COMPLEXITY AND MARKEDNESS IN OPTIMALITY THEORY

by

Trisha Kathleen Causley

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Linguistics,
University of Toronto

© Copyright by Trisha Kathleen Causley 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-41121-4
Abstract of the Dissertation

Complexity and Markedness in Optimality Theory

by

Trisha Kathleen Causley
Doctor of Philosophy 1999
Department of Linguistics
University of Toronto

This thesis explores the issue of segmental representations in Optimality Theory. I make two claims relating to this issue. First, I argue that input segmental representations may vary from language to language, and that these differences influence the patternning of different segment types in the phonology. Second, I argue that segmental markedness is evaluated in terms of representational complexity: the more complex a representation is, the more marked it is. Markedness relations are thus encoded in the segmental representations, so that marked segments involve more structure than unmarked segments.

The variation that is found in segmental markedness across languages is expected under the proposed theory, since representations are constructed in response to contrasts in the inventory. Since segmental representations are influenced by contrast, variation in inventories entails variation in segmental representations.

The representations that I propose are built in response to the competing considerations of structure minimization and the need to make contrasting specifications representationally distinct. Structure is only added to a specification under the pressure of contrast. Since additional structure
means a more complex representation, and markedness is linked to complexity, the markedness status of a particular segment may vary depending on the number and type of contrasts in an inventory.

My proposal entails that input representations play a more central role in the phonology than is currently assumed in the Optimality Theory literature. I argue that, despite claims to the contrary, all cross-linguistic variation cannot be made to fall out from output constraint ranking alone. Instead, phonemic inventories are necessarily defined in the input, and therefore there is no universal set of inputs.
Acknowledgements

Finishing this thesis means the end of graduate school, a period covering the greater part of my adult life. A lot has happened in the past five years and a lot of people have made a contribution to my work, my training, and my survival in the programme.

One of the best things about being at U of T is the people that I had the great fortune to spend time with. The graduate students, in particular, made the department a fun and supportive place to hang out and to grow up. Thanks especially to Abdul-Khalig Ali, Craig Chambers, and Susan Gropp for the support during the year of Omigod-I-can't-believe-they-let-me-into-grad-school and for the years that followed. The following year’s new recruits paid us deserved respect as The People Who Knew Last Year’s Gossip, and they turned out to be pretty cool as well. So I’m thankful to the department for admitting decidedly wonderful people like Luba Butska, Julio Cebrian, Carolyn Smallwood, and Susanna Bejar. Of course, one of the more fortuitous meetings is the one I had with David Bennett, who signed himself up for a full-time position as my sounding board/crying shoulder/document formatter. Poor guy.

I have also had some really good guidance and support from various faculty-types. Keren Rice, my supervisor, has been a tremendous help to me throughout my time at U of T both inside and outside of her supervisory role. Keren is one of those rare people who seem to have boundless amounts of energy and, lucky for me, that energy can be just infectious enough to get you through tight spots.

Thanks to Ron Smyth for his friendship and valuable counsel, to Elan Dresher for helpful and insightful discussions about phonology and other stuff, and to the rest of the faculty in the department for various courses and assistance. Peter Avery was involved with my thesis from the earliest stages right up until the defense and I was thrilled to have his input and encouragement, and I was honoured that he would come all the way from York for countless meetings with me.

My thesis committee made the whole defense relatively painless and surprisingly unembittering experience. Stimulating comments and questions raised by Glyne Piggott, Emmanuel Nikiema, Elan Dresher, Peter Avery, and Keren Rice have given me directions to progress with this research.

Finally, I would like to thank my family for putting up with me for all these years, when it would have been so easy for them to just move away and not tell me.
Table of Contents

Chapter 1: Introduction ........................................................................................................... 1
  1.1 The relevance of inputs .................................................................................................. 2
  1.2 Markedness as representational complexity ................................................................. 4
  1.3 Linking markedness to inputs ....................................................................................... 5
  1.4 Structure of the thesis .................................................................................................. 6

Chapter 2: Markedness, Faithfulness and *Structure ......................................................... 8
  2.1 Max-Dep Asymmetries and *Structure constraints .................................................... 10
    2.1.1 Segmental Correspondence ................................................................................. 10
    2.1.2 Featural Correspondence .................................................................................. 12
    2.1.3 Max_{sy}lable and Dep_{sy}lable .................................................................... 16
    2.1.4 Max_{foot} and Dep_{foot} ............................................................................. 18
    2.1.5 Max_{phi} and Dep_{phi} ............................................................................. 19
    2.1.6 Summary ........................................................................................................ 20
  2.2 *Structure and Uni-Directional Faithfulness ............................................................... 22
    2.2.1 Dep and the Prohibition of Structure ................................................................ 22
    2.2.2 Markedness as non-Faithfulness *Structure constraint set .................................. 25
    2.2.3 Contrasting Views of Markedness .................................................................... 26
  2.3 On the role of Markedness .......................................................................................... 28
    2.3.1 Choice of Epenthetic Segments ....................................................................... 29
    2.3.2 Implicational relationships in inventories/positions ......................................... 29
    2.3.3 Shape of inventories ....................................................................................... 30
    2.3.4 Assimilation and markedness ........................................................................ 32
  2.4 Conclusion ................................................................................................................ 33

Chapter 3: A representational theory of features ................................................................. 35
  3.1 The representations ................................................................................................... 36
  3.2 Representational Complexity and Phonological Activity ........................................... 39
    3.2.1 Markedness and phonological activity ............................................................ 40
    3.2.2 Determining Inputs ....................................................................................... 47
  3.3 The Relevance of Input Specifications ........................................................................ 54
  3.4 Variability in Markedness .......................................................................................... 60
    3.4.1 C-Place Variability ....................................................................................... 61
    3.4.2 Variability in Markedness II: V-Place .............................................................. 67
    3.4.2.1 Front vowels as unmarked ........................................................................ 67
    3.4.2.1.1 Front vowels as phonologically inactive ........................................... 67
    3.4.2.1.2 Front vowels as epenthetic segments ............................................. 70
    3.4.2.1.3 Front vowels as reduplicative defaults ........................................... 70
    3.4.2.2 Back vowels as unmarked .......................................................................... 71
    3.4.2.2.1 Back vowels as phonologically inactive ........................................... 72
    3.4.2.2.2 Back vowels as epenthetic segments .............................................. 73
    3.4.2.2.3 Back vowels as reduplicative defaults ........................................... 74
    3.4.2.3 Central Vowel is unmarked ....................................................................... 74
    3.4.2.3.1 Central vowels as phonologically inactive ..................................... 74
3.4.2.3.2. Central vowels as reduplicative defaults..........................76
3.4.2.4 Markedness variability: summary ........................................76
3.4.3 Markedness variability and inventory ......................................77
3.5 Building Inputs: the relationship between *STRUCTURE and contrast .........................................................................................78
3.6 Phonetic Implementation and Inventory Optimization ...............85
  3.6.1 From Representations to Phonetic Realizations .....................85
  3.6.2 Flemming (1995) .................................................................87
  3.6.2 Dispersion Theory and Phonetic Realizations .......................92
  3.6.3 Going it with Dispersion Alone ..........................................95
3.7 Conclusion ..............................................................................97
Chapter 4: The problem of constraint-ranking as explanation ..........99
  4.1 Markedness variability and fixed rankings ..............................99
  4.2 Unexpected invariability: the problem of unexplained fixed rankings .......................................................................................105
    4.2.1 Factorial Typologies and free re-ranking ............................106
    4.2.2 Markedness and Phonological activity in Optimality Theory ...............................................................................................108
  4.3 Constraint ranking and lost generalizations ............................118
    4.3.1 Variability in Markedness II: V-Place .................................119
    4.3.2 Markedness variability and inventory .................................119
      4.3.2.1 Variability in emergence of the unmarked environments .......................................................................................120
      4.3.2.2 Variability in target and trigger behaviour ....................122
      4.3.2.3 Ranking of Faithfulness and relationship to inventory ...............................................................................................129
  4.4 Conclusion ..............................................................................131
5. Coalescence, Markedness, and the Relevance of Inventory .........132
  5.1 Coalescence and Input structure ..............................................133
    5.1.1 Preliminary Assumptions: A unified approach to Vowel Coalescence/Elision/Assimilation ..................................................133
    5.1.2 Front Vowels are Unmarked/ Back Vowels are marked .................................................................135
    5.1.3 Back Vowels are Unmarked/Placeless ..................................139
    5.1.4 Central Vowels are Unmarked/Placeless: Front and Back have place .................................................................141
    5.1.4 Summary ............................................................................148
  5.2 Where is the inventory? ..........................................................149
    5.2.1 Richness of the Base and the universal input set ..................149
    5.2.2 The relevance of input inventories ....................................151
    5.2.3 Phonemic inventories characterize the input ......................157
  5.3 Conclusion ..............................................................................158
6. Height Coalescence ....................................................................159
  6.1 The Patterns ............................................................................160
    6.1.1 Three height systems and e-coalescence .........................161
  6.3 An alternative approach to height specifications ....................170
Chapter 1: Introduction

This thesis deals with issues surrounding segmental representations in Optimality Theory and the role that they play in the phonology. In particular, it addresses the issue of segmental representation as it relates to markedness and markedness-linked behaviours of segments. It also addresses the issue of input representations and their relevance in phonological explanation.

There are two basic claims of this thesis. First, input segmental representations for the same segment may differ from language to language, and therefore there is no universal set of input forms. The input representations vary because they are constructed in response to contrasts in the inventory (Steriade (1987b), Avery & Rice (1989), Rice and Avery (1993), Goad (1993), Rose (1993), Dyck (1995), Causley (1997b, 1998a)). Since segmental representations are influenced by contrast, variation in inventories entails variation in segmental representations across languages.

Variation in the representation of a particular segment results in that segment patterning differently in different languages. Thus, input representations are an important determiner of segmental behaviour. If the inputs can be shown to play such a role in the phonology, it follows that the locus of phonological explanation is not found in constraint ranking alone, but must also reside in the input representations.

The second claim involves the issue of segmental markedness. I argue, following work by Avery and Rice (1989), Rice and Avery (1993), Rice (1993, 1995) that segmental markedness should be evaluated in terms of representational complexity. Under this view of markedness, the representations of more marked elements have more structure than those of unmarked elements. Marked elements will behave differently in the phonology because they have more structure. The additional structure means
that marked elements will be dispreferred by structure minimization constraints. This same structure entails that marked elements will have more structure to spread, and will represent graver violations of Faithfulness when deleted.

Together, these two claims form my proposal for the treatment of segments in Optimality Theory. The input representations form a large part of the segmental phonology, but they themselves are built in response to the competing considerations of structure minimization and the need to make contrasting specifications representationally distinct. The segmental representations that result determine the markedness relations between segment types and consequently drive the markedness-linked behaviour of different specifications.

1.1 The relevance of Inputs

A claim that is frequently made in the Optimality Theory literature is that languages differ in terms of constraint ranking alone; well-formedness constraints are universal and all possible input forms are available to all languages. Thus, Prince and Smolensky (1993) formulate the principle of Richness of the Base, rejecting the possibility of language-particular input representations.

Richness of the Base (P&S 1993:191)

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.

This thesis takes issue with this position, and argues for language-particular input representations for segments. Segmental representations are built in response to considerations of contrast and complexity. Therefore,
languages may differ in their representation of a particular sound depending on the type of inventory the sound occurs in.

The set of possible representations for a particular segment type differs in terms of degree of structure: one possible representation may have additional structure that is not included in an alternative representation. The presence vs. absence of this structure in the input has consequences for the patterning of a segment type in the output phonology. These consequences provide empirical evidence of the non-universality of inputs.

One generalization that comes up several times in this thesis is that, under circumstances where there is a potential for feature loss, unmarked features are often not preserved by Faithfulness, while marked elements are (cf. Kiparsky 1985). Rather than assume an inexplicable bias on the part of Faithfulness constraints, I argue that features whose output presence is not favoured by Faithfulness are to be treated just like segments whose output status goes unnoticed by Faithfulness: they are simply not part of the input. In this way, the behaviour of different segment types provides cues as to the input representation for different specifications.

In Chapters 5 and 6 I demonstrate through in-depth discussion of place and height coalescence in vowel systems that languages may differ to some degree in terms of which feature types are absent from the input. I show that these differences are to a large extent determined by the number and type of constraints in a system. Thus, the presence vs. absence of input features is linked to the shape of the segmental inventory.

Under a view where inputs are universal, the different behaviour of segment types in different languages cannot be explained in terms of different input representations. Instead groups of constraints which are ostensibly independently ranked must be arranged in a particular ranking in order for
these behaviours to fall out. This leads to an inexplicably constrained set of possible ranking configurations. Further, the different behaviour of segment types cannot be tied to inventory type under this approach, since the constraints determining inventories are independently ranked from those determining behaviour. However, such relationships do arise, and therefore additional constraints on ranking configurations emerge.

1.2 Markedness as representational complexity

In most current Optimality Theoretic research, segmental markedness is conceived of as an inherent property of feature specifications, and the relative markedness of particular segment types is governed by the relative ranking of constraints prohibiting those features. Thus, when the constraint banning feature [a] (i.e. *a) is ranked above the constraint banning the feature β (i.e. *β), the feature [a] is more marked than [β]. In Chapter 2 I argue that this view of segmental markedness results in a hybrid theory, with markedness being evaluated in terms of structural complexity at segmental and supra-segmental levels while at the sub-segmental level, relative markedness is defined in terms of constraint ranking.

A purely structural view of markedness allows for a unified theory, where markedness is consistently viewed in terms of a preference for simpler representations over more complex ones. Further, in Chapter 4 I show that evaluating markedness in terms of markedness constraint ranking allows for either too much or too little variability in the relative markedness of segment types cross-linguistically. This view also fails in capturing generalizations regarding the attested variations in markedness relations and the types of segmental inventories that are associated with those variations.
1.3 Linking markedness to inputs

The proposals for a structural definition of markedness and the focal role of input representations are intimately connected. As noted in §1.1, the phonological activity of a particular segment type is closely connected to its markedness status. In this thesis, phonological activity is viewed as the consequence of Faithfulness constraints driving the maintenance of feature specifications in the output. The link between markedness and the perseverance of input specifications in the output falls out from the representations. The degree of representational complexity in marked specifications makes them less preferred by constraints on structure minimization. At the same time, Faithfulness considerations mean that a marked specification, having more input structure to be faithful to, is more likely to be preserved in the output. In contrast, the relatively simple structure in a less marked specification will be preferred by structure minimization constraints, which means that these specifications will be selected in situations where simpler structure is preferred, e.g. in epenthesis, reduplication, and neutralization. However, this same lack of structure means that these specifications are often over-written in assimilation and coalescence in favour of more marked elements since loss of the more complex specification would involve Faithfulness violations.

The relationship between markedness behaviour of a segment type and the shape of the inventory also falls out from the representations ((Steriade (1987b), Avery and Rice (1989), Rice and Avery (1993), Goad (1993), Rose (1993), Dyck (1995), Causley (1997b, 1998a)). Since contrasts help to determine the input representations for different specifications, contrasts drive the type and degree of complexity involved in the representation of a given segment type. This contrast-motivated complexity involves violations of structure
minimization constraints, and therefore the representation will be marked. Hence the link between inventory shape and the markedness behaviour of different segment types.

1.4 Structure of the thesis

In Chapter 2 I lay out the logical foundations for pursuing a markedness-as-complexity approach within the OT framework, particularly given the current means of evaluating markedness at supra-segmental levels of the phonology. In Chapter 3 I discuss in detail the proposal for determining input representations and constraining the variability in segmental representations cross-linguistically. I describe how the current proposal provides an elegant account of the link between segmental markedness and phonological behaviour, as well as the link between segmental inventory and markedness. Chapter 4 addresses the feasibility of proposals attributing all cross-linguistic variability to constraint ranking. I argue that such an approach predicts more variability in markedness behaviour than what is cross-linguistically attested. Further, it fails to capture generalizations linking inventory to the markedness status of particular segment types.

Chapter 5 and 6 explore the issue of coalescence and inventory. Coalescence provides an excellent empirical arena for viewing the effects of Faithfulness, since the featural make-up of the output segment is entirely determined by Faithfulness and, I argue, the representations of the input segments. I explore two types of vowel coalescence in a variety of languages with different inventory types. Chapter 5 involves a detailed account of place coalescence in vowels, and Chapter 6 deals with height coalescence. These chapters provide additional empirical support for the proposal relating
markedness to segmental representations which are determined by structure minimization and contrast.
Chapter 2: Markedness, Faithfulness and STRUCTURE

Within the OT framework, grammars are conceived of as a set of ranked universal constraints on output forms. Constraints are violable; therefore although universal, they are not always surface-true. Violation of a particular constraint is restricted to situations where the violation is the only way of satisfying another constraint that is more highly-ranked. Thus, any output constraint violation must render a form more "optimal" in terms of a constraint that dominates the violated constraint.

Constraints fit into one of two categories: Faithfulness constraints and Structural/Markedness constraints (Prince and Smolensky (1993), Smolensky (1993)). Faithfulness constraints govern the identity relationship between different representations (e.g. Input-Output, Reduplicant-Base). In a general sense, Faithfulness requires that two representations be identical. Specific Faithfulness constraints penalize particular deviations such as non-identity in the linear order of elements or feature specifications which are non-identical.

Structural constraints militate against any dispreferred or marked structures. For example, the universal preference for syllables to have onsets is encoded in the structural/markedness constraint ONSET, penalizing marked onsetless syllables. A syllable with an onset satisfies ONSET, while a syllable without an onset violates ONSET. In this way, a marked structure is distinguished from an unmarked structure in terms of constraint violations.

To capture the generalization that, all else being equal, the simpler a form is, the less marked it is, there are also structural/markedness constraints which literally penalize structure, and therefore work to minimize complexity in outputs. Thus, with respect to these constraints a structurally
simpler form (e.g. with fewer segments) is preferred over a more complex form (e.g. with more segments).

Phonological alternation results from the conflicting requirements of Faithfulness constraints that require identity between input and output forms and Structural/markedness constraints that require that an output form be optimally unmarked. The interaction of these two constraint sets makes up the phonology of a language.

