

Consider the virtual form of the word *tsiillosa ‘abuse-ACC/GEN.SG’. The virtual form must be *tsiillusaC. The ghost consonant provides the reason for the underapplication of Apocope, and the short high u in the latic syllable provides the impetus to monophthongize the diphthong in the first syllable. For expository reasons, we represent the optimization of the virtual form only in (525) below. Only candidates satisfying the constraints on Consonant Gradation are shown.

(525)
11.5. THE PHONOLOGY OF TRISYLLABIC STEMS

<table>
<thead>
<tr>
<th>/tsiellusa + C/</th>
<th>*CV</th>
<th>*CV-Max-V</th>
<th>*Ghost VS-Max-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tsiillusaC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. tsiillusa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. tsiellusaC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. tsiellusa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. tsiillusaC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. tsiillusa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. tsiellusaC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h. tsiellusa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. tsiillusaC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j. tsiillosa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k. tsiellusaC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l. tsiellusa</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now let’s deal with the relation between the virtual form as established in (525) and the actual surface form. The mapping to the surface form is mediated by VS-Max-V. This preserves the vowel in the surface form, forcing the underapplication of Apocope. In the surface form, virtual [u] is unfaithfully mapped to surface o, since short high vowels in latus position are disallowed in surface forms. Virtual [u] is faithful to its underlying height specification if we allow UV-Ident[high] to outrank *{u}[^i]. Surface u, however, is disallowed, hence *{u}[^i] has to outrank VS-Ident[high]. This renders the environment for Monophthongization non-surface-apparent, and so the fact that we find a monopthong on the surface rather than a diphthong must be an overapplication effect. Overapplication may be understood as being enforced by through the ranking VS-Dep-V ≫ VS-Max-V, on the assumption that diphthongization primarily involves some kind of epenthesis.

(526)
The non-interacting opacities may be distributed between more than one virtual form, but there is no evidence for this. I will assume both opaque generalizations are encoded into a single representation.

### 11.6 Concerns about the Richness of the Base

We have achieved these results at the cost of ignoring certain kinds of input, and so the grammar is fragile in the sense proposed in §2.4.1. For example, we have simply assumed that Q3 is not present in input forms. Taking these inputs into consideration would have greatly increased the complexity of the analysis, since we would have had to develop constraints to handle them. However, the reasons for neglecting these inputs go beyond the merely expository: admitting them into the grammar of Grade Alternation would force adding to an already burgeoning pantheon of constraints. Allowing inputs with a consonant centre in Q3, e.g. /kollii/ ‘gold’, for example, would have necessitated adding a constraint to drive Lenition in order to get out the fact that the consonant centre is in Q2 in preceding a closed syllable. We can extend the list to other examples of ‘rogue’ input. There is a trade-off at work here, and we have to decide which of the two possible options is the more costly. On the one hand, we can enrich the constraint set, with all that would entail for factorial typology, and on the other, we can curtail ‘underlying’ contrasts, apparently doing violence to the Richness of the Base.
11.6. CONCERNS ABOUT THE RICHNESS OF THE BASE

Ultimately, it appears OT must incorporate constraints on underlying forms. Once the theory of candidates is enriched by including virtual forms within the output, it is not difficult to see how restrictions on underlying forms can be incorporated into the model. The underlying form is simply a kind of virtual form, which serves as the raw material for alternation-inducing phonological processes.

It may be objected that all this really achieves is to sell OT’s output-orientation down the river, and for this reason it would be worthwhile to pursue the alternative tack of boosting the first line of defence by enriching Con. This strategy seems pretty untenable though. Although Lenition might be a plausible constraint on functional grounds, we would have to propose other far less natural constraints to perform the task of neutralizing other rogue inputs. In this event, it seems preferable to live with constraints on the underlying form.

Let’s consider a few examples of the kind of constraints we would have to contrive to do the job. It is not immediately obvious, for example, how we would deal with inputs combining Coda Maximization and Balance, such as /arv:vu:/ ‘will’. Such structures are well-formed in the language, yet, in the open syllable environment, we would have to contrive some constraint which demands the mapping to the appropriate distribution of length in the Strong Grade, e.g. /arv:vu:/ → /ar.v:vu:u:.