In this chapter, I explore two formally parallel types of Faithfulness constraints, and demonstrate that only one of these constraint types is required to govern identity between input and output representations. The other constraint type, I argue, is better formulated as a type of structural constraint against the complexity of representations. This anti-complexity constraint type forms a family of markedness constraints which plays a crucial role in the theory of markedness proposed in this thesis. These constraints evaluate markedness at different levels of the phonology, and guide the building of representations in the input.

Along with this reformulation comes two proposals: one for a unidirectional mapping under correspondence (Faithfulness) and one for a revised view of segmental markedness within OT. I argue that these revisions result in a more unified, simpler theory that accounts for the same phenomena as the standard theory (c.f. Prince and Smolensky (1993), McCarthy and Prince (1993, 1995)).

This chapter is structured as follows. First, I examine the formally parallel Faithfulness constraint sets MAX and DEP, constraints against deletion and epenthesis. In comparing the effects of these constraints at different levels of representation, I demonstrate that although they appear to be similar in spirit, they are distinct in their domains of relevance: MAX is
irrelevant at prosodic levels of representation while DEP is irrelevant at the subsegmental level. The reason MAX is irrelevant at the prosodic level is because there is generally no input structure to be Faithful to at these levels. The reason DEP is irrelevant at the subsegmental level is because all of its requirements are already enforced by segmental markedness constraints. Then, I suggest a reformulation of DEP as a structural/markedness constraint. Since this creates an overlap with the existing markedness constraint hierarchy currently assumed in most OT research, I suggest abandoning the substantive constraint hierarchy in favour of the unified general constraint on structural complexity. The proposed division of labour between constraint types means that Faithfulness mapping constraints (i.e. the MAX family) work only to preserve input structure. The anti-complexity constraint curbs the proliferation of structure. Finally, I examine the ramifications of the proposed view of markedness within the OT framework, and outline briefly how this proposal deals with markedness related generalizations.

2.1 MAX-DEP Asymmetries and *STRUCTURE constraints

2.1.1 Segmental Correspondence

Correspondence Theory (McCarthy and Prince (1995)) presents a view of Faithfulness in which elements contained in correspondent strings are governed by a set of correspondence constraints requiring such things as identity between correspondent pairs and a one-to-one mapping of the elements in two strings (for our purposes, Input and Output). The mapping requirement is enforced by two constraints, MAX and DEP, given in (2) and (3) (taken from M&P 1995: 264).

---

1Together, these constraints help to establish the exact nature of input representations (c.f. §3.2.2).
(1) **I/O Correspondence**

\[
\text{INPUT} \quad /p_1 \ a_2 \ u_3 \ k_4 \ t_5 \ a_6 / \\
\text{OUTPUT} \quad [p_1 \ a_2 \ u_3 \ k_4 \ t_5 \ a_6]
\]

(2) **MAX**

Every element of the Input has a correspondent in the Output.

(3) **DEP**

Every element of the Output has a correspondent in the Input.

The MAX constraint requires every input element have some output element coindexed with it in the output.\(^2\) Thus, this constraint prohibits deletion of input elements. The formally parallel DEP constraint is the "no insertion" counterpart to MAX: it requires that all output elements be coindexed with some input element. The effect of these constraints can be seen in the following double tableau illustrating the evaluation of a faithful and non-faithful candidate for each constraint.

(4)

<table>
<thead>
<tr>
<th>Input /k_1a_2t_3/</th>
<th>MAX</th>
<th>Input /k_1a_2t_3/</th>
<th>DEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. [k_1a_2]</td>
<td>* (&lt;t_3&gt;)</td>
<td>1. [k_1a_2t_3a]</td>
<td>* (a)</td>
</tr>
<tr>
<td>2. [k_1a_2t_3]</td>
<td></td>
<td>2. [k_1a_2t_3]</td>
<td></td>
</tr>
</tbody>
</table>

The first candidate on the left lacks a correspondent for the third input segment, incurring a violation of MAX. This violation is indicated with an asterisk in the MAX column. The second candidate has correspondents for all three input segments and therefore satisfies MAX. The first candidate on the right contains a segment (also shown in parentheses in the DEP column) which is not coindexed with any input segment; this represents a DEP

---

\(^2\)For the purposes of this thesis, I limit my discussion to relations between input and output representations and do not discuss reduplication.
violation. The second candidate contains no output segments which lack input correspondents and therefore satisfies DEP.

The constraints MAX and DEP seem to be logical co-conspirators in ensuring a one-to-one mapping of input and output elements at the segmental level. However, an examination of the effects of these two constraints at other levels of representations reveals them to be quite distinct: while MAX is well-motivated at the subsegmental or featural level to preserve input elements (e.g. Lombardi (1995), Causley (1996a), Causley and Smallwood (1997)), the effect of DEP violations is more difficult to detect at the featural level as they always coincide with markedness *Feature violations. Furthermore, above the level of the segment, the DEP constraint serves to constrain the amount of prosodic structure in a form, while a parallel MAX constraint remains largely ineffectual since prosodic structure is typically left out of the input representations.

This irrelevance of MAX at certain levels of representation points to a key concern of this thesis, which is addressed in detail in Chapter 3: MAX has no stake in any structure that is not part of the input. This means that all other structure in an optimal candidate must be motivated by some well-formedness constraint. Thus, the treatment of elements by MAX can help to define inputs: if MAX works to preserve an element, it must be part of the input. If MAX disregards some element, allowing it to appear or not at the whim of some other constraint, then this is evidence that this element is not part of the input representation.

2.1.2 Featural Correspondence

As an implementational assumption, M&P (1995) understand the correspondence relation to hold only between segments, thus only segments are governed by MAX and DEP. However, many researchers currently assume
a version of correspondence that is more general, holding at other levels of representation. For example, it is argued in Lombardi (1995), Causley (1996a), (1997a), Walker (1997), Causley and Smallwood (1997), Lombardi (1998) and assumed in Lamontagne and Rice (1995) that features must enter into a correspondence relation with one another. The analyses in these works involve the mapping constraint MAX(Feature) to ensure that input features have output correspondents.

Given the parallel nature of MAX and DEP, if we find evidence for a MAX(Feature) constraint, it seems reasonable to look for motivation for a DEP(Feature) constraint. Similar to the MAX and DEP constraints in (2) and (3), MAX(Feature) will prohibit the deletion of input features while DEP(Feature) will prohibit the insertion of features in output candidates. The effects of these constraints are depicted in the evaluation of schematic candidates in (5).

\[
\begin{array}{|c|c|}
\hline
\text{Input} /\text{CVC}/ & \text{MAX}(F) \\
\hline
\text{F}_1 & * (<F_2>) \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|}
\hline
\text{Input} /\text{CVC}/ & \text{DEP}(F) \\
\hline
\text{F}_1 & * (F_2) \\
\hline
\end{array}
\]

The first candidate on the left lacks an output correspondent for one of the input features, F_2, violating MAX(F). The first candidate on the right incurs a DEP(F) violation as it contains an output feature (F_2) which has no input correspondent.

Thus, just as DEP(Segment) rules out epenthetic segments, DEP(F) rules out "epenthetic" features. Interestingly, for every violation of DEP(F) by an
epenthetic feature there is also a violation of a non-Faithfulness markedness constraint. The featural markedness constraint set as proposed in Prince and Smolensky (1993) is a set of constraints that penalize specific feature specifications in the output. The general form of these constraints is given in (6).

(6) $^\ast F$

Segments should not bear F in the output.

Thus, the feature specific constraint $^\ast \alpha$, penalizes every instance of the feature $\alpha$ in a form, regardless of whether it has a correspondent in the input. Because markedness penalizes all output feature specifications, $DEP(F)$ violating features will be a subset of the features violating $^\ast F$. Thus in the tableau in (7), the first candidate bears a feature which violates $DEP(F)$ because it is has no input correspondent, and violates $^\ast F$ because all features are penalized by $^\ast F$. Note that the output feature $F_1$ only violates $^\ast F$ since it is part of the input and therefore not a violation in the eyes of $DEP(F)$.

(7)

<table>
<thead>
<tr>
<th>Input /CVC/</th>
<th>$DEP(F)$</th>
<th>$^\ast F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>1. CVC \ F_1 \ F_2</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>2. CVC \ F_1</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

With this subset relation arises a difficulty: how do we ascertain the effects of $DEP(F)$ as distinct from the markedness constraints? From the earliest research in OT, we know that markedness constraint violations are often compelled by Faithfulness: this is the explanation for why marked

---

3These constraints form sub-hierarchies within the $^\ast F$ constraint family. I discuss this hierarchical organization in §2.3.
structures are ever selected as optimal. However, DEP(F) will never itself compel a *F violation since there is no potential for these constraints to conflict given the inherent overlap in their requirements: *F says "don't have any F's in the output," and DEP(F) says "don't have any non-underlying F's in the output." Since these two constraints overlap to such a great extent, it seems questionable that they are co-existent in the grammar. I return to this point in section 1.2 and suggest the unification of these constraints.

Leaving aside the DEP(F)/*F overlap for the moment, evidence for correspondence at both the segmental and subsegmental levels of representation compels us to turn to the levels of representation above the segment to examine the effects of MAX and DEP. Parallel to the mapping requirements these constraint types make on roots and features, MAX will require underlying structure be preserved and DEP will require that any output structure be "sponsored" through correspondence with input structure. With respect to prosodic structure, however, we can expect to find some asymmetries in the effects of these constraints since prosodic structure is typically analysed as GEN-supplied structure and therefore non-underlying.4 This means that constraints requiring mapping of input structure to output structure (MAX-Prosody) will have little relevance to the evaluation of candidates: if there is no input structure to be preserved, then any candidate will satisfy this constraint. On the other hand, the only

4I leave aside for the present discussion languages with lexical accent systems. In these languages, metrical structure may be part of the input and therefore may be subject to Faithfulness constraints requiring identical structure in output candidates.

The issue of moraic structure is controversial. If they are to be considered input elements in their own right that seek output correspondents, then they could be treated in a manner parallel to segments and features. If they are not part of the input but instead are provided by GEN, then they can be treated as another level of prosodic structure. Another alternative is the Wt-Ident approach espoused in McCarthy (1996) that requires that the weight associated with a particular vowel in the input be identical to the weight associated with its output correspondent.
candidate that will not violate DEP-Prosody constraints is one without any (GEN-supplied) prosodic structure; obviously candidates without any prosodic structure whatsoever will be ruled out by other constraints requiring segments to be parsed into syllables, syllables into feet, etc.

Above the level of the segment, MAX-Prosody constraints will never be violated and are therefore not relevant in candidate selection, unless prosodic structure is lexically specified. DEP-Prosody constraints, on the other hand, will be violated by any candidate with prosodic structure. Thus, in addition to the subsegmental asymmetry, we have another asymmetry above the level of the segment in the nature of the formally parallel mapping constraint types MAX and DEP. This asymmetry is due to the nature of prosodic structure: since it is generally driven by well-formedness constraints,\(^5\) it need not be part of the input representation. Since MAX concerns itself only with output structure that has correspondents in the input, all well-formedness driven structure is ignored.

In the following section, I will demonstrate that at the levels of representation of syllable, foot, and phrase the DEP constraint is well motivated while the effect of its MAX counterpart is undetectable, except in cases where prosodic structure is lexical.

2.1.3 MAXsyllable and DEPsyllable

It is generally assumed that syllable structure need not be included in the input representation. Instead, the syllabification of the optimal output form will be determined by output constraints. If there is no input syllable structure, then a constraint such as MAXsyllable will be satisfied by all output candidates and will never serve to select one candidate over another. In contrast with MAXsyllable, DEPsyllable will be violated and will be able to

\(^5\)As noted above, this is excepting systems with unpredictable prosodic structure.
distinguish between better and worse candidates. Although all candidates will violate DEPsyllable to some degree, minimal violation will entail that the syllable structure in the optimal candidate be minimal. The following tableau demonstrates the effects of these constraints in evaluating a set of schematized output candidates. I use "." to indicate syllabification and, by extension, the presence of syllable structure.

(8)

<table>
<thead>
<tr>
<th>Input /CVCVCV/</th>
<th>MAXsyllable</th>
<th>DEPsyllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CV.CV.CV</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>2. CVC.CV</td>
<td>**</td>
<td></td>
</tr>
</tbody>
</table>

The first candidate is parsed into three syllables while the second contains two syllables with syncopation of the second vowel. Notice that MAXsyllable is indifferent to how many syllables a candidate has. DEPsyllable, however, is violated twice (in counting each syllabic constituent as a violation) by the syncopated candidate and three times by the non-syncopated candidate. Thus, a high-ranking DEPsyllable will serve to constrain the number of syllables in a form.

The effect of a syllable-minimizing constraint such as DEPsyllable can be seen in languages with syncope. Myers (1994) analyses Yawelmani syncope as a reflection of a high-ranking constraint against syllable structure. Short vowels are deleted between single consonants, provided that the consonants can be successfully syllabified. Thus, a constraint (or constraints) requiring well-formed syllables dominates the constraint against syllable structure, and this ranking determines the applicability of syncope in a form. The rule formulation of syncope is given in (9a) and is shown applying in (9b).

---

6Note, however, that the second candidate incurs a MAX(Segment) violation since one of the input vowels does not have an output correspondent.
(9) Yawelmani Syncope (Myers 1994 from Archangeli 1984)

a. \( V \rightarrow \emptyset / VC \_\_\_ CV \) [-long]

b. \( \text{p'um'na?} \) \( \text{"full-blooded one"} \)
   \( /\text{p'um'-in-a?} / \)
   full-mediopass.-noun

Thus, the Yawelmani form in (b) could be an example of schematic candidate #2 in the tableau in (8) which is selected for minimally violating DEPsylbable (despite a Maxsegment violation).

To summarize briefly, syllable structure is not relevant to Max because it is not part of the input. The DEPsylbable constraint, however, serves to reduce the number of syllables in the output.

2.1.4 Maxfoot and Depfoot

At the foot level, we see the constraints behaving in a similar fashion: Maxfoot has little to say about the number of feet a candidate has while Depfoot pushes for the minimum number of constituents. The following tableau demonstrates this.

---

7As noted above, it may be that lexical accent languages contain feet underlyingly and therefore Max(Foot) could influence candidate evaluation. The direction taken in McCarthy (1996) is to refer to the identity of correspondent elements, specifically their status as heads of metrical constituents: output correspondents of elements that are in a head position in the input must themselves be in a head position in the output (Head-Match).
Aside from deletion, one way to constrain the number of foot constituents in a form is to ensure that feet are maximal. In languages which require binary feet (presumably because a constraints guaranteeing binarity are highly ranked), the effect of such a constraint will not necessarily be entirely obvious. However, in systems characterized by unbounded feet, for example, the effects of DEPfoot are more apparent.

2.1.5 MAXφ and DEPφ

Again, at the level of the phonological phrase (φ), MAXφ would be completely vacuous since there are no input structures to map to the output. DEP(φ), however, works to achieve parsimony by constraining the number of constituents and the amount of structure.

Thus given an input with several phonological words, DEP will, all other things being equal, prefer an output candidate with the fewest phrasal constituents possible.

We see the effect of such a constraint implicit in a phonological phrasing algorithm such as that proposed in Ghini (1993) for Italian. The analysis in Ghini (1993) specifies where phrase boundaries must occur at
certain syntactic boundaries; within those boundaries, the number of phrasal constituents is constrained, and a candidate such as the second in (11) is dispreferred.

For example, in (12) a phrase boundary must coincide with a syntactic boundary between the two phonological words in the sentence. Notice that in (13), that this same syntactic boundary doesn't occur between the words of the sentence. When the phonological words in (13) are parsed into phonological phrases, notice that they do not form a single phrase each to give four phonological phrases but instead form only two phrases. This could be viewed as the effect of a constraint minimizing the number of phonological phrases in a form.

(12) La verità vince. The truth wins. (Ghini 1993: 71)
(la#verità)φ(vince)φ
the-truth wins

(13) Berró vino di botte vecchia. I will drink old barrel wine.
(berró)φ(vino)φ(di#botte)φ(vecchia)φ
I will drink wine of-barrel old

2.1.6 Summary

In this section, I have described asymmetries in the effects of MAX and DEP constraints at different levels of representation. It seems that MAX constraints are motivated in domains where material is present in the input, while DEP-type constraints are required where material is not present in the input to constrain the amount of structure in a form.

Although MAX-type constraints are required to maintain input elements at the subsegmental level, below the level of the segment it is difficult to see the effect of DEP-type Faithfulness distinct from independently motivated markedness constraints. Above the segmental level, where input
structure is not present, the effects of MAX-type constraints are not obvious. At some levels of prosodic structure, these types of constraints will never have the opportunity to be violated (and therefore can have no effect on candidate evaluation). Because of lack of input prosodic structure, DEP-type constraints penalize all structure, since there is no input structure to distinguish from the GEN-supplied structure. The non-isomorphy in domains of MAX and DEP effects is schematized in the diagram in (14).

(14)

\[
\text{MAX} \quad \phi \quad \text{Ft} \quad \sigma \quad \text{DEP} \\
\text{Rt=SEGMENT} \quad \text{Feature}
\]

This mis-match in the effects of the mapping constraints comes down to the nature of the input: root nodes and features are included as part of the input, and therefore are governed by MAX. Prosodic structure is not part of the input, and therefore is not governed by MAX. DEP constraints serve to restrict the amount of structure in an output. However, the occurrence of subsegmental elements is already restricted by markedness constraints. The next section argues that the current formulation of MAX is a useful constraint for ensuring that outputs reflect input forms, but the current formulation of DEP needs to be revised.
2.2 \textbf{STRUCTURE and Uni-Directional Faithfulness}

2.2.1 \textsc{DEP} and the Prohibition of Structure

In the previous section, I discussed the Faithfulness constraint DEP at different levels of representation and I argued that the effect of DEP can be seen from the level of the segment to the phonological phrase. At all levels of representation, the formalization of DEP remains the same, as in (3). At the segmental level, DEP penalizes any segment without an input correspondent. However, its effect will be somewhat different at prosodic levels given that most prosodic structure is constructed by GEN and therefore does not correspond with input elements. That is, DEP continues to penalize non-underlying structure at the prosodic level yet its effect is to penalize \textit{all} structure, since it is all given by GEN.

DEP-Prosody, therefore, is not serving necessarily to distinguish underlying from non-underlying structure. Instead, it simply constrains the amount of prosodic structure in an output form. In this respect, it is less like a Faithfulness constraint than it is a structural or markedness constraint, assigning a violation to any structure. The categorization of a constraint as a Faithfulness vs. a Structural/Markedness constraint is a reflection of its function in the system: Faithfulness constraints work to achieve identity between two representations, while Structural/Markedness constraints require output forms to be as unmarked as possible. As demonstrated in the previous section, DEP-Prosody constraints are perhaps better defined as constraints against markedness (i.e. in terms of complexity of structure) than constraints against the non-identity of forms, since identity is not at issue when output material has no input correspondents.

Given this non-Faithfulness characterization of the DEP constraints governing prosodic structure, can we extend this view to the segmental level,
removing DEP(Segment) from the Faithfulness constraint set and re-formulating it as a Structural/Markedness constraint?

This line of inquiry does well to make use of an idea taken up by Myers (1994) who describes a Faithfulness constraint *STRUCTURE as a constraint prohibiting structure at different levels of representation. McCarthy and Prince (1993) make the initial proposal for a general family of constraints against structure. Working within Containment-based OT, McCarthy and Prince suggests that the anti-epenthesis Faithfulness constraint FILL can be seen as a part of a more general constraint family:

FILL belongs to [a] class of constraints which militate against the presence of structure *STRUC, ensuring minimal structural development in response to any dominant Parse considerations. In a fully general account, this would include filled as well as empty or partly empty nodes, not to mention auto-segmental links, grid positions, and so on. The same constraint family is active in syntax and even semantics. For example, Chomsky's suggestion that X' nodes appear only when accompanied by a sister falls naturally under *STRUC (Chomsky 1986:4). (M&P 1993: fn 16, page 15)

In Myers' work this constraint subsumes the Standard Theory (P&S (1993)) FILL constraint and is extended to constrain the number of segments, syllables, and association lines in a form. I adopt this *STRUCTURE constraint as defined in Myers (1994), recategorizing it as a non-Faithfulness Structural/Markedness constraint which subsumes the DEP family of constraints. In doing so, I propose a uni-directional mapping of elements from Input to Output. This means that the MAX constraint still holds, requiring the preservation of input elements. However, there is no mapping from Output to Input and no Faithfulness constraint explicitly prohibiting epenthetic elements. Instead, all structure is penalized by *STRUCTURE which
may be relativized to separate constraints governing the number of metrical and syllabic constituents\(^8\), root nodes, and feature specifications in a form.

Epenthesis will still be penalized, but by a general constraint against structure. Since input elements are to be preserved to satisfy Max, the violations of *\text{STRUCTURE} incurred by them are motivated. Violations incurred by epenthetic elements must be motivated by some other structural constraint (such as Onset). Thus, as demonstrated in the tableau in (15) there still remains a distinction between input and GEN-supplied elements.

(15)

<table>
<thead>
<tr>
<th>Input /k₁a₂t₃/</th>
<th>Max</th>
<th>*\text{STRUCTRT}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. [k₁a₂]</td>
<td>*! (&lt;t₃&gt;)</td>
<td>**</td>
</tr>
<tr>
<td>2. [k₁a₂t₃]</td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>3. [k₁a₂t₃a]</td>
<td></td>
<td>*** *! (a)</td>
</tr>
</tbody>
</table>

In the above tableau, we see the interaction of Faithfulness (Max) and *\text{STRUCTURE} at the root level. The anti-complexity constraint *\text{STRUCTURE} would prefer that output candidates have the fewest number of root nodes possible. Of the candidates included here, then, the first candidate is the best as far as *\text{STRUCTURE} is concerned. However, this candidate lacks an output correspondent for one of the input root nodes, thereby incurring a violation of Max. Since Max is higher-ranked, this candidate is ruled out, and only the remaining two faithful candidates are possible outputs. The difference between these two last candidates is that the third one involves an extra segment which has no input correspondent. The decision between these candidates is made on the basis of *\text{STRUCTURE}: the second candidate is optimal since violation of *\text{STRUCTURE} is minimal. Thus, Faithfulness can

---

\(^8\)In Chapter 6 I propose that *\text{STRUCTURE} is also relativized to constraints governing different types of structure, namely branching vs. embedded dependency.
motivate violations of *STRUCTURE, but beyond being Faithful, *STRUCTURE will rule out any extraneous structure.

The interaction of these two basic constraint types will play a major role in later discussions of input representations. The appearance of an element in the face of well-formedness violations means that Faithfulness is the motivating constraint. The intervention of Faithfulness on the part of an element that is dispreferred in the output (i.e. all elements by *STRUCTURE) is evidence for its presence in the input. Thus, the behaviour of Faithfulness is a diagnostic for distinguishing input vs. well-formedness driven structure.

What does this proposal for eliminating DEP imply for subsegmental mapping? If *STRUCTURE holds at the level of the segment and above, does it also constrain subsegmental structure? This is the issue to which I now turn.

2.2.2 Markedness as non-Faithfulness *STRUCTURE constraint set

In Section 2.1, I touched briefly on the problem of overlap between a DEP(Feature) constraint and the markedness constraints barring feature specifications. I said that the markedness constraints would penalize every candidate that the DEP constraint would penalize. Markedness constraints, in penalizing every feature specification, also penalize the non-underlying features that would violate DEP(F). In this respect, the markedness constraints are very similar to the non-Faithfulness *STRUCTURE constraint: all structure is a violation. Like segmental and prosodic structure, features from the input are distinguished from those supplied by GEN in that the features with input correspondents are governed by the Faithfulness constraint MAX.

My proposal in this section is to follow up on the similarities between *STRUCTURE and markedness constraints and suggest a reformulation of the markedness constraint set as the more general *STRUCTURE constraint. My proposal makes use of representations (from Rice and Avery (1993), Rice
(1993, 1994, 1996)) which directly encode markedness as structural complexity. The more structure a representation has, the more marked it is in the sense that it violates *STRUCTURE more seriously than an unmarked, less complex representation.

2.2.3 Contrasting Views of Markedness

There is a long-standing assumption in the segmental phonology literature, beginning with observations of Prague School linguists Trubetzkoy (1939) and Jakobson (1941), that feature values are not all of equal status. Instead, intrinsic markedness relations exist between different features: some features are less marked than others. This marked-unmarked distinction is reflected in various asymmetries regarding the distribution and patterning of different feature types.

In Optimality Theory, these generalizations are captured through a markedness constraint hierarchy (Prince and Smolensky (1993)). The featural markedness constraints are ranked in a fixed hierarchy, where constraints against marked features dominate constraints against unmarked features. The marked vs. unmarked status of features comes from the independently recognized markedness relations, where certain features such as coronal are treated as unmarked relative to non-coronal specifications (see Paradis and Prunet (1991) and references therein). The unmarked status of coronals means that the constraint banning coronals is lower ranked than the constraint banning other place types: *dors, *lab>>*cor.  

The *STRUCTURE constraint set resembles the markedness hierarchy in that they both require unmarked forms: structurally unmarked forms are the ones which are least complex in terms of structure, and therefore incur the

---

9Lombardi (1995, 1997) proposes an extended hierarchy, including *pharyngeal at the right end. I discuss this proposal in detail in Chapter 4.
least number of violations of structural constraints. The markedness constraint hierarchy also compels the least marked form, although not because a marked form incurs more violations of a single constraint. Instead, a more marked feature specification incurs a violation of a higher-ranked markedness constraint.

The proposal put forth here is to adopt structural representations for these feature specifications that directly encode the markedness relations captured in the fixed markedness constraint hierarchy. With these representations, the degree of markedness is reflected in the structural complexity of a specification. The representations for Place specifications, described in detail in the Chapter 3, are given in (16).10 While the markedness relations between different manner specifications have not received much attention in the Correspondence Theory literature,11 Rice and Avery (1993) and Rice (1992), provide evidence for situating more sonorous elements at the marked end of the scale, while obstruent elements are identified as less marked. Thus the representations in (17) depict obstruent segments as structurally less complex than nasals, which are in turn less complex than lateral segments (the question of rhotics is ignored here).

(16) Representation of Place:

<table>
<thead>
<tr>
<th></th>
<th>coronal</th>
<th></th>
<th>labial</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>laryngeal</td>
<td></td>
<td>labial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rt</td>
<td></td>
<td>Place</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Place</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peripheral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Labial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>coronal</td>
<td></td>
<td>dorsal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rt</td>
<td></td>
<td>Place</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peripheral</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dorsal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

10 In Chapter 3 I describe the variability that is permitted in these representations.
11 Prince and Smolensky (1993) do discuss a scale of preferences for segments of different sonority to associate with different syllable positions.
Coupled with these representations (or any set of representations encoding markedness as complexity), the *STRUCTURE constraint will replace the markedness constraint hierarchy. The *STRUCTURE constraints, as well as constraining structure at other levels of representation, will serve to limit the complexity of output segmental representations. Like the interaction of Faithfulness with the markedness constraint hierarchy, violations of *STRUCTURE will be compelled by the presence of input structure. However, where Faithfulness does not play a role in the evaluation of output form, *STRUCTURE will still ensure the emergence of the unmarked.

The interaction between these two constraint types (Faithfulness and *STRUCTURE) and the appropriate representations allow for an elegant theory of markedness and segmental patterning, described in brief below, and in detail in subsequent chapters.

2.3 On the role of Markedness

The markedness constraints as presented in Prince and Smolensky (1993) play a role in the emergence of several phenomena. The fixed hierarchy of constraints (*dors, *lab >> *cor) means that coronals will be the least marked segment type, and should pattern uniquely in several ways. In this section I review some of the generalizations that have been linked to markedness and have received an account under the markedness hierarchy

---

view. Then, I describe how a representational view of markedness copes with the same patterns. This discussion is only a precursor to the discussion of segmental markedness behaviour that follows in subsequent chapters, and we will return to these issues later to provide a more in-depth exploration than is given in the overview here.

2.3.1 Choice of Epenthetic Segments

Cross-linguistically, the most unmarked feature specifications are preferred in epenthetic segments. This pattern may be explained in terms of the markedness constraint set: the insertion of a less marked feature violates a lower-ranked markedness constraint. As is the case whenever Faithfulness is not at issue, the unmarked is preferred. Epenthesis is thus one of the places where we see the much discussed "emergence of the unmarked" (McCarthy (1995)).

A representational view of markedness also accounts for the choice of unmarked elements for epenthesis. *STRUCTURE is to be violated minimally whenever possible. Unfettered by Faithfulness considerations, *STRUCTURE will ensure the least complex segments (i.e. those bearing the least complex featural specifications) are selected as optimal.

2.3.2 Implicational relationships in inventories/positions

In the phonemic inventories of the world's languages, there are several implicational relationships that may be drawn between certain feature specifications: if an inventory contains the marked specification for a particular feature, then the unmarked value of that feature is also generally part of the inventory. For example, the presence of a high front nasal vowel

---

13Another context where we expect to see the emergence of the unmarked (often abbreviated as TETU) is in reduplication (McCarthy and Prince (1993, 1995), McCarthy (1995)).

14There are some counterexamples to this tendency. For example, Hawaiian lacks coronal stops in its inventory (Maddieson (1984), Rice and Avery (1993), Rice (1996)).
in a given inventory implies the presence of its less marked oral counterpart. Similarly, if a language has labial place of articulation in stops, it also has the less marked coronal place for stops. These same implicational relationships hold largely within a particular position or domain: if a particular position allows the marked member(s) of a feature class, it also allows the less marked member(s).\textsuperscript{15}

These facts fall out of the fixed hierarchy of markedness constraints since a ranking that motivates the violation of a higher-ranked markedness constraint also allows the violation of a lower-ranked markedness constraint. That is, if an output form bears a marked feature, we know that Faithfulness must be ranked above the markedness constraint prohibiting that feature. If Faithfulness, is ranked above the constraint $^F_{\text{marked}}$, it is necessarily ranked above $^F_{\text{unmarked}}$, since the ranking between the markedness constraints is fixed as $^F_{\text{marked}} \gg^F_{\text{unmarked}}$. Since Faithfulness therefore dominates $^F_{\text{unmarked}}$, this unmarked feature is permitted in the output.

In a representational view of markedness, the implicational relationships fall out of representations and the effect of the $^\text{STRUCTURE}$ constraint. Because the representations are built up from the contrasts in the inventory,\textsuperscript{16} faithfulness to more complex structure means that less complex structures contained in the input will also be protected by Faithfulness. That is, once a certain amount of structure (and $^\text{STRUCTURE}$ violations) is permitted, any less structure is also permitted.

2.3.3 Shape of inventories

\textsuperscript{15}Note that the issue of positional contrast is significantly more complicated than I imply here. Markedness can certainly be relativized to a particular position so that a feature which is marked in one position may be the unmarked specification in another position (e.g. see Hamilton (1996) on this issue).

\textsuperscript{16}Chapter 3 explores the role of contrast and $^\text{STRUCTURE}$ in the building of representations. Informally put, Lexicon Optimization and $^\text{STRUCTURE}$ will ensure that the building of representations is done in a step-wise fashion, building only one level of structure at a time.
Inventories are inclined to have a "triangular" shape (see, e.g. Trubetzkoy 1969). As for consonants, the number of place contrasts available to a particular manner class tends to be inversely proportional to the sonority of that manner class (see, e.g. Rice and Avery (1993)). Roughly speaking, if there is a difference in the number of place contrasts available to obstruents and sonorants, the least sonorous manner of articulation (stops) typically have the greatest number of place contrasts, while the more sonorous have the least place contrasts. In vowels, a similar pattern obtains: the more sonorous low vowels tend to admit fewer distinctions (in terms of backness, roundness, etc.) than the less sonorous high vowels.

While these facts have not received much direct attention in the OT literature, an account is readily available through the markedness constraint sets. We could imagine that the interaction of markedness scales determines the "markedness ceiling" for segments given a particular ranking (McCarthy 1994; for a non-OT account along similar lines see Hamilton (1996)).

A representational view of markedness also succeeds in predicting inventory shape. With representations (such as in (17)) where more sonorous segments involve more structure, we can understand why fewer place contrasts are found with more sonorous segment types. The notion of local constraint conjunction (Smolensky (1993, 1997), Ito and Mester (1996), Alderete (1996)) comes into play in establishing a complexity ceiling for an individual segment type. Local conjunction, simply put says that local multiple violations of a constraint are worse than non-local violations. We return to the issue of constraint conjunction and the determination of inventories in Chapter 6.

Since sonorants already contain more structure than obstruents, the addition of more structure in terms of place specification will create extremely
complex segments, and multiple local violations of *STRUCTURE. The obstruents, being the least complex in terms of manner, incur fewer violations of *STRUCTURE and so are the targets for elaboration of place distinctions. Local Conjunction of *STRUCTURE inhibits the elaboration of already-complex segment types by penalizing excessive structure locally, i.e. on a single segment.

2.3.4 Assimilation and markedness

Assimilation patterns in many languages treat unmarked and marked elements differently: marked elements are often triggers but not targets of assimilation while unmarked elements are often targets but not triggers of assimilation (Kiparsky (1985)). Thus, in place assimilation patterns, we often see a [coronal][labial] sequence becoming [labial][labial] while a [labial][coronal] sequence remaining [labial][coronal]. Why are unmarked elements susceptible to assimilation while marked elements resist assimilation?17

As discussed in Chapter 4, the markedness constraint hierarchy does not provide a satisfactory answer to this question. In fact, the patterns are completely unexpected given that the assimilation patterns work to preserve more marked feature at the cost of losing a less marked feature, creating less harmonic forms. Faithfulness may appear to be a direction to follow in search of a solution. However, such a solution requires the stipulation of a "reverse hierarchy" where Faithfulness is more sensitive to marked elements (i.e. \( \text{MAX}_{\text{marked}} \gg \text{MAX}_{\text{unmarked}} \)).

The asymmetries in assimilation patterns are completely expected, however, under a representational view of markedness. Assimilation of a more marked segment to a less-marked one entails loss of input structure

---

17This question is raised in Avery and Rice (1989), Smolensky (1993), Kiparsky (1994), and Alderete et al. (1997) amongst others, and is the subject of Causley (1996b), who suggests a representational OT account of the facts.
(MAX violations) while assimilation of less-marked to more marked involves no MAX violations.

Since labials have more input structure than coronals, assimilation of a labial to a coronal is worse than assimilation of a coronal to a labial because there is more structure to map (be faithful to) in labials. The assimilation of a coronal segment does not involve a Max violation because a specification for coronal does not exist in the input. Thus, coronal place of articulation will assimilate to a Dorsal or Labial consonant, even though the reverse pattern produces a more harmonic, less marked form. These assimilation patterns simply represent cases where Faithfulness works to keep the maximum input structure in the output, despite the violations of *STRUCTURE constraints. Thus, the link between markedness and the behaviour of a segment type in assimilation is simply a result of the requirements of Faithfulness being met. This same point will surface several more times in this thesis in the analysis of various other phonological processes.

2.4 Conclusion

In this chapter, I have proposed a purely representational view of markedness, relating degree of markedness to structural complexity. In addition, I have argued for a unidirectional mapping under correspondence, eliminating the correspondence constraint DEP. While DEP formerly constrained epenthetic elements, the non-Faithfulness constraint *STRUCTURE now constrains structure at all levels of representation. In requiring minimal complexity in output forms, *STRUCTURE forms a general markedness constraint obviating the need for the feature-specific markedness constraint hierarchy. A *STRUCTURE view of markedness means the emergence of the unmarked is the emergence of the simplest representation.
The presence vs. absence of structure in the input is a crucial
determiner of segmental behaviour in the output: Faithfulness can only
intervene on the part of input structure. The intervention of Faithfulness
means preservation in the output, which is realized as resistance to
assimilation and, as we will see in the following chapter, other processes
resulting in feature loss.
Chapter 3: A representational theory of features

In the previous chapter, I outlined a proposal for the treatment of markedness as a structurally defined notion. In terms of segment structure, this means that the markedness status of specifications is encoded in the representations. In this chapter I illustrate the advantages of a representational view of segmental markedness in accounting for links between the phonological patterning of segments and their markedness status.

An important generalization that receives an explanation under this view is the relationship between the shape of a particular segmental inventory and the phonological behaviour of different segment types in that inventory. I argue that contrast plays a role in determining representations (Steriade (1987b), Avery and Rice (1989), Rice and Avery (1993), Goad (1993), Dyck (1995)) and therefore has important consequences for both the markedness and the phonological behaviour of a particular segment type. Further, I argue that such a view of segmental markedness also provides an explanation for the limited degree of variability that is seen in the markedness relations between some segment types; see, for example, Dyck (1995), Rice (1996), Causley (1998a).