Inputs fully equipped with an excrecent vowel (sonorant-initial clusters in the Strong Grade) are another case in point. What constraint could we posit to drive the mapping /ar.ro.vuu+s/ → /arv:vuus/? What about inputs which are in Q4, e.g. kolklii? Presumably, these surface as alternations between Q3 and Q2. Q4 is neutralized to Q3 by an undominated constraint *Q4, but how do we ensure Q2 in the Weak Grade? If mappings such as these are to be direct and unmediated by any underlying form, we end up having to posit ad hoc constraints which are not phonetically sensible, as well as weakening the typology by adding constraints to Con. The problem is a general one. By Richness of the Base, possible output forms are also possible input forms. If constraints are required to mediate between alternating forms directly, we end up having to add constraints to the set, and the threat of combinatorial explosion looms. The Richness of the Base is simply unhelpful here. Countenancing certain forms overtaxes the grammar. The Saami data throw into stark relief a conflict of interest between reigning in the burgeoning constraint set and maintaining the Richness of the Base. The analysis presented in this chapter has erred on the side of grammar minimization at the apparent cost of doing violence to the Richness of the Base. The most judicious strategy for future research will be to refine the notion of the input. We have come across examples of language-specific restrictions on the input before, in §3.4. The paradox was resolved by introducing a further level of representation, identified as the virtual form. The crucial distinction is between the ‘input’ and the ‘underlying form’. Whereas rule-based theory conflates the two, in OT they are crucially distinct.

Whereas in Classical OT, the ‘underlying form’ is derived from the surface form by Lexicon Optimization (Prince and Smolensky 1993: 191ff.), it appears that certain facts render these minimalist assumptions untenable: the ‘underlying form’ must be admitted as a distinct term of the output array. While the mapping from the underlying form to the surface form must be cumulative, the mapping from the input to the surface form need not be.
11.7 Conclusions

After developing a fortition-based OT account of Grade Alternation, this chapter addressed two instances of rule sandwiching involving Balance (the distribution of voicing in prestops and preycotic shortening). It was also argued that Uniqueness of the virtual form cannot be maintained in the face of the existence of multiple interacting opacity. We argued instead for a replacement of the Uniqueness principle developed in Chapter 3 with a model in which the output array contains in principle an arbitrary number of virtual representations. This lays the theory prone to accusations of overgeneration, but there are strong indications that the solution to the problem of overgeneration lies within learnability as opposed to the formal theory. This can be shown unproblematically in the case of simple opacity, but more research needs to be done in order to carry over this result to cases in which more than one opaque process interact.
Chapter 12

Conclusion

This dissertation has addressed what is probably the central challenge to OT: how to square the existence of opaque interactions between phonological processes with OT's parallelism and output-orientation.

While there are other OT approaches to opacity, Sympathy Theory (McCarthy 1999) serves as the chief foil for launching Virtual Phonology. Both Sympathy Theory and Virtual Phonology go in for full parallelism: in contrast to Stratal OT (Kiparsky 1998), the grammar consists of a single constraint hierarchy. In contrast to Sympathy Theory, though, Virtual Phonology enriches the theory of the candidate for evaluation: the output is not a unique representation, but a linear array of forms over which correspondence relationships are defined. This linear array minimally contains a designated surface representation and may contain one or more virtual forms encoding opaque phonological generalizations.

Sympathy Theory and Virtual Phonology diverge in two predictions which crucially favour Virtual Phonology. First, it was argued that, contra McCarthy, rule sandwiching phenomena are in fact robustly attested in a variety of languages, including Yawelmani Yokuts (§3.3.1), Modern Hebrew (§3.3.2), Mohawk (3.3.3), Yidip (§3.4.6), and Saami (Chapters 7 and 11). North Saami in particular furnishes particularly striking evidence of the phenomenon. Sympathy Theory, however, cannot generate the rule-sandwiching pattern, and this flaw was traced to the fact that sympathetic selectors behave as if top-ranked for the purposes of sympathetic selection. In Virtual Phonology, the role of sympathetic selection is played by faithfulness constraints predicated on the correspondence between the input and the virtual form. There is an crucial architectural difference between the two theories in this regard, though, because, unlike the sympathetic selector, the $\mathcal{I}V$-faithfulness constraints are defeasible. It is by virtue of this defeasibility that the rule sandwiching phenomenon falls out unproblematically in Virtual Phonology.

Second, Sympathy Theory turns out to be incompatible with multiple opacity, whether of the interacting or non-interacting kind. An examination of multiple opacity in Yawelmani Yokuts in §3.2.4 revealed that the Sympathy-theoretic approach to multiple opacity was fatally flawed.

Whereas multiple non-interacting opacities can be dealt with unproblematically within Virtual Phonology, multiple interacting opacity can only be captured provided that the output array may include more than one virtual form.
Chapter 3 started out with the *Uniqueness* restriction that the number of virtual forms in the output array was precisely one. This restriction was relinquished in Chapter 11 in the face of facts from West Finno-Saami.