Finally, I demonstrate how such a model is complemented by a dispersion-type model of phonetic implementation (Lindblom (1986), (1990), Flemming (1995)). I show that while my proposal is compatible with a dispersion-type phonetic implementation component, the representational component cannot be replaced by dispersion theory, contra Flemming (1995) and Ní Chiosáin and Padgett (1997).
3.1 The representations

As seen in the previous chapter, markedness can be evaluated in terms of structural complexity: more complex representations incur more violations of \(^*\text{STRUCTURE}\) and therefore may pattern differently from less complex elements. Under this view, segmental representations need to reflect the relative markedness of different segment types in terms of relative complexity. For example, if we have two segment types \(a\) and \(b\), and \(a\) is more marked than \(b\),\(^1\) \(a\) should have a representation that is more complex than \(b\),\(^2\) as shown schematically in (1).

\[
\begin{array}{c}
\text{a=marked} \\
\text{b=unmarked} \\
\text{a} \\
\text{b} \\
| \\
| \\
x \\
y \\
x
\end{array}
\]

The more complex segment will incur more violations of \(^*\text{STRUCTURE}\), and therefore, all else being equal, the less complex segment will be preferred. As mentioned in the previous chapter, this means that the unmarked segment will be chosen over the marked one for epenthesis and other situations where Faithfulness is not at issue. It also means that the unmarked segment type will be implied in an inventory that allows the marked segment type. The relationship between markedness and phonological patterning is discussed in more detail in subsequent sections.

Given this view of markedness as structurally defined, what is required of any theory of segment structure is that markedness relations between specifications and segment types are encoded in the relative complexity of the

\(^1\)In §3.4 I discuss how markedness may be evaluated for a particular segment type.

\(^2\)Although, as we will see in §3.4, a segment can be marked by virtue of being completely lacking in structure.
representations. For the purposes of this thesis, I adopt the following representations (from Avery and Rice (1989), Rice (1994, 1996)) and feature system for consonantal place, although other feature sets and/or configurations could be consistent with the current proposal.

(2) Representation of Place:

\[
\begin{array}{cccc}
\text{laryngeal} & \text{coronal} & \text{labial} & \text{dorsal} \\
\text{Rt} & \text{Rt} & \text{Rt} & \text{Rt} \\
\mid & \mid & \mid & \mid \\
\text{C-Place} & \text{C-Place} & \text{C-Place} & \text{C-Place} \\
\mid & \mid & \mid & \mid \\
\text{Peripheral} & \text{Peripheral} & \text{Peripheral} & \text{Peripheral} \\
\mid & \mid & \mid & \mid \\
(Labial) & (Dorsal) & (Labial) & (Dorsal)
\end{array}
\]

These representations reflect the relative markedness of the different place types: coronal is less marked than labial and dorsal specifications (see, e.g. references in Paradis and Prunet (1991)) and correspondingly has less structure than labial and dorsal. Laryngeals also often appear to be unmarked relative to other segment types, frequently appearing in environments where we see the emergence of the unmarked. This too is encoded in its representation: it lacks any C-Place structure and therefore incurs no violations of *STRUCTURE for its place specification.

The fact that laryngeal place structure is less complex than all other C-places raises the question of why coronal should ever pattern as the least marked segment in a language. This question is explored in §3.4, which deals with variability in markedness relations across languages. In addition to variable markedness relations between coronal and laryngeal segment types, there is some evidence that labial and dorsal segment types may enter into a variable markedness relation with each other (see, e.g. Rice (1994)). In some languages labial segments may be considered unmarked relative to dorsal
segments, while in other languages the reverse appears to be true. Such a relationship is represented through the lack of a Peripheral dependent on the less marked segment type. Since either labial or dorsal segments may pattern this way, it appears to be variable which segment type lacks a dependent. The exact representations are decided on on a language-particular basis. This type of variability is also discussed in subsequent sections.

The vowel place specifications are parallel to the C-Place representations above, although slightly more complicated in the range of variability permitted. First of all, in a system with a front-central-back place contrast, the representations are as in (3).\textsuperscript{3} Since central vowels pattern as the least marked in these systems, they bear the least amount of place structure. In systems without central vowels, however, the front and back vowel specifications are subject to variability. As discussed in greater depth in §3.4, either the front or the back vowel will behave as the least marked vowel. This variability in markedness relations suggests a variability in relative complexity of representations, therefore either the front or the back vowel in front-back systems may lack a V-Place dependent, rendering them identical to the central vowel in a three-place system. Again, as with the labial-dorsal variability mentioned above, which vowel will be unmarked appears to be decided on a language-particular basis.

\textsuperscript{3}I have left out the issue of systems with front rounded vowels, which always pattern as marked relative to other vowel types. Because of its universally marked status, front rounded vowels under this proposal would have a complex representation, bearing both coronal and peripheral specifications.
(3) Representation of V-Place:

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rt</td>
<td></td>
<td>Rt</td>
<td>Rt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPlace</td>
<td>VPlace</td>
<td>VPlace</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Coronal) (Peripheral)

In the sections that follow, I discuss various issues surrounding these proposed representations. I begin with how these representations allow a unified account of the related generalizations regarding the markedness of a given specification and the patterning of that specification in spreading and coalescence phenomena. I follow this with a discussion of some of the arguments down-playing the role of underspecified representations in Optimality Theory in §3.3. The assumption of selectively specified representations are crucial to virtually every point in this thesis. Therefore, the whole work should be taken as an argument for this view of segment structure.

3.2 Representational Complexity and Phonological Activity

As discussed in the previous chapter, segmental markedness is reflected in the phonology of a language in several ways. Inventory shape, patterns of neutralization, and choice of epenthetic segments have all been linked to universal preferences for particular segment types.

In this section I describe how the proposed view of markedness together with representations such as those proposed above allows a straightforward account of markedness-linked asymmetries in the phonological activity of different features.
Phonologically "active" features are those features that spread from one segment to another in assimilation and harmony processes, and those features that are maintained under coalescence of two segments.4 Phonologically "inactive" or inert features are those features which fail to spread and are often lost in coalescence. Phonological activity is not random. Instead, the activity of a feature is related to its markedness status: marked features tend to be "active" while unmarked features are "inactive" (Kiparsky 1985). In this section, I present data from assimilation and coalescence demonstrating the activity of marked places of articulation in consonants (i.e. labial, dorsal) and the relative inactivity of unmarked places of articulation (i.e. coronal). Next, I demonstrate how, under my proposal, this link between markedness and phonological activity is completely expected.

3.2.1 Markedness and phonological activity

Cho (1991) describes a process of place assimilation in Korean in which dentals assimilate to labials, palatals and velars. Thus, the place feature of dentals is inactive in this process, while the place features of labials, palatals, and velars are active.

(4)

a. /pat + ko/ [pakko]5 to receive and
b. /kotpalo/ [kopparo] straight
c. /kæt+ci/ [kæcci] Let us uncover
d. /hankan/ [haŋkan] the Han river
e. /han + bán/ [hamban] once

Parallel to the Korean patterns, Avery and Rice (1989) discuss place assimilation processes in Catalan and English which demonstrate the inactivity of the feature coronal, and the relative active behaviour of non-

---

4Chapters 5 and 6 focus on coalescence as a type of phonological activity.
5This data is also presented in Avery and Rice (1989).
coronal place features. In the Catalan data discussed (taken from Kiparsky (1985)), coronal nasals, laterals, and stops assimilate to the place of the following consonant. Labial and dorsal nasals and stops fail to undergo such assimilation in the same environment. Kiparsky (1985) and Avery and Rice (1989) note that the non-coronal nasals do assimilate within their own place. For example, plain labials will assimilate to labiodentals as in (6c). In (5)-(8) I present the examples involving nasals (from pp.187-8).

(5) /n/
   a. unassimilated         so[n] amics     they are friends
   b. labial                so[m] pocs       they are few
   c. labiodental           so[r] feliços    they are happy
   d. dental                so[n] dos        they are two
   e. alveolar              so[n] sincers    they are sincere
   f. postalveolar          so[n] rics       they are rich
   g. laminopalatal         so[n, 3]ermans   they are brothers
   h. palatal               so[n]ljures      they are free
   i. velar                 so[ŋ] grans      they are big

(6) /m/
   a. so[m] amics           we are friends
   b. so[m] pocs            we are few
   c. so[r] feliços         we are happy
   d. so[m] dos             we are two

(7) /ŋ/
   a. ti[ŋ] pa              I have bread

(8) /ñ/
   a. a[ŋ] felic            happy year
In (5), the final nasal of son surfaces as a coronal when preceding a vowel-initial word (a) or when preceding a coronal-initial word (d-f). When a labial-initial or dorsal-initial word follows, however, the final nasal shows up as a labial or dorsal respectively. In (6)-(8), we see that a final nasal which is labial, dorsal, or palatal does not assimilate to the place of a following consonant.

Avery and Rice (1989) also discuss a similar rule of nasal place assimilation in English. A nasal may optionally assimilate to the place of the initial consonant of a following word. They discuss examples involving the preposition in which I repeat in (9).

(9) labial  \textit{i[m]} Brussels
labiodental \textit{i[n]} France
dental \textit{i[n]} there
alveolar \textit{i[n]} Toronto
velar \textit{i[n]} Kingston

(10) labial  \textit{fro[m]} Belgium
labiodental \textit{fro[n]} France
alveolar \textit{fro[m]} Toronto
velar \textit{fro[m]} Kingston

As in the Catalan and Korean cases, only coronal nasals undergo this assimilation; labials and dorsals do not assimilate to the place of a following coronal or dorsal segment. Thus, only the feature coronal is inactive in this process, while place features of labial and velar nasals are active.

Another process which demonstrates the phonological activity of features is coalescence. In coalescence, active features are maintained in the output while inactive features are lost. In the Lower dialect of Ahtna, an

---

6Some analyses of coalescence treat what I call coalescence as two separate processes:
Athapaskan language spoken in Alaska, a coronal oral or nasal stop coalesces with a following continuant to give a stop which is homorganic with the continuant. The consonantal inventory in (11) and data in (12) and (13) come from Kari (1990).

(11) Ahtna Consonants

<table>
<thead>
<tr>
<th>Manner of Articulation</th>
<th>Labial</th>
<th>Alveolar</th>
<th>Lateral</th>
<th>Alveo-Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>p</td>
<td>t</td>
<td>t'</td>
<td>ts</td>
<td>k</td>
<td>q</td>
<td>?</td>
</tr>
<tr>
<td>plain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aspirated</td>
<td>(pʰ)</td>
<td>tʰ</td>
<td>tʰ'</td>
<td>tʰs</td>
<td>kʰ</td>
<td>qʰ</td>
<td></td>
</tr>
<tr>
<td>glottalized</td>
<td>t'</td>
<td>t'ʰ</td>
<td>t'ʰs</td>
<td>k'</td>
<td>q'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricatives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiced</td>
<td>(v)</td>
<td>l</td>
<td>z</td>
<td>r</td>
<td>g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiceless</td>
<td>hw</td>
<td>ñ</td>
<td>s</td>
<td>x</td>
<td>h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasals</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(12) Coalescence

<table>
<thead>
<tr>
<th>t + y → k</th>
<th>Expected form</th>
<th>Lower dialect</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>t + y → k</td>
<td>[tv]en</td>
<td>[k]en</td>
<td>he is acting as a shaman</td>
</tr>
<tr>
<td>t + r → q</td>
<td>a[r]an’</td>
<td>a[q]an’</td>
<td>it is moldy</td>
</tr>
<tr>
<td>th' + x → q’h</td>
<td>[th’x]atsisi’</td>
<td>[q’h]etsisi’</td>
<td>I will make it</td>
</tr>
<tr>
<td>[th’x]estna?</td>
<td>[q’h]estna?</td>
<td>he started working</td>
<td></td>
</tr>
<tr>
<td>n + r → ñ</td>
<td>[nr]t’æn’</td>
<td>[n]t’æn’</td>
<td>you saw him</td>
</tr>
</tbody>
</table>

(13)

| χ + k → χ + k           | ò[χk’]r atiix | ò[χk’]r atiix | he is barely walking around |
| k + r → k + r           | ñe[κ’r]astets’| ñe[κ’r]astets’| I was walking with a cane |

assimilation and deletion. This point is take up in more detail in Chapters 5 and 6.

7In Athapaskan literature, stops and affricates are both treated as stops and this is reflected in the chart. Also, lateral is included as a place of articulation since it patterns with other places in the range of laryngeal and manner distinctions laterals enter into.
Note that in each case in (12), the first consonant in the underlying cluster is a coronal stop. It gives its stop and laryngeal feature to the output consonant but we see no evidence from these forms for its place feature. The place feature of the second underlying consonant, however, is maintained in the surface consonant.

The forms in (13) demonstrate that clusters in which the first consonant is non-coronal do not undergo coalescence. Instead, both consonants are maintained in the Lower dialect surface form. Thus, the coalescence only applies when the first consonant is coronal, and only the coronal place feature is inactive in this process.

To summarize at this point, in the assimilation and coalescence patterns seen in this section, the feature coronal behaves as inactive in that it is not a trigger for assimilation and its features do not remain in coalescence and assimilation. In contrast, other place features are active, patterning as triggers for assimilation and persevering under coalescence. This is consistent with the generalization linking phonological activity of a feature to markedness: unmarked features are inactive while marked features are active.

Given that marked segments are in general dispreferred when an unmarked segment is possible, one might expect the direction of the asymmetry to be reversed, and the unmarked specification to be active. Thus, the direction of this asymmetry is surprising considering that the spreading of a marked specification to an unmarked segment results in a marked output segment.

I will revisit this point in the next chapter, and demonstrate the difficulty the existing asymmetry poses for current view of markedness in OT. In the view of markedness espoused here, the direction of the asymmetry is
expected. The link between phonological activity and markedness is straightforward result of the theory: a specification is phonologically active for the same reason it is marked. The presence of a specification in the input means that particular specification is likely to be preserved by the Faithfulness constraint MAXF. The preservation of this feature is alone evidence of its activity. Furthermore, that feature is available for spreading in response to constraints that require it. Thus, markedness and activity are connected: the presence of a feature specification means that this specification is active, but it also means that it is more marked as it entails more structure and therefore more violations of *STRUCTURE.

In a previous account of these assimilation patterns, Avery and Rice (1989) suggest that the unique patterning of the coronal nasal reflects a lack of place specification in the underlying representation of the segment (see also Kiparsky (1985), and papers in Paradis and Prunet (1991)). The segment can get a place specification in the phonology via spreading of place from a neighbouring consonant, or failing that, the feature coronal will be filled in by redundancy rules in the phonetic implementation component.

The analysis argued for here is similar to the one provided by Avery and Rice (1989): marked specifications are active because they are present in the input. The unmarked specification coronal is inactive because is absent in underlying specifications. The interaction of representations with Faithfulness has the asymmetries fall out automatically. Assimilation of a more marked segment to a less marked one entails loss of input structure (MAX violations) while assimilation of less-marked to more marked involves no MAX violations. For example, since labials have more input structure than coronals, assimilation of a labial to a coronal is worse than assimilation of a coronal to a labial because there is more structure to map (be faithful to)
in labials. Thus, coronal place of articulation will assimilate to a dorsal or labial consonant (as in (14)), even though the reverse pattern produces a more harmonic, less marked form. These assimilation patterns (represented schematically in (14b)) simply represent cases where Faithfulness works to keep the maximum input structure in the output, despite the violations of *STRUCTURE constraints. The tableau in (15) demonstrates the evaluation of a candidate set for a schematic coronal-labial input sequence.

(14a) \( t + p \Rightarrow pp, *tt \)
\( p + t \Rightarrow pt, *tt \)

(14b) \[
\begin{array}{ll}
\text{Rt} & \text{Rt} \\
\text{Place} & \text{Place} \\
\text{Peripheral} & \text{Peripheral}
\end{array}
\]

\(*Rt \quad Rt \quad R /
\quad / \quad / \quad /
\quad \text{Place} \quad \text{Place} \quad \text{Place}
\]

(15) Input: /...Rt + Rt.../

<table>
<thead>
<tr>
<th>Candidates</th>
<th>SPREAD*</th>
<th>MAXF</th>
</tr>
</thead>
</table>
| a. Rt Rt
   \[ \text{per} \] | *! | |
| b. Rt Rt
   \[ \text{per} \] | | |
| c. Rt Rt
   \[ ? \] | *! | <per> |

---

8Here, I use the constraint SPREAD (from Kiparsky (1994)) to represent some higher-ranking constraint which would compel two segments to share a specification.
The first candidate fails to assimilate and, therefore, is ruled out by the spreading constraint. The second candidate has the output correspondent for the labial specification doubly-linked to both consonants in the output, satisfying both the SPREAD and MAXF constraints. In contrast, the third candidate lacks an output correspondent for the input place specification, and is ruled out by the Faithfulness constraint, MAXF.

To summarize, a specification is marked because it is structurally complex: complexity in representations leads to more violations of *STRUCTURE. A specification is also active because it is structurally complex: having additional structure in the input means that that specification can resist being lost in the output and it can spread to other segment types. Thus, it is not simply an accident that marked specifications are active in the phonology, but instead it falls out from the relationship between input representations and markedness. In the next section, I describe briefly how the degree of specification of input representations is determined from the output patterning of segment types.

3.2.2 Determining Inputs

The possible representations for a particular segment type differ in terms of degree of structure: one possible representation may have additional structure that is not included in an alternative representation. The presence vs. absence of this structure in the input has consequences for the patterning of a segment type in the output phonology. These consequences provide empirical evidence of the non-universality of inputs. How do we ascertain whether additional structure is included in a particular segmental representation? My answer is to follow the same method reasonably used to determine the input forms of morphemes.
We can tell that the input representations of a pair of morphemes differ by examining how Faithfulness constraints respond to them in different morphological contexts. At the segmental level, if an output form of a morpheme contains a particular segment which cannot be shown to satisfy the requirements of any well-formedness constraint (e.g. ONSET) then we know that it is part of the input, and maintained by Faithfulness. Thus, in the case of a hypothetical output form of a morpheme [slak], we know that the second segment [l] is part of the input form of the morpheme. This must be true because it serves the needs of no well-formedness constraint. In fact, it violates a universal constraint against consonant clusters. If we were relying simply on well-formedness to select an output for this morpheme, we might expect a form such as *[sak] to be optimal. For [slak] to be chosen over *[sak], there must be a constraint that is higher-ranking than the anti-cluster constraint (given in (16) as *CC) that is satisfied by [slak] and not *[sak].

(16)

<table>
<thead>
<tr>
<th></th>
<th>??</th>
<th>*CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. slak</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. sak</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

As already noted, it is difficult to see how candidate (a) is preferred over (b) in terms of well-formedness. The only way to compel the presence of [l] in the output form is to posit an /l/ in the input,9 and therefore the higher-ranked constraint in (16) must be a Faithfulness constraint.

If a segment is only present in the output form of a morpheme when it satisfies some well-formedness constraint, we know that it is not part of the

---

9This method for determining inputs is of course more than reminiscent of the familiar principle of predictability followed in constructing underlying representations in derivational phonology. Despite its familiarity, I include such a discussion to demonstrate that the same principles are applied in the determination of segmental input representations.
input representation. This is because we know that, all else being equal, Faithfulness must work to preserve input elements. If a segment fails to appear in any instantiation of a morpheme where it is not required for well-formedness, then that element is not part of the input.

To illustrate with an example, I present a hypothetical evaluation in (17). In the hypothetical form in (17), the stem is affixed with a suffix /-d/ and, in the optimal output, appears with a final vowel [tata-d]. Given another possible candidate form *[tat-d], one might ask why this second candidate fails to be the optimal output for this form. That is, what constraint is fatally violated in this candidate that is not violated in the actual output form?

(17) i. Input /tat-d/ ii. Input /tata-d/

<table>
<thead>
<tr>
<th></th>
<th>*CC</th>
<th>FAITH</th>
<th>*STRUCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tata-d</td>
<td></td>
<td></td>
<td>!11</td>
</tr>
<tr>
<td>b. tat-d</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>*CC</th>
<th>FAITH</th>
<th>*STRUCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tata-d</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| b. tat-d  | *(!12)| *(!)  | *

In the tableaux in (17), these two candidates are evaluated. Since we have not yet determined the input form of the stem, we do not know if the input form contains a final vowel. Thus, two tableaux are given: tableau (i) demonstrates the evaluation of the candidate pair given an input form which lacks a stem-final vowel; tableau (ii) evaluates the same candidates given an input which includes a stem-final vowel. With any input, there is a well-formedness constraint against consonant clusters that is violated in (b) and not in (a). This is likely the reason for the dispreferred status of (b). If the

---

10 Let us assume for the purposes of the present discussion that we have already ascertained that the vowel in question is not part of the input form of the affix.

11 I have included the *STRUCTURE constraint because it becomes relevant in the next tableau. Here, the only violations that are shown are the ones which separate the two candidates. Thus, since the first candidate contains an additional vowel, it incurs an additional violation of *STRUCTURE.

12 Since we cannot establish at this point the relative ranking of *CC and FAITH, I include parentheses around the fatal violation markers (!) in (17ii) to indicate that the fatal violation belongs to the constraint which is higher-ranked.
input contains a final vowel, as in (ii), then the (a) candidate may also preferred over (b) for reasons of Faithfulness. If the final vowel is not part of the input, the Faithfulness constraint is violated in the optimal output form, since it contains a final vowel. Thus, the correct output form can be explained with either possible input: if the input does not contain a final vowel, then (b) is ruled out by *CC. If the input does contain a final vowel, then (b) may be ruled out either by *CC or Faith.

The input form of the stem can be determined when we look at the unaffixed output form of the stem [tat]. If the input form of the stem is /tata/, then we need to explain why the final vowel is deleted in this form.13 If the input form of the stem is /tat/, then no motivation for the output [tat] is necessary: it is a Faithful mapping. The following tableaux demonstrate the evaluation of the actual output candidate [tat] relative to another possible output *[tata]. Tableau (i) shows an evaluation relative to an input lacking a final vowel, while tableau (ii) shows an evaluation given a vowel-final input stem.

i. Input /tat/

<table>
<thead>
<tr>
<th></th>
<th>*CC</th>
<th>Faith</th>
<th>*STRUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. tata</td>
<td></td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

ii. Input /tata/

<table>
<thead>
<tr>
<th></th>
<th>*CC</th>
<th>Faith</th>
<th>*STRUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tat</td>
<td></td>
<td></td>
<td>!</td>
</tr>
<tr>
<td>b. tata</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

In (i), the identity between the first candidate and the input form means that this candidate satisfies Faithfulness. The second candidate contains an

---

13One reason such a deletion is unexpected is that it creates a closed syllable, which is considered to be dispreferred by a constraint against codas (NOCODA (P&S 1993)). On the other hand, there is the possibility of a different constraint requiring for example that stems end in a heavy syllable (c.f. McCarthy (1997) on Rotuman incomplete/complete phase distinction). However, such a templatic-type constraint would have to be well-motivated elsewhere in the patterning of other stems for it to be a viable analysis (c.f. footnote 9 on predictability above). Let us assume, for the present discussion, that such evidence does not exist in this hypothetical system.
additional vowel that is not contained in the first candidate, and since it has no high-ranking well-formedness constraint recommending it,\textsuperscript{14} it is ruled out. If the input is vowel-final, as in (ii), the first candidate would be unfaithful, and should not be chosen over the faithful second candidate. The black hand demonstrates that, given this input, Faithfulness should prefer the wrong candidate.

Thus, the absence of the final vowel in the input form of \([\text{tat/tata}]\) is fairly clear: it only occurs in output forms where it is motivated by a well-formedness constraint,\textsuperscript{15} and therefore its presence in the output is easily explained. Where it is not required by a well-formedness constraint, it does not occur. Contrast this with the "questionable [l]" of the previous example in (16). It is difficult to imagine a well-formedness constraint that would motivate its presence in the output, therefore Faithfulness must the reason for its presence. If Faithfulness is requiring its output presence, then it must be part of the input.

I summarize the characteristics of input elements in the left column of the table in (18). On the right side are characteristics of elements which are not part of the input.

\textsuperscript{14} Obviously, a constraint against codas would prefer the second candidate over the first. However, since the first candidate is in fact the optimal one, we know that NOCODA must not be highly-ranked enough to be decisive. The point of these tableaux is not to argue for some particular ranking, but rather to show how one input-output pair could be optimal.

\textsuperscript{15} Of course, the well-formedness constraint must be appropriately ranked relative to other constraints in order to be satisfied in this way. For the purposes of this discussion, it is important only that the output segment be in some way favourable to some non-Faithfulness constraint.
It is this same reasoning applied to segment structure that forms the basis of my argument for (under)specification in input segmental representations. Where the presence of output features can only be motivated by Faithfulness constraints, it is assumed that they are contained in the input specification. If a particular set of features occurs only in predictable environments and is not consistently preserved by Faithfulness, it is assumed that they are not contained in the input specification.

As discussed in the preceding section, under circumstances where there is a potential for feature loss (e.g. assimilation, coalescence, etc.), unmarked features are often not preserved by Faithfulness, while marked elements are (c.f. Kiparsky (1985)). Rather than assume an inexplicable bias on the part of Faithfulness constraints towards marked features, I argue that features whose output presence is not favoured by Faithfulness are to be treated just like segments whose output status goes unnoticed by Faithfulness: they are simply not part of the input. In this way, the behaviour of different segment types provides cues as to the input representation for different specifications.

For example, I have claimed that the input representation for the place representation for labial consonants involves a Place node and a dependent peripheral specification (c.f. the tableau in (15)). How is this input representation determined? The answer to this is parallel to the reason the [l] in the hypothetical output [slak] must be part of the input to that form:
because the only way for an output to appear with this marked structure is if Faithfulness requires it. If Faithfulness requires the output presence of some element, it must be part of the input. To parallel the arguments in the [slak] case, one might ask why the form with a peripheral specification (i.e. [p]) is chosen over one with an simple C-Place (i.e. [t]). As far as the markedness constraint *STRUCTURE is concerned, the less complex candidate is preferable. What kind of constraint could compel the additional structure of a labial specification? Faithfulness will, provided that the additional structure is part of the input.

I have also claimed that some input specifications must be underspecified. How do we know that the place specification of a coronal consonant is a simple C-Place in the input? The susceptibility of coronals to be over-ridden in assimilation and coalescence provides a clue. When other constraints such as SPREAD (c.f. tableau in (15)) require that a specification be spread to/shared by an input coronal segment, then the output form of the coronal undergoes that assimilation; Faithfulness does not intervene. With other place types such as labials and dorsals, Faithfulness ensures that SPREAD does not have its way. The fact that coronal place specifications are ignored by Faithfulness is evidence of their lack of a place dependent. This is parallel to the first characteristic of underspecified inputs shown in the table in (18). The behaviour of coronal place also fits the second characteristic of absent elements described in (18): well formedness constraints will ensure that the phonetic realization of plain C-place representations is coronal.\footnote{These same constraints also allow for plain C-place to be realized as a velar, as discussed in §3.6.} The constraints which determine this phonetic realization of underspecified representations are discussed in detail in §3.6. I claim that this complete
predictability is also support for the absence of a specified place feature on coronals in the input.

Underspecified input representations allow an elegant explanation for the special treatment of certain segment types by Faithfulness. The fact that these segment types are unmarked as well as being designated as special by Faithfulness comes down to the complexity of their representations, both in the output and in the input. In the next section I address some arguments in the literature against the importance of input representations as a determinant of output phonology.

3.3 The Relevance of Input Specifications

One of the major claims of this thesis is that input specifications have an impact on the output. However, this claim runs counter to the view espoused in most current OT work which seeks an explanation for any phonological patterning in the interacting requirements of output constraints. Thus, required underspecification of input representations has been rejected as a possible explanatory tool.

One of the main arguments for rejecting underspecification theories within OT is that it is an unnecessary complication of the theory, since the constraint system alone can account for apparent underspecification effects:

Central to the optimality-theoretic enterprise is the hypothesis that explanation can be achieved through output constraints alone. Therefore, neither underspecification, nor anything else, can be meaningfully required of inputs.

(Itô, Mester, and Padgett 1995: 588-9)

Itô, Mester, and Padgett (1995) examine one arena in which underspecification has played an essential role: the behaviour of redundant features. They show that constraint interaction will select the correct output representation, even where there are several possible input representations.
Itô, Mester, and Padgett (henceforth IMP) address the issue of nasal voicing in Japanese and the complications it presents for a theory of underspecification. Like other sonorants, nasals are redundantly voiced and therefore may be expected to be inert with respect to triggering or blocking of phonological processes which refer to voicing. This expectation seems to be fulfilled in the interaction of Japanese Rendaku and Lyman’s Law. Rendaku is the term for the voicing of the initial obstruent in the second member of a compound, as demonstrated in the examples in (19) (taken from IMP (1995: 574)).

(19)

a. ori + kami → origami paper folding  
b. oo + sumoo → oozumoo grand sumo tournament  
c. yama + tera → yamadera mountain temple

Rendaku is over-ridden by Lyman’s Law, which prohibits more than one voiced obstruent per root. Therefore, where the second member of a compound already has a voiced obstruent, the initial-obstruent will not be voiced by Rendaku. Some examples (taken from IMP (1995: 575)) are given in (20).

(20)

a. širo + tabi → *širodabi white tabi  
b. ore + kugi → *oregugi broken nail  
c. mono + sizuka → *monojizuka tranquil

Although phonetically voiced, sonorants do not behave like voiced segments with respect to Lyman’s law. That is, Rendaku still applies when there is a sonorant in the second member of the compound. Thus, the m in the second morpheme in (19a) does not block the voicing of the initial obstruent.
This patterning of sonorants falls out naturally if it is assumed that sonorants do not have an underlying specification for [voice]. Instead, this is filled in by a redundancy rule ordered after the application of Rendaku and Lyman’s Law. However, another constraint in Japanese phonology requires that nasals have a voice specification. This constraint ensures that all nasal-obstruent clusters are voiced. Some examples of this constraint at work are given in (21), taken from IMP (1995: 575).

(21)

a. yom + te → yonde reading
b. ŝin + te → ŝinde dying

It seems that the feature [voice] of the nasal is shared by (or spread to) the obstruent in these clusters. This in itself is not a problem for an underspecification account: the nasal-obstruent voicing constraint could hold after the application of the redundancy rule filling in voicing of nasals (as in Itô and Mester (1986)). The problem is that for the purposes of Rendaku and Lyman’s Law, these clusters are treated as though they are voiced. If Rendaku and Lyman’s Law apply before the redundancy rule inserting [voice], they should not be able to see the voicing on a cluster which results from the redundancy rule.

IMP argue that such a paradox does not arise in a non-derivational, constraint-based approach such as the OT analysis that they suggest. In their analysis, they exploit the relationship between feature redundancy and licensing in terms of antagonistic constraints. One constraint requires the presence of the redundant feature, while the other constraint requires proper licensing of features (License(φ)). Conflict arises between these two constraints because of a restriction on licensing where a feature that is redundant (i.e. inherent) to a segment cannot be licensed by that segment. In the case of
nasals, the fact that they are inherently (i.e. redundantly) voiced means that they cannot license [voice], yet the constraint NasVoi requires [voice] on nasals. An example tableau illustrating the interaction of these constraints is given in (22).

(22)

<table>
<thead>
<tr>
<th>Candidates</th>
<th>License</th>
<th>NasVoi</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. kami</td>
<td></td>
<td>*!</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>b. kami</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The question now is which candidate (and therefore which ranking) is correct. This is answered by the patterning of this type of form with respect to Rendaku and Lyman's Law. Example (19a), in which Rendaku has applied, demonstrates that the m is not treated as a voiced segment. This indicates that the first ranking in (22) is the right one since it appears that (b) is the correct representation for the morpheme and it is the first ranking (License >> NasVoi) which selects (b) as the output candidate.

The next case to address is the one in (21) in which nasal segments appear to share the feature [voice] with an obstruent. A tableau evaluating several candidates is given in (23).
This cluster behaves like a voice segment with respect to Rendaku and Lyman's law, so it will not have the same representation as *kami*, but instead must carry the feature [voice]. Since *License* is highly ranked, we cannot simply allow the feature to be unlicensed as in (b). However, in the case of nasal-obstruent clusters, there is a third option in which the feature [voice] may be present to satisfy *NasVoi* and at the same time be properly licensed: the obstruent can license the feature [voice] which it shares with the nasal.

A key point that IMP make is that whatever voice specification nasals have in the input representations, the constraints as they are will force the same output representation. This output representation may itself be underspecified, as they note:

> Underspecification, then, is an emergent property of the output. 17 Whether redundant voicing is underlyingly present or not, the outputs converge on the same output core (parsed substructure), with /m/ not linked to voice. (Itô, Mester, and Padgett 1995: 29)

Thus, the authors do not seek to get rid of underspecification altogether, since it is crucial in the case of single nasal consonants, but instead they wish to show the irrelevance of input specifications. The analysis still makes a distinction between representations based on the temporary absence of a

---

17The highlighting is mine.
feature, since presumably the /m/ in kami will be voiced in the phonetic component. There are therefore three places where nasals can get voicing: in the underlying representation, the phonetic spellout component, and from GEN. Only the last has consequences for the phonology.

IMP thus show that in the case of nasal voicing in Japanese, the input may be irrelevant given the tension between the presence of redundant features and licensing requirements of such features. However, their analysis does not show that Optimality Theory is incompatible with underspecified input representations except in as much as constraints hold only of output forms and not of inputs. In fact, they demonstrate that such underspecification is compatible with their analysis, even if it adds nothing to it descriptively. Therefore the inputs to tableau (23) could easily involve underspecified inputs where the nasals have no underlying feature [voice] but instead get such a feature from GEN in order to satisfy NasVoi where possible (i.e. where License may also be satisfied). In this case, assuming the presence or absence of [voice] in the input does not affect the selection of the optimal candidate.\(^{18}\) In fact, IMP point out that given the assumption of Lexicon Optimization of P&S (1993), single nasal consonants will be underspecified in the input. In §3.5, I return to the important issue of how output constraints can help drive underspecified representations, and how those representations affect the output.

In their summary, IMP argue that output constraints alone can account for the asymmetries in phonological activity: feature-specific Faithfulness

---

\(^{18}\)This follows from the fact that lexical optimization (Prince and Smolensky (1993), Itô, Mester, and Padgett (1995)) requires that the choice between several possible input representations be made on the basis of identity with the optimal candidate: since the optimal candidate in the case of single nasal candidates is one without [voice], this candidate will be closest or most faithful to an input representation that also lacks [voice]. We return to the issue of lexicon optimization in §3.5.
constraints can be ranked in such a way as to render marked feature specifications active and unmarked feature specifications inactive. However, I will argue in Chapter 4 that this type of account fails to offer an explanation for why this ranking should be true cross-linguistically, especially since it results in the selection of anti-harmonic output representations.

In contrast, if we follow the approach proposed here, segmental representations are directly linked both to markedness and to phonological activity. The degree of representational complexity in marked specifications makes them less preferred segments for situations where *STRUCTURE has its way (e.g. in epenthesis, reduplication, positions of neutralization). At the same time, Faithfulness considerations mean that the amount of structure involved in a marked input specification makes a segment less likely to be a target for assimilation and more likely to be a trigger. The difference in the degree of specification of marked vs. unmarked forms must be encoded in input representations. As discussed in the previous section, the intervention of Faithfulness on the part of marked elements means that those elements must be present in the input representations. On the other hand, unmarked elements, which are inactive in the phonology, are not favoured by Faithfulness and therefore must be absent from input representations.

3.4 Variability in Markedness

In the discussion of representations in §3.1, I noted that although markedness of specifications is always tied to the degree and type of structure involved, the representation of a particular specification is somewhat variable. A particular segment type may pattern differently with

---

19Chapter 6 explores the question of structure type in greater detail. There I propose a distinction in the evaluation of vertical complexity of structure (i.e. branching structure) versus horizontal complexity of structure (i.e. embedded structure).
respect to markedness considerations in different languages under different conditions. In this section I show that, while there is a certain degree of variability, that variability falls out from the proposed view of markedness in a constrained and principled fashion.