Opacity is also implicated in two other challenges to the OT vision of grammar. These are (i) morphologically conditioned phonological alternations, and (ii) language-specific restrictions on the input. Morphologization is a problem for OT’s universalism and bias towards functionally grounded constraints, while the existence of apparent language-specific restrictions on the set of inputs to the grammar conflicts with the principle of the *Richness of the Base*.

The perspective taken here on morphologized alternations is that they are different in degree, but not in kind, to opaque phonological alternations. Morphologized alternations are properly analyzed in the same terms as phonological opacity. As McCarthy states (McCarthy 1999: 383), rejecting the stance adopted by some researchers, notably representatives of Natural Generative Phonology such as Hooper (1976): “opaque generalizations have exactly the same character as transparent generalizations, except for being opaque.” It is tempting to rework this statement to make an analogous point about the status of morphologically conditioned alternations: “morphologized generalizations have exactly the same character as opaque generalizations, except for being morphologized.” Just as the Natural Generative Phonologists took opaque generalizations to be somehow different in kind to transparent generalizations, it is entirely current to see morphologized alternations as somehow qualitatively distinct from overtly phonological ones. This stance has characterized most twentieth century thinking about morphologized alternations, and this dissertation represents a break with this tradition. It is admittedly a radical stance, and unlikely to appeal to some. The viability of the proposal depends, ultimately, on whether language learners actually do impute functional bases to the alternations they hear in the stimulus, and “reconstruct” appropriate conditions into abstract phonological representations.

Certain alternations only make sense given specific assumptions about the underlying form. This was the case with the Soft Mutation of $\delta z$ to $\delta f$ in Abersoch Welsh, considered in §3.4.2, and it is true for many of the Saami alternations we have discussed. It appears that, at some level, OT must allow for restrictions on the ‘underlying form’. Although we have not developed an articulated theory to deal with cases such as this, Virtual Phonology does at least hold out the promise of a solution to this problem.

The *Richness of the Base* is not an empirical hypothesis about the set of inputs to the grammar, but a logico-methodological assumption. As such it is a non-negotiable tenet of Optimality Theory. Assuming certain types of input to be systematically absent is therefore arbitrary and unsatisfactory. Ultimately, the theory of grammar must find ways of dealing with rogue inputs.

Objections may be raised to the effect that what has been done here is ultimately nothing more than a reification of the derivation. As Johnson (1972) has shown, any derivation may be computed non-derivationally, since the crucial stages of the derivation may be reified as distinct representations. The rules themselves may then be replaced by constraints on the relationship between these representations. The translation of the derivational to a static constraints-based account is trivial in itself, and the question arises as to whether anything significant is gained in this way. Indeed, the move raises the spectre of globality: since all of the relevant representations are copresent, we either have to stipulate
which representations correspond and which don’t, or we have to neutralize any unwanted global effects in some way. Put like this, why *not* revert to rule-based serialism?

There are some very cogent reasons why not: Optimality Theory is still the best theory for dealing with phenomena such as conspiracies, rule persistence, and opportunistic application. Irrespective of derivationalism’s adroitness in handling opacity, it simply has nothing insightful to say about any of these phenomena. West Finnmark Saami, in addition to being extraordinarily rich in opaque interactions, offers some striking examples of all three kinds of interaction, and it might be appropriate to remind ourselves what these are.

- **Conspiracy**
  Gemination and shift of syllable boundary conspire in satisfaction of CODA\textsubscript{MAX} in §9.3.

- **Persistence**
  Vowel Apocope may apply twice in the derivation, once before the application of Fortition, and a second time post-lexically (§7.3). Prestopping is another example of this kind (§7.1.8 and §10.3).

- **Opportunism**
  Several processes apply opportunistically, when CODA\textsubscript{MAX} cannot be satisfied. These are Onset Maximization and Epenthesis in §9.3.3. Both illustrate the emergence of lower-ranked constraints.

This is the first attempt at providing a generative analysis of any significant portion of Saami phonology. Very much work remains to be carried out, and Saami offers a typological diversity which should provide an excellent testing ground for phonological theory for many years to come. It is hoped that future work will map the typology of grade and quantity in all dialects of Saami in order to attain a better understanding of these fascinating phenomena as well as the deep questions concerning grammar they invite us to contemplate.
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


Hale, Mark, and Charles Reiss. 2000. Phonology as Cognition. Ms., Concordia University, Montréal.


BIBLIOGRAPHY


BIBLIOGRAPHY