There are two ways in which markedness may vary between languages or contexts. One is in the complexity of the representation of a specification; the other is in terms of the relative ranking of a very limited type of segmental well-formedness constraint. First, I will demonstrate that much of the variability in consonantal place markedness is a simple case of output constraint interaction involving a very limited and well-motivated set of constraints. Second, I will discuss variability in vowel place markedness and demonstrate how variability in representations results in the selection of different vowel types as default or unmarked vowels. Much of the variability in representations can be directly associated with the number of contrasts in a given inventory, and therefore is predictably constrained.

3.4.1 C-Place Variability

It has been widely noted in segmental phonology literature that coronal segments are less marked than labials and dorsal segments with respect to assimilation, transparency, and epenthesis. However, as addressed in Lombardi (1995,1997) coronals are not always the least marked segment types in a given language. Instead, glottals, /ʔ/ and /h/, often appear to be the most unmarked segment type cross-linguistically. For example, Lombardi (1996) points out examples where place neutralization in coda position results in a glottal: in Slave (Athapaskan, Rice 1989) coda place neutralization results in /h/, and in Kelantan Malay, consonants neutralize to /ʔ/ in coda position. Also as pointed out in Lombardi (1997) many languages prefer /ʔ/ or /h/ for epenthesis and other emergence of the unmarked environments (e.g.
reduplication). The question raised here is, if coronals are the least marked segment type according to the markedness constraint hierarchy, why are glottals treated as least marked in some languages?

A representational view of markedness allows an elegant account of the markedness generalization: both coronals and glottals (/ʔ/ and /h/) are accorded special status with respect to markedness, each under different conditions in different languages (Causley 1997b, Rice and Causley 1998); see also Steriade (1987a) and Rose (1996) on laryngeals and placelessness.

The proposal made here is that markedness be encoded as structural complexity: the more marked a segment is, the more structure it has ((Steriade 1987b), Avery and Rice (1989), Rice and Avery (1993), Goad (1993), Dyck (1995)). The degree of structural complexity of a segment is constrained by *STRUCTURE, a general family of constraints on representational complexity. *STRUCTURE always prefers the least complex segment type. Thus, given the representations in (2) above (repeated below for convenience), according to *STRUCTURE, laryngeals are actually the least marked segment type since they are completely lacking in Place structure.

(2)

\[
\begin{array}{cccc}
\text{laryngeal} & \text{coronal} & \text{labial} & \text{dorsal} \\
\text{Rt} & \text{Rt} & \text{Rt} & \text{Rt} \\
\text{Place} & \text{Place} & \text{Place} & \text{Place} \\
\text{Peripheral} & \text{Peripheral} & \text{Peripheral} & \text{Dorsal}
\end{array}
\]

This means that in some languages and in some positions, this complete lack of structure will be preferred, due to a high-ranking *STRUCTURE. For example where *Place constraints hold, such as in coda position, a laryngeal will not incur a violation and therefore may be the only segment permitted in a coda.
Similarly, if a language prefers that GEN provide the absolute minimum amount of epenthetic material where necessary, a laryngeal is the best choice. The tableau in (24) illustrates the evaluation of a schematic candidate set for a VC input.

\[
\begin{array}{|c|c|c|}
\hline
/VC/ & \text{Onset} & \ast \text{Structure} \\
\hline
1. \text{VC} & \ast! & \\
\hline
\tilde{\ast} 2. \tilde{\gamma}^{20}\text{VC} & & \\
\hline
3. \text{tVC} & \ast! & \\
\hline
\end{array}
\]

The highly ranked Onset constraint rules out the first candidate which lacks an onset. The second and third candidates both involve epenthetic consonants. The difference between these two last candidates is in the amount of structure inserted by GEN. Since the laryngeal involves less place structure than /t/, there are fewer *Structure violations incurred by the second candidate than the third. By *Structure, this second candidate is optimal.

In many languages, however, laryngeals are not the preferred segment type for the emergence of the unmarked (TETU) environments, and are often absent entirely from an inventory. Thus, it seems that complete lack of structure is not the only consideration in determining the markedness of a segment. Instead, some languages seem to require consonants to have some place structure. Evidence for this requirement is found in languages where epenthetic consonants share place structure with an adjacent segment, despite the dispreferred linked structure that results. In languages where this

\[\text{\textsuperscript{20}I use the symbol /\gamma/ to represent the class of laryngeal segments rather than a particular segment in the class.}\]

\[\text{\textsuperscript{21}This sort of requirement is seen in languages such as Navajo, where epenthetic consonants share place with a following vowel. In the following paradigm (from Young and Morgan 1987) the epenthetic consonants in the output forms are bolded and the input forms are given in}\]
requirement holds, laryngeals are marked. Where /?/ is marked, /t/ will be the most unmarked. The special status of /t/ also falls out from its representation. Since it has a place node, it has more structure than /?/, therefore it is more marked. However, it is not deficient in the way that /?/ is because it has a place node. Thus, where [place] is required, /t/ will be less marked than /?/.

As Lombardi (1997:164) notes, "[C]oronal epenthesis is only seen when something additional to Place markedness is active." This additional constraint could be a SEGPLACE constraint, in the same family as the SEGHEAD constraint from Ito and Mester (1993). This constraint would require each segment to have some place structure. The tableau in (26) demonstrates the effect of such a constraint.

(25) SEGPLACE
  Every Root node dominates a place node.

(26)

<table>
<thead>
<tr>
<th>/VC/</th>
<th>ONSET</th>
<th>SEGPLACE</th>
<th>*STRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. VC</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ?VC</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. tVC</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>4. pVC</td>
<td></td>
<td>**!</td>
<td></td>
</tr>
</tbody>
</table>

slashes. Note that the epenthetic consonant agrees in place with the following vowel. Since the set of forms in (b) involve an initial consonantal prefix, no epenthesis is required. See Causley (1997a) for detailed discussion.

(a) yiêteeh /i-i-t-teeh/ 'I bury (animate obj.)'
    niêteeh /ni-i-t-teeh/ 'you bury'
    yiêteeh /i-i-t-teeh/ 's/he buries'
    woieteeh /oh-i-t-teeh/ 'you dpl. bury'

(b) niêteeh /ni-i-t-teeh/ 'I lie down'
    niêteeh /ni-i-t-teeh/ 'you lie down'
    niêteeh /ni-i-t-teeh/ 's/he lies down'
    noieteeh /n-oh-i-t-teeh/ 'you dpl. lie down'
Similar to the tableau in (24), the high-ranking ONSET constraint rules out any non-epenthesizing candidate (given the VC shaped input). The last three candidates all contain an epenthetic consonant in the onset position. The second candidate has an epenthetic //?/. While the lack of place structure of //?/ is preferable to *STRUCTURE, it violates the higher-ranked SEGPLACE and is ruled out. The choice between the remaining candidates is made on the basis of *STRUCTURE: since the last candidate has a labial epenthetic consonant, it incurs more violations of *STRUCTURE than the candidate with a coronal consonant.

To summarize, a structural view of segmental markedness allows us to define an unmarked set of segments (/t/ and /?/) in a principled and constrained way. Importantly, this account only makes use of the anti-complexity constraint *STRUCTURE and the well-motivated segmental well-formedness constraint SEGPLACE. This approach expects only two segment-types to be unmarked and not some random class of segments, while proliferation of additional markedness constraints leads to the potential for a system to choose any segment type as unmarked. In Chapter 4 I show that such a difficulty arises in a markedness hierarchy account of these facts.

As a final point on variability in consonantal place markedness, I touch briefly upon the question of velar consonants as unmarked segment types. Rice (1996) presents evidence for the unmarked status of velar consonants in many languages. Although this may at first appear to be a contradiction to the generalization that only coronals or laryngeals may be selected as least marked, it is in fact unproblematic when one considers the convincing arguments made by Rice in support of her claim that velars and coronals share a single underlying representation. In the languages Rice discusses, coronals are either in complementary distribution with or alternate with
velars, a segment type which is representationally distinguished from dorsals. The representation shared by coronal and velar segments is that of the coronal specification given above in (2). In Rice's model, the different phonetic realizations of segments bearing this specification are a result of the application of a redundancy rule inserting the specification coronal in the case of coronal surface segments, or the phonetic interpretation of a bare place node in the case of velar surface segments. I follow Rice here in assuming that unmarked coronal and velar segments share the representation in (2). The different phonetic realizations of this single representation receive an account in §3.6 below.

I now turn to the question of variability in vowel place markedness, demonstrating how, as with consonant place markedness, this variability is subject to principled restrictions on what segment types may be unmarked.

3.4.2 Variability in Markedness II: V-Place

In §3.2 we saw that unmarked elements pattern differently from marked elements in that they are inactive (i.e. targets and non-triggers) in assimilation, while marked elements are active (i.e. triggers and non-targets). There are also asymmetries in vowel harmony processes that are parallel to the ones seen in assimilation, and therefore it seems likely that these asymmetries may also be attributed to featural markedness. In vowel harmony, it is often the case that one feature specification will spread to other vowels, while another feature of the same class fails to spread and furthermore, may be lost when a vowel undergoes the spreading. Parallel to the role of markedness in assimilation patterns, we may take the activity of one specification to be a diagnostic for its status as a marked element relative to an inactive specification, which is taken to be unmarked. In addition to spreading and assimilation processes, we saw that markedness also plays a
role in epenthesis and other emergence of the unmarked (TETU) situations, such as in reduplication. In TETU environments, unmarked segments are selected over marked ones. Thus, the selection of a segment as epenthetic, or as a reduplicative default (c.f. 2.3) may be an indication of the unmarked status of that segment in a particular language.

In this section, I take all of these "markedness diagnostics" together to demonstrate the unmarked status of a vowel place specification in a language. I show that different languages treat different place specifications as unmarked, and that the choice of unmarked specification is to some extent inventory-driven (Rose (1993), Walker (1993), Rice (1995), Causley (1998a), Rice and Causley (1998)). Perhaps under-recognized in the segmental markedness literature is the tendency for back vowels to pattern as the unmarked vowel type, in addition to front and central vowels (Goad (1993), Rice and Causley (1998)). For each of these vowel types, I present examples of languages which treat them as the least marked place type.

3.4.2.1 Front vowels as unmarked

In many languages, front unrounded vowels behave as though they are unmarked: they serve as the target for harmony but fail to trigger harmony themselves; and they appear as epenthetic segments and as default segments in reduplication.

3.4.2.1.1 Front vowels as phonologically inactive

Snider (1989) describes a process of vowel harmony in Chumburung, a Kwa language spoken in Ghana, which spreads the features back/round leftward to a preceding front vowel. In (28a-c), the final front vowel of the

---

22There are several conditions on this process that I leave out for reasons of clarity. First, the process does not apply if the final syllable of the first word is closed. Second, the vowels /a/ and /ɛ/ do not undergo this process, although /ɛ/ behaves like other front vowels in failing to appear in phonological words with back/round vowels. Snider suggests that the harmony process is therefore parasitic on the feature [atr]. Third, there is also a process of ATR-harmony
first word assimilates to the place of back vowel in the first syllable of the
second word. Notice that in (d), a final back vowel does not assimilate to the
front vowel in the first syllable of the following word.

(27)

\[
\begin{array}{c|c|c}
   i & u \\
   i & r \\
   e & o \\
   e & ç \\
\end{array}
\]

(28)

a. \text{ibrr } + ku & \text{iburu ku} & \text{ten voices} \\
b. \text{okpo } + ku & \text{okpo ku} & \text{witch's husband} \\
c. \text{an } + jono & \text{an } jono & \text{our dog} \\
d. \text{jon } + w & \text{jon } + w & \text{dog's mucus} \\

In Chumburung, the place feature of the front vowels is inactive,
behaving only as a target for harmony, while the place feature of the back
vowels is active, behaving as the trigger of the harmony. Thus, from this
process, the front specification appears to be the unmarked place feature for
vowels in this language.

In Arapaho, a Western Algonquin language, a front vowel /e/ becomes
a back vowel /o/ when it is adjacent to /o/ or is separated from /o/ by a glide
or laryngeal segment. The data in (30) are from Picard (1977: 238).

---
operating on these forms, which will be described further in Chapter 4.
Thus, in Arapaho, the place specification of the front vowel is inactive while the place specification of the back vowel is active.

Place specifications in front vowels also seem to be inactive in vowel harmony in Yawelmani. The inventory of Yawelmani is given in (31).

The language demonstrates a process of vowel harmony that spreads the place specification of back vowels to suffix vowels of the same height.

In this process, front vowels undergo the spreading, but do not themselves trigger spreading. Thus, the front vowel place specifications may be considered inactive, while back vowel specifications are active.
3.4.2.1.2 Front vowels as epenthetic segments

Since epenthesis is a context where we often see the emergence of the unmarked, we expect that unmarked vowel places will characterize epenthetic vowels.

In Yawelmani, front vowels appear as epenthetic vowels breaking up clusters of three consonants, as shown in the stem alternations in the following paradigms (from Kenstowicz (1994:109)).

\[(33)\]

\[
\begin{array}{cccc}
\text{future} & \text{dubitative} & \text{gerundive} & \text{non-future} \\
pa\text{?t-en} & pa\text{?t-al} & pa\text{?it-mi} & pa\text{?it-hin} & \text{fight} \\
lihm-en & lihm-al & lihim-mi & lihim-hin & \text{run} \\
logw-en & logw-al & logiw-mi & logiw-hin & \text{pulverize} \\
\end{array}
\]

Since a cluster of three consonants arises from affixation of a consonant-initial suffix to a stem ending in two consonants such as the ones shown here, epenthesis occurs with the addition of the suffixes \(-mi\) and \(-hin\). No epenthetic /i/ occurs in the forms with a vowel-initial suffix, such as with the suffixes \(-en\) and \(-al\).

3.4.2.1.3 Front vowels as reduplicative defaults

Alderete et al. (1996) present an analysis of certain examples of fixed segmentism in reduplication which attributes the quality of invariable segments in reduplication to phonological rather than morphological requirements. They argue that the segments that appear as the non-copying elements of a reduplicative affix are often unmarked,\(^{23}\) an expected consequence of markedness constraints out-ranking Base-Reduplicant Faithfulness constraints. Thus, just as "the emergence of the unmarked"

\(^{23}\)Cases where invariant segments in the reduplicant are not unmarked (e.g. \textit{table-slimable}) are argued to be "true" cases of pre-specified segments, and are governed by morphological constraints.
ensures that unmarked elements are chosen as epenthetic segments, unmarked segment types also emerge in reduplication.

Yoruba (Pulleyblank 1988: 265) exhibits so-called fixed segmentism in the formation of deverbal nouns. Deverbals are formed through the affixation of the prefix C\- where C is a copy of the first consonant of the verb. The vowel /i/ appears as an invariant element in the prefix. Therefore, in (35a-e) the prefix vowel is always /i/, regardless of the quality of the stem vowel.

(34) Yoruba inventory (Pulleyblank 1988)

```
i  u
 e  o
 s  o
 a
```

(35) a. gbóná  gbí-gbóná  be warm, hot/ warmth, heat
b. dára  di-dára  be good/goodness
c. gbé  gbí-gbé  take/taking
d. rí  rí-rí  see/act of seeing
e. mu  mí-mu  drink/drinking

The invariance of /i/ in this reduplicative affix in Yoruba, assuming the position of Alderete et al. (1996), is easily seen as a reflection of its unmarked status relative to other vowels in the inventory.24

3.4.2.2 Back vowels as unmarked

The unmarked patterning of front vowels has been widely discussed (e.g. Pulleyblank (1988), Abaglo and Archangeli (1989), Paradis and Prunet (1991) and references therein, Hume (1992), McCarthy (1994), Rice (1995), and

---

24Pulleyblank (1988) presents several other types of evidence that /i/ is the unmarked vowel in Yoruba, such as susceptibility to assimilation, etc.
others). However, in addition to front vowels patterning as unmarked, there are also languages in which back vowels appear to be the unmarked vowel type, in that they are targets but not triggers of assimilation, and they surface in emergence of the unmarked environments such as epenthesis and reduplication.\textsuperscript{25}

3.4.2.2.1 Back vowels as phonologically inactive

Topping (1973: 51-53) describes a process of vowel harmony in Chamorro in which an initial back vowel of a stem assimilates to the place of a front vowel in a preceding particle. The Chamorro vowel inventory is given in (36).

(36) Chamorro inventory (Topping 1973)

\[
\begin{array}{c}
\text{i} & \text{u} \\
\text{e} & \text{o} \\
\text{æ} & \text{a}
\end{array}
\]

(37) a. guma? \hspace{1cm} house \hspace{1cm} i gima? \hspace{1cm} the house  
 b. foggon \hspace{1cm} stove \hspace{1cm} ni feggon \hspace{1cm} the stove  
 c. okso? \hspace{1cm} hill \hspace{1cm} gi ekso? \hspace{1cm} at the hill  
 d. tungo? \hspace{1cm} to know \hspace{1cm} in tingo? \hspace{1cm} we know  
 e. tungo? \hspace{1cm} to know \hspace{1cm} en tingo? \hspace{1cm} you know  
 f. otdot \hspace{1cm} ant \hspace{1cm} mi-etdot \hspace{1cm} lots of ants  
 g. lagu \hspace{1cm} north \hspace{1cm} sæn-lægu \hspace{1cm} northward  
 h. godde \hspace{1cm} to tie \hspace{1cm} g-in-edde \hspace{1cm} thing tied

In the forms in (37a-h), the initial vowels in the stems appear as back vowels when they appear in isolation, but as front vowels when preceded by a particle with a front vowel (as in (a-g)) or when they occur with a front vowel

\textsuperscript{25} Many of the ideas and observations in this section grow out of joint work with Keren Rice.
infix, as in (h). Thus, the back vowels serve as targets for assimilation, a characteristic we have ascribed to segments bearing unmarked specifications.

3.4.2.2.2 Back vowels as epenthetic segments

Paradis (1992) describes /u/ as the default (or epenthetic) vowel for Fula. The Fula vowel inventory is given in (38).

(38) Fula inventory (Paradis 1992)²⁶

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>ε</td>
<td>o</td>
</tr>
</tbody>
</table>

The epenthetic vowel occurs in contexts where syllabification of underlying consonants would violate one or more of several sonority constraints governing syllabification. One place where it occurs is between stems ending in a consonant cluster and certain consonant-initial suffixes, such as the infinitive suffix -dε.

(39) a. moml-u-dε to rub
    b. wors-u-dε to knot
    c. fort-u-dε to spread out
    d. naafl-u-dε to pray
    e. sokl-u-dε to need
    f. naam-dε to eat
    g. un-dε to pound

Notice that /u/ does not occur in forms in (f) and (g), where the stems do not end in a consonant cluster.

²⁶A process of [a̰]r harmony in the language gives the following phonetic vowel inventory.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>ε</td>
<td>o</td>
</tr>
<tr>
<td>ε</td>
<td>o</td>
</tr>
<tr>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>
Other languages reported to have /u/ as a default vowel specification are Telugu and Kannada, described in Bright (1975) as the *enunciative vowel*:

...loan words from Indo-Aryan or English thus regularly take an added vowel, for example, *pariṣattu* "society", *bassu* "bus". This vowel is usually pronounced as [i] in Tamil, Malayalam, Tulu, and in some Kannada dialects of North Karnataka; it is usually pronounced as [u] in other dialects of Kannada and in Telugu.

(Bright 1975:13)

### 3.4.2.2.3 Back vowels as reduplicative defaults

Alderete et al. (1997) treats the invariant appearance of /u/ in Marathi reduplicants as a morphological requirement as opposed to derived from markedness. However, since some languages treat back vowels as unmarked in other respects, we may take this case of fixed segmentism to derive from phonological considerations of markedness.

(40) Marathi (Apte 1968)

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>saman</td>
<td>luggage/ luggage, etc.</td>
</tr>
<tr>
<td>dhak</td>
<td>fear/ fear, apprehension, and the like</td>
</tr>
<tr>
<td>kh?ra</td>
<td>true/ true, genuine, etc.</td>
</tr>
</tbody>
</table>

### 3.4.2.3 Central Vowel is unmarked

In languages with central vowels, the central vowel behaves as though it is the unmarked vowel type in terms of phonological activity (Rice (1993), Walker (1993), Causley (1998), Rice and Causley (1998)) and in emergence of the unmarked environments such as epenthesis and reduplication (Rose (1993)).

#### 3.4.2.3.1 Central vowels as phonologically inactive

In Turkish vowel harmony, the high central vowel /i/ is an assimilation target, while front and back vowels are assimilation triggers. The Turkish vowel inventory is given in (41).
(41) Turkish vowel inventory

<table>
<thead>
<tr>
<th>front</th>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>rounded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>ü</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>e</td>
<td>ö</td>
<td>a</td>
<td>o</td>
</tr>
</tbody>
</table>

In the data in (42) the central vowel of the suffix assimilates to the place specification of the stem vowel.

(42) | nominative sg. | genitive sg. /-in/ | gloss |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. kiz</td>
<td>kiz-in</td>
<td>girl</td>
<td></td>
</tr>
<tr>
<td>b. sap</td>
<td>sap-in</td>
<td>stalk</td>
<td></td>
</tr>
<tr>
<td>c. ip</td>
<td>ip-in</td>
<td>rope</td>
<td></td>
</tr>
<tr>
<td>d. el</td>
<td>el-in</td>
<td>hand</td>
<td></td>
</tr>
<tr>
<td>e. pul</td>
<td>pul-un</td>
<td>stamp</td>
<td></td>
</tr>
<tr>
<td>f. son</td>
<td>son-un</td>
<td>end</td>
<td></td>
</tr>
<tr>
<td>g. yüz</td>
<td>yüz-un</td>
<td>face</td>
<td></td>
</tr>
<tr>
<td>h. köy</td>
<td>köy-un</td>
<td>village</td>
<td></td>
</tr>
</tbody>
</table>

Non-central vowels, on the other hand, do not undergo assimilation when they occur in a suffix, as in the data in (43) below from Clements and Sezer (1982). Notice that the place of the bracketed vowel is constant regardless of the quality of the stem vowel.

(43) a. fixed Coronal /istan/
| arab-[i]stan-i | Arabia (accusative) |
| mool-[i]stan-i | Mongolia (accusative) |
| türk-[i]stan-i | Turkestan (accusative) |

b. fixed Peripheral /adur/
| gid-ed[u]r-sun | let him keep going |
| kos-ad[u]r-sun | let him keep running |
| gül-ed[u]r-sun | let him keep laughing |
| bak-ad[u]r-sun | let him keep looking |
Thus, in Turkish, central vowels behave as the unmarked vowel type, while front and back vowels behave as marked.

3.4.2.3.2. Central vowels as reduplicative defaults

Central vowels may also appear in cases of reduplicative fixed segmentism, where markedness considerations often compel the occurrence of the least-marked elements.

In Thai (Noss 1964) the central vowel [a] appears in a type of total reduplication. Notice that a [a] occurs in each reduplicant, regardless of the quality of the vowel in the base.

(44)

| kin      | kin-kan²⁷  | to eat/wining and dining       |
| faj      | faj-faj    | fire, electricity/electrical system |
| thēw     | thēw-thāw  | row, section/general vicinity |
| mōo      | mōo-māa    | pot/pots and pans |

A central vowel /i/ also appears as the fixed element in Bamileke reduplication, shown in (45), below.

(45) Bamileke (Rose 1993, citing Hyman 1973)

| za       | ziza       | to eat       |
| to       | tito       | to punch     |
| keen     | kikee      | to refuse    |
| ben      | pipen      | to accept    |

3.4.2.4 Markedness variability: summary

To summarize briefly, it appears that languages differ with respect to which vowel place specification is the least marked. The range of choices for unmarked place, however, is not unrestricted. For example, languages do not appear to choose front rounded vowels as least marked. Instead, the choice seems to be limited to front unrounded, back rounded, or central unrounded

²⁷Some forms have optional realizations which copy the place feature of the vowel in the base, e.g. km-ken “wining and dining “.
vowels (see Rice and Causley (1998)). Furthermore, the choice does not appear to be completely free within this set, but is constrained by the shape of the inventory. In the languages discussed in §3.4.2, only those which lack a central vowel have a front or back vowel patterning as unmarked. When a language has a front-central-back contrast, the central vowel is chosen as the least marked (Rose (1993), Walker (1993), Rice (1995), Causley (1998a) Rice and Causley (1998)). Thus, the markedness status of different vowel specifications is linked to the inventory.

3.4.3 Markedness variability and inventory

Relating the markedness of a specification to the presence of contrasts in an inventory presents a different problem depending on the view of markedness that is assumed.

If markedness is structurally encoded, as assumed here, then the differences in one language's treatment of /i/ as unmarked, and another's treatment of it as bearing a marked specification has to do with differences in the representation of /i/. Marked specifications are more complex than unmarked specifications, therefore the complexity of /i/ is greater in one language than in another. Furthermore, if the markedness of /i/ is related to the inventory of the language, then the complexity of the representation is related to the inventory.

In contrast to this view is the view that input representations are unimportant in determining output candidates. Instead, all patterns arise out of the interaction of output constraints. In Chapter 4, I examine the generalizations regarding inventory shape and variability in the markedness of segment types from the perspective of a purely output constraint-based phonology.
In the next section, I outline exactly how input representations are built in response to inventory shape as well as considerations of complexity. All aspects of markedness behaviour are derivable from these input segmental representations. As we saw in §3.4.1, since markedness is represented in terms of structural complexity, a constraint such as *STRUCTURE will work to ensure the least complex segment type is inserted in epenthesis. The same considerations of complexity will also ensure that the least complex segment type is chosen as the unmarked fixed segment in reduplication.

As seen in §3.2, the relationship between phonological activity and markedness is easily captured under a representational view of markedness: marked segment types have specifications to spread and to be faithful to in the output. Unmarked segment types, on the other hand, lack specifications to spread, and have less structure to be faithful to in the output.

3.5 Building Inputs: the relationship between *STRUCTURE and contrast

I have argued that a general type of constraint which prohibits representational complexity governs segmental markedness. These constraints operate at different levels of the grammar, falling under the general family name *STRUCTURE.

As discussed in Chapter 1, the effects of these constraints are mitigated by the requirements of Faithfulness which ensure that input structure is preserved in the output. Since *STRUCTURE prohibits all output structure, a question that may be posed is whether it can be highly-ranked enough in a language that no output structure is permitted.

An obvious answer to this question is that in order to be effective for communication, languages have to be able to make at least a minimum number of distinctions (c.f. Kean (1975), Flemming (1995) on the functional
goal of maximizing the number of contrasts in a system). To make distinctions at the segmental level, at least two different segmental representations are required. Thus, to distinguish two segments or segment types, we need two representations as in (46).

\[(46)\]

<table>
<thead>
<tr>
<th>Segment A</th>
<th>Segment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root</td>
<td>Root</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>

Given the requirement of richness, *STRUCTURE cannot eradicate all the contrasts in a language by prohibiting any segmental structure. However, it can ensure that distinctions are represented with a minimal amount of structure. Thus, two distinctions can be made as in (46), or to minimally violate *STRUCTURE while still representing a contrast, only one of the segments need be elaborated, giving two possibilities as in (47a) and (47b).

Either one of these systems may be optimal as far as *STRUCTURE is concerned.

\[(47a)\]

<table>
<thead>
<tr>
<th>Segment A</th>
<th>Segment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root</td>
<td>Root</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

\[(47b)\]

<table>
<thead>
<tr>
<th>Segment A</th>
<th>Segment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root</td>
<td>Root</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b</td>
</tr>
</tbody>
</table>

Thus, the representation of a particular segment may change depending on whether it receives the additional structure in a two-way

---

28 However, *STRUCTURE may be relativized to a particular syllabic, metrical, or morphological position and therefore may be ranked so that it prohibits any contrasts in a particular position, creating a situation of neutralization (see, e.g. Steriade (1997)).

29 This is parallel to the claim made in Rice and Avery ((1995) that segment structure is built up monotonically in the acquisition of a system.
contrast. What is relevant to our concerns here is that the additional structure has implications for the markedness patterning of a segment. In one system, Segment A will pattern as unmarked (47b), while in the other system, Segment B will pattern as unmarked (47a).

In a system where three different distinctions need to be made, *STRUCTURE allows an additional level of structure to be added to one of the representations, as in (48).

<table>
<thead>
<tr>
<th></th>
<th>Segment A</th>
<th>Segment B</th>
<th>Segment C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root</td>
<td></td>
<td></td>
<td>Root</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td></td>
</tr>
</tbody>
</table>

Notice that, in this type of system with a three-way contrast, both Segment A and Segment B bear more than minimal structure, while a third segment is unmarked. Thus, in this system both A and B are expected to pattern as marked, while Segment C will be unmarked.

While we have discussed how *STRUCTURE works to minimize output structure and ensure, for example, the insertion of the minimally complex segment type in epenthesis, we have also looked at the complexity of the input representation of a segment as a determiner of markedness behaviour. For example, two segment types may behave differently depending on their input representations. If segmental representations are derived through considerations of minimal structure, and input segmental representations are also subject to these considerations, then *STRUCTURE must also influence input representations. The question raised by this observation is how an output constraint can be brought to govern input representations.

The answer to this question is found in the principle of Lexicon Optimization (Prince and Smolensky (1993), Inkelas (1994), Itô, Mester, and
Padgett (1995)). Lexicon Optimization allows inputs to be in some situations indirectly governed by output constraints. Itô, Mester, and Padgett (1995) define Lexicon Optimization as follows:

Of several potential inputs whose outputs all converge on the same phonetic form choose as the real input the one whose output is the most harmonic. (Itô, Mester, and Padgett (1995: 593))

This definition may seem unhelpful, given that all outputs are optimal under some evaluation, otherwise they would not be outputs. However, consider the possibility that any input may be posited for a particular output form. Thus, we could posit, that for a particular output [ABC] there are infinite possible inputs, including the completely identical input /ABC/ and the non-identical /ADC/. Thus, we have two possible input-output pairs as in (49).

(49) \[
\begin{array}{ccc}
\text{Input} & /ABC/ & /ADC/ \\
\text{Output} & \{\text{ABC}\} & \{\text{ABC}\}
\end{array}
\]

How do we decide which mapping is the actual one we are dealing with, and therefore which input is the correct one? There is no question that [ABC] is the optimal output, since it emerges as the output form. However, the output [ABC] is in a sense more harmonic under one mapping than under another. Since the harmony of a candidate is measured in terms of constraint violation, the more harmonic mapping is the one which allows the output candidate to incur the fewest and least serious violations. If we consider that these constraints include those governing the identity relationship between

---

30 Recall from the discussion in §3.2 that the determination of alternating forms is somewhat more complicated than suggested here. If a form alternates, both alternants and their conditioning contexts must be considered. For the present discussion, I limit the discussion to non-alternating forms.
inputs and outputs, the best mapping (other considerations aside) is the most Faithful mapping, that is the one with the identical /ABC/ input.

For our problem of discerning input segmental representations, I give the following tableaux in (50). These tableaux represent two different possible mappings given two different possible inputs: the tableau in (i) involves an input with fully specified representations, while (ii) involves an input with minimally specified input representations.

(50)

(i) INPUT

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>*STRUCT</th>
<th>FAITH</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
<td><em>(a)</em></td>
<td><em>(&lt;b&gt;)</em></td>
</tr>
</tbody>
</table>

(ii) INPUT

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>*STRUCT</th>
<th>FAITH</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td><em>(a)</em></td>
<td></td>
</tr>
</tbody>
</table>

Assuming that *STRUCTURE has pared down the output segmental representations so that only one member of a contrasting pair bears a specification, the question now is what the representations are in the input. The tableau on the left in (i) shows a mapping between fully specified input representations and the minimally specified output, while the tableau on the right in (ii) shows a mapping between minimally specified input representations and the identically specified output representations. Lexicon Optimization tells us to choose the input that helps the output be as harmonic as possible. Looking at the two mappings, then, we see that in both mappings, the output violates *STRUCTURE to the same degree. However, in terms of Faithfulness, the two mappings are not equal: the fully specified input representations mean a non-Faithful output, while the minimally specified input representations allow the output to be fully Faithful. Thus, the optimal input is the minimally specified one on the right in (ii).
Thus, \*STRUCTURE serves to guide the building of input representations. However, as mentioned above, the representations are still somewhat under-determined given considerations of representational complexity alone.\textsuperscript{31} In the case of a three-way contrasting system, two of the three segment types receive non-minimal specifications. What determines which segment-types receive specification? To take a more concrete case, in a vowel system with a three-way place contrast, the three place types are front, central, and back. Evidence from the phonology suggests that the additional specification is added to the front and back vowel types, and not the central vowel. Why is this? I assume that place specifications are drawn from a universal feature set (i.e. \{coronal, peripheral\}, (Rice 1994)),\textsuperscript{32} and there is no place feature associated with central vowels.

There is also a question with the two-way contrasting systems as to which segment type receives the additional structure, as discussed in §3.4.2. In this case, the phonology will provide the learner with cues as to which segment type is unmarked, and which is marked. Learners need not actively draw conclusions based on the markedness patterning of segment types, however, since Lexicon Optimization will again drive the construction of a unique input segmental representation. For example, recall from the discussion in §3.2 that the loss of an input specification involves violating

\textsuperscript{31}Note that I am only considering a single dimension (e.g. place) of a segment's representation. The complete representation of a particular segment cannot be determined from considerations of minimality alone since, for a particular segment, minimal specification can be achieved in more than one way. For example, in a three-vowel inventory, vowels may be distinguished minimally in a few different ways, depending on whether height or place features are specified. For the purposes of this discussion, I address only considerations of minimality within a particular class (e.g. place or height) and do not explore the interaction of feature types. For a detailed discussion of such interactions and the relationship between the ordering of contrasts and the building of segmental representations, see Dresher (1998).

\textsuperscript{32}The status of these features as universals in phonology has phonetic grounding in that their phonetic realizations (as primary place features) represent opposite ends of a continuum of F2 frequency (Ladefoged (1993), Flemming (1995)).
Faithfulness. If a specification is consistently over-ridden in the phonology, then the learner could do one of two things. She could posit a fully specified representation that involves the loss of input elements in the output when those elements are over-ridden. This would mean positing consistent violations of Faithfulness. Conversely, she could posit an underspecified representation for the unmarked segment in the input. This would mean that when that segment assimilates to another segment type, no features are lost and no violations of Faithfulness incurred. In terms of Lexicon Optimization, the underspecified representation is preferred, because it allows the optimal output to be more harmonie.

To summarize to this point, it is the interaction between contrasts in a system and *STRUCTURE constraints that guides the building of representations. Contrast between two segments entails a difference in representations. *STRUCTURE ensures that this difference is minimal. That is, structure is added to a segment only under the pressure of contrast and even then only one of a contrasting pair need receive the additional structure. This means that, where two segments contrast, one will have a 'mark' and one will be unmarked. When an additional contrast is introduced, the unmarked member of the first pair will need to be distinguished from the third element and will receive a mark of its own (c.f. Rice and Avery (1995) on monotonic addition of structure in acquisition).

This view is built upon the assumption that languages have their own input inventory, and that contrasts are built into the input representations of segments. In contrast to this view is the view currently assumed in most OT work that input inventories are universal, while language-particular output inventories are determined by output constraints. In Chapters 5 and 6 I explore various problems with this approach.
In the next section, I address this issue of phonetic realization for the representations and demonstrate how the place of articulation for underspecified phonological representations is determined by inventory shape and potential for auditory distinctiveness.

3.6 Phonetic Implementation and Inventory Optimization

3.6.1 From Representations to Phonetic Realizations

So far in this chapter I have demonstrated how the current proposal accounts for variability in the markedness relations between segment types, and for the correlation between that markedness variability and the number of contrasts in an inventory.

One major issue that has yet to be addressed is the relationship between the proposed phonological representations and their phonetic realizations. In some cases, the proposed representations are fairly abstract, and their phonetic realizations variable. In particular, I have proposed representations for vowel place specifications that allow three different vowel types to surface from a single representation. As argued in §3.4, evidence from phonological patterning suggests that one of three vowel places may be treated as the least marked vowel in a system, subject to the considerations discussed in §3.4.

Since markedness is evaluated in terms of structural complexity, the unmarked place will be the least complex phonologically, whatever its phonetic realization. I have proposed that the least marked vowel place specification is a simple V-Place node, lacking any dependents, as in (51). A learner is driven to the construction of this representation of the least marked vowel in response to the structure minimizing constraint *STRUCTURE. Thus, since /i/ is the least marked vowel in Turkish and Bamileke (see examples in (42) and (43)), its phonological representation in these languages is as below in
(51). The representations of /i/ and /u/ in these systems follow from the need to distinguish them from each other and the central vowel.33

\[
\text{(51)} \quad \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array}
\]

\[
\text{(52)} \quad \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array} \rightarrow \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array} \rightarrow \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array} \rightarrow \begin{array}{c}
\text{i} \\
\text{i} \\
\text{u}
\end{array}
\]

coronal \quad \text{peripheral}

In Turkish and Bamileke, the phonetic realization for the representation in (51) is /i/. However, in Yoruba and Yawelmani which lack a central vowel in their inventories, the least marked vowel type is a front vowel (see examples in (32) and (33)). This means that in these languages, front vowels have the representation in (51). Thus, the phonetic realization for this representation in Yoruba and Yawelmani is /i/, as in (53).

\[
\text{(53)} \quad \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array} \rightarrow \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array} \rightarrow \begin{array}{c}
\text{i} \\
\text{u}
\end{array}
\]

peripheral

In Chamorro and Fula, there are only two place contrasts in the vowels, and the least marked vowel is /u/. Since /u/ is the least marked vowel type (see examples in (36) and (38)), it has the least complex vowel place representation, namely the one in (51). Thus, the phonetic realization in Chamorro and Fula for the representation in (51) is /u/, as in (54).

\[
\text{(54)} \quad \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array} \rightarrow \begin{array}{c}
\text{i} \\
\text{u}
\end{array}
\]

---

33For ease of illustration, I use actual vowels [i, i, u] to represent vowel place types. Therefore, when a phonetic realization is said to be [i], I mean a front vowel in general. The issue of vowel height is ignored for the present, to be addressed in detail in Chapter 6.
Thus, the representation in (51) has three different realizations. Given this single phonological representation, how can we determine what the ultimate phonetic realization will be in a particular language? In this section I will argue, adopting proposals from Dispersion Theory (Flemming 1995) that considerations of inventory optimization drive the ultimate positioning of the unmarked vowel in the vowel space.\(^{34}\)

First I outline briefly the model proposed by Flemming and highlight a few of the predictions made which are relevant to the question of vowel positioning. Then I demonstrate how aspects of this model are directly compatible with the proposals of this thesis. I show how key suggestions from the model work in conjunction with the current proposal, provide an explanation for the phonetic realizations of unmarked vowels. Finally, I argue that, while the two models are in some ways compatible, Flemming's model cannot stand alone as a theory of phonological processes without missing important inventory-based generalizations.

### 3.6.2 Flemming (1995)

Following models such as Lindblom's (1986, 1990) Theory of Adaptive Dispersion, Flemming (1995) suggests that inventory shape is driven by the conflicting requirements of contrast: maximizing the number of contrasts and maximizing the distinctiveness of contrasts. The preference for a rich segmental inventory necessitates fitting multiple sounds into a finite

---

\(^{34}\)For a related approach to phonetic realizations of underspecified vowel representations, see Dyck (1995).
auditory space, while at the same time, the need for auditory distinctiveness requires that contrasting sounds be far apart in that auditory space. Thus the conflict between these considerations: the greater the number of sounds in a fixed space, the smaller the distance between those sounds. Conversely, the greater the distance between sounds, the fewer the number of contrasts possible.

The diagram in (55) schematically depicts a system with a three-way contrast. This system has certain advantages in terms of richness of contrast; however, the distinctiveness or distance between those contrasts is not as great as in the system depicted in (56). This system makes a simple two-way contrast, and therefore it is possible to situate the sounds further apart in auditory space. Thus, the system in (55) prioritizes the richness of contrast over auditory distinctiveness, while the one in (56) prefers auditory distinctiveness over richness of contrast.

(55) \[ \begin{array}{ccc} x & y & z \end{array} \]

(56) \[ \begin{array}{cc} x & z \end{array} \]

Flemming proposes two sets of constraints governing these two requirements: Maintain N Contrasts and Minimal Distance. The first type consists of a set of constraints each specifying the maintenance of a number (N) of contrasts. The second constraint requires that contrasts are maximally distinct, and therefore as far apart as possible along a particular auditory dimension.\(^{35}\)

\(^{35}\)The Minimal Distance constraint set involves a set of parameterized constraints specifying the minimally required distance between sounds. Distance is measured in terms of difference in feature values for a particular dimension, e.g. F2. For reasons of simplicity, I do not specify different Minimal Distance constraints in this discussion, but instead treat them as a single constraint.
For the purposes of the present discussion, I will not go into the details of his analysis, which involves proposals for a scalar-valued feature set. However, I will discuss some results of his proposals that have direct relevance to the project of this thesis.

Since the constraint sets are re-rankable with respect to one another, different inventory types fall out from different rankings. Where the requirements of distinctiveness win out over maintenance of contrasts, a single sound occupies an entire dimension, as shown in the following tableau.

The presence of three vowel places, as in the first candidate inventory means that Maintain 3-way Contrast is satisfied. The presence of two vowel places, as in the second candidate inventory, means that Maintain 3-way contrast is violated, but Maintain 2-way Contrast is satisfied. However, both the first and the second inventories do not provide the maximal auditory distinctiveness required by the high-ranking MINDIST. The final candidate

\[\begin{array}{|c|c|c|}
\hline
\text{candidate} & \text{MINDIST} & \text{Maintain 2-way contrast} \\
\hline
1. \text{i - i - u} & **! & * \\
2. \text{i - u} & * & * \\
3. \text{ii} & * & * \\
\hline
\end{array}\]

Note that there is some logical limit on the extent to which either of these constraint types are satisfied. That is, there is an upper limit on the number of contrasts that any human language will make, no matter how lowly ranked the Minimal Distance contraints. Similarly, there is a minimum number of contrasts a language must make. Thus, no language ranks Maintain N-way Contrasts so low that only a single vowel or consonant exists in the language.

For the sake of simplicity, I represent the evaluation of MinDist satisfaction as a single violation for every additional vowel in the inventory since the presence of the extra vowel reduces the auditory distance between vowels.

The vowel sound that results from a lack of contrast is shown here as a central vowel. The preference of a central vowel over other vowel types is determined under Flemming's model by an additional constraint type which minimizes effort expended in articulation. Since a central vowel presumably involves the least amount of tongue displacement, it is chosen as the best vowel by effort minimization.
inventory, involving only a single vowel place, maximally satisfies MINDIST. Since there is no place contrast, there is no need to keep the vowels distinct, therefore MINDIST is not violated as it is in candidate #1 and #2.

Where the maintenance of contrasts is given higher priority over maximizing distinctiveness, contrasts exist in the inventory. The number of contrasts is determined by the highest value of $N$ in the Maintain $N$-way Contrasts constraint that dominates MINDIST.

Thus, in the tableau in (58), the domination of Maintain 2-Way Contrast over MINDIST means that a two-way contrast exists at the expense of whatever auditory crowding results from the presence of two vowel types.

\[
\begin{array}{|c|c|c|c|}
\hline
& \text{Maintain 2-way contrast} & \text{MINDIST} & \text{Maintain 3-way contrast} \\
1. & i - i - u & **! & \text{**} \\
2. & i - u & \text{*} & \text{} \\
3. & i & \text{?!} & \text{} \\
\hline
\end{array}
\]

The first inventory type in this tableau makes a three-way place contrast. This means that Maintain 3-way Contrast is satisfied. However, the auditory crowding that results means a violation of MINDIST, which is higher-ranked than the contrast maintenance constraint. This candidate is therefore ruled out. The final candidate, on the other hand, makes no contrasts thereby satisfying MINDIST, but violating the higher-ranking Maintain 2-way Contrast. The second candidate wins out, satisfying Maintain 2-way Contrast while minimally violating MINDIST.

Where both contrast maintenance constraints dominate MINDIST, a three-way contrast is made in the output candidate inventory, as in (59).
Thus, the ranking of the contrast maintenance constraints relative to MINDIST determine the number of contrasts that are made in an inventory.

In addition to determining contrasts in a system, the MINDIST constraint set also serves to determine which vowels will be part of a given inventory. A constraint such as Maintain Contrast only states how many contrasts must exist within a particular dimension, and does not specify which particular vowels must contrast. For example, any system with two vowel places will satisfy Maintain 2-way Contrast. Therefore any of the systems in (60) would satisfy such a constraint.39

(60) a. [i u] b. [i u] c. [i i]

The choice between these three possibilities is made by the MINDIST constraint: the system in (a) is preferred over (b) and (c) because the contrast is maximally distinct. It is this function of Minimal Distance that I will argue is

<table>
<thead>
<tr>
<th></th>
<th>Maintain 2-way contrast</th>
<th>Maintain 3-way contrast</th>
<th>MINDIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>i - i - u</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>2.</td>
<td>i - u</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>3.</td>
<td>i</td>
<td>!</td>
<td>*</td>
</tr>
</tbody>
</table>

39There are, of course, additional possibilities which I leave out in the interest of simplifying the example. Other possibilities include systems with any two of the vowels from the following set: [i y i u u]. In fact, any intermediate vowels could also be chosen, given an adequately refined feature set. Flemming represents the distance between the above five vowels in terms of the following values for a feature set based on the frequency of the second formant (F2):

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>y</th>
<th>i</th>
<th>u</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>highest F2</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>high F2</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>low F2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>lowest F2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Distance is measured in terms of difference in values for features of a given dimension. Thus, [i] and [u] are maximally distinct since they have no feature values in common. If a third vowel is part of a system, the choice of [i] as the third vowel is determined by MinDist: [y] would be too close to [i] on the F2 scale, and [u] would be too close to [u].
relevant to the determination of phonetic inventories from the proposed phonological representations.

3.6.2 Dispersion Theory and Phonetic Realizations

Flemming's model takes an output-based view of contrast: contrasts in the output occur because output constraint rankings allow them to surface. In comparison, the proposal made here involves a different view of contrast, one that requires that contrasts be built into input representations, and therefore maintained in the output via Faithfulness. Thus, the contrast maintenance constraints presented in Flemming (1995) are redundant given the proposed view of contrast. However, the Minimal Distance constraint makes predictions about output inventory realizations that are interesting when viewed in light of the representations proposed here.

I have argued that the least marked place specification in vowels is represented as a V-place node that lacks any dependents. The phonetic realization of a vowel with this specification is underdetermined by this representation, and may in fact be any one of three places: front, central, or back. However, while the phonetic realization of this representation is variable, it is not random but rather is governed by considerations of inventory. I will show here that the inventory considerations are actually the requirements of auditory distinctiveness within a system.

Recall from the discussion at the beginning of this section that there are three possible phonetic mappings for the unmarked V-place representation, given again in (61).

(61) Root
   └ VPlace
In an inventory with a three-way place contrast, the unmarked vowel is a central vowel, as in (62). Why should this representation have this phonetic realization in this particular inventory type? The simple answer to this is that, since there is a front (i.e. coronal) and a back (i.e. peripheral) vowel in the inventory, the best place to situate a third vowel is centered between the two fixed vowel places.\textsuperscript{40} In terms of Flemming's Minimal Distance constraint, the central vowel [i] is the best choice because it is maximally distinct from both [i] and [u].

\begin{equation}
\text{(62)} \quad \begin{array}{c}
\text{Root} \\
\text{VPlace} \\
\text{coronal}
\end{array} \quad \begin{array}{c}
\text{Root} \\
\text{VPlace} \\
\text{peripheral}
\end{array} \quad \Rightarrow \quad [i \ i \ u]
\end{equation}

In languages with a two-way place contrast, one of the two V-place representations will bear a feature (either coronal or peripheral). When the marked vowel place is peripheral, the best choice for the placement of the under-determined vowel is front, as in (63). This allows the two vowels to be maximally distinct and satisfies Minimal Distance.

\begin{equation}
\text{(63)} \quad \begin{array}{c}
\text{Root} \\
\text{VPlace}
\end{array} \quad \Rightarrow \quad [i \ u]
\end{equation}

When the marked vowel place in a two-way system is coronal, the best placement of the unmarked vowel is as a back vowel, since this allows the contrast to be realized with maximal distinctiveness.

\textsuperscript{40}I assume the phonetic realization of phonologically coronal and peripheral vowels is fairly strictly determined by the presence of these place dependents. Thus, their realizations are fixed in a way that the realization of underspecified representations are not.
Thus, languages with two-way place contrasts often have the same phonetic inventory, even though their representations of vowel place are different. The same vowel representation may have different realizations, and will appear as a different vowel phonetically depending on what other representations exist in the system. However, the unmarked vowel place will demonstrate similar phonological patterning, despite its phonetic realization.

There are, however, phonetic inventories where vowels are not situated the maximal distance apart in phonetic space. For example, in Japanese, the phonetic vowels [i] and [u] in this two-place system are not as far apart in phonetic space as [i] and [u]. In Flemming’s proposal, the requirement for maximal distance conflicts with a desire to minimize effort in articulation. If we consider the lip-rounding involved in the articulation of [u] as requiring effort not involved in the articulation of [u], then we have a reason for selecting [u] over [u]. Thus, provided that MinimalDistance is dominated,\(^{41}\) inventories with less than optimally spaced vowels may be selected as a phonetic inventory.

A similar account can be made for consonant place. As discussed above in §3.4.1, we find variability in the selection of coronals over laryngeals as the least marked segment type. I argued that this variability is due to the variable ranking of a segmental well-formedness constraint, SEGPLACE. In addition to

\(^{41}\)Recall that Flemming’s MinDist constraint is actually as set of constraints parameterized to particular degrees of distance. Thus, for the phonetic inventory [i-u] to be chosen as optimal, only the most MinDist constraint demanding the maximal auditory distance need be dominated by the constraint minimizing effort.
this variability, I also noted that velar segments in some languages appear to be the least marked place in consonants (Rice (1996)). Following Rice (1996) I assume that velars\textsuperscript{42} share a representation with coronals. Parallel to the representation of the unmarked vowel place discussed above, the unmarked consonant place is a plain C-place node lacking any dependants. The interpretation of this unmarked C-place as two different phonetic segment types is determined by the same type of considerations as those determining vowel place realizations. The difference between coronals and velars in terms of acoustic properties is that velars, in addition to being auditorily similar to dorsals, are also close to labial consonants in terms of the second and third formant frequency transitions of their release. Coronals are therefore more distant from both dorsals and labials in terms of auditory distance. Thus, as far as the constraint MinimalDistance is concerned, an inventory with an undetermined unmarked consonant place that contrasts with either a labial or a dorsal segment, or both, is best to have the unmarked segment realized as a coronal. That way, auditory distance is maximized. On the other hand, if MinimalDistance is low ranked, the interpretation of the consonant will be velar, preferred for conserving effort through minimization of tongue displacement from a position of rest.

3.6.3 Going it with Dispersion Alone

Flemming (1995) does not intend for Dispersion Theory to be simply a model of phonetic implementation. Instead, he argues, Dispersion Theory unifies the phonology and phonetics module in a single component. The auditory representations he proposes are the only representations necessary to articulate phonological generalizations regarding inventory shape, segmental

\textsuperscript{42}Recall that in Rice's terminology velars are representationally distinct from dorsals, a more complex segment type.
patternning, markedness, etc. These representations are based on acoustic properties of individual sounds and are invariant from language to language and position to position. Therefore, this proposal does not provide a means by which to capture the generalizations that have dominated the discussion of this chapter, and the entire thesis: languages differ in segmental markedness relations and markedness is linked to phonological patternning. The markedness variability between languages is subject to constraints, and is to a large extent driven by contrasts in an inventory. How is this captured in a view with static representations of segment specifications?

In Flemming’s model, assimilations occur to maximize the duration of a particular feature type. Thus, assimilation of one segment to another involves the extension of a particular auditory feature at the expense of another, in order to satisfy a constraint such as MAXDURF (“Maximize the duration of F”). This constraint type is completely independent of the constraints determining inventory (i.e. the MAINTAINCONTRAST and MINIMALDISTANCE) constraint, therefore we might expect no relationship between inventory shape and the ranking of MAXDURF. However, as we have seen, phonological activity of features in assimilation phenomena are often quite closely connected to the presence or absence of contrasts in an inventory (see §3.4.2). Thus, since such a relationship does exist, this theory of phonology misses an important generalization.

Although coalescence phenomena are not treated in detail until Chapters 5 and 6, I will briefly point out a difference in the analysis allowed by

---

43Given the degree of detail in the feature sets or scales Flemming proposes, the representations for an individual segment or segment type are fixed across languages. However, it is not inconceivable that the feature set could be even more fine-grained (e.g. Highest F2, Medium High F2, Lower High F2) and a more precise acoustic characterization of sounds possible. These featural characterizations would nonetheless be unchanging for a particular sub-type of sound; for example, higher high front vowels would have the same feature values across all languages that had them in their inventory.
my proposal than the one permitted by Flemming (1995). The evidence presented there will show that, parallel to the patterns discussed in this chapter, in vowel coalescence phenomena the most marked features are maintained in the output forms. Here, the markedness of a segment type is once again linked to inventory shape. The view of markedness that I have argued for allows a straight-forward account of these patterns, directly relating markedness of a specification to the preservation of that feature in the output.

Flemming's theory cannot make the same link. Coalescence phenomena would be handled in his theory by a set of Output-Output constraints (Flemming (1995) and see, e.g. Benua (1997), Steriade (1997)). The contrast maintenance constraints proposed by Flemming (1995) fulfill many of the functions of input-output Faithfulness. In fact, Flemming rejects the notion of underlying representations, and therefore all input-output mappings. The identity between morphologically related forms is instead guaranteed by Paradigm Uniformity constraints which require identity between morphologically related forms. These Paradigm Uniformity constraints can motivate the maintenance of a particular feature given the right relative ranking. However, this ranking is unrelated to either the markedness of the feature value or the shape of the inventory, a relationship that seems too important to miss.

3.7 Conclusion

This chapter has laid out the basic proposals for this thesis. I have shown that a representational view of markedness allows an easy explanation for generalizations linking the markedness of a particular segment type, its behaviour with respect to different phonological processes, and the shape of the inventory in which it occurs. I have also addressed issues concerning the relevance of underspecified representations within OT, as it is maintained
throughout this thesis that such representations are not only compatible with the theory, but are crucial to its goals of explanatory adequacy. I have also discussed variability in markedness relations and shown that this variability is limited by the principles of the proposal. Finally, I have shown that the phonetic interpretation of underspecified representations may be seen as a reflection of universal preferences for auditory distinctiveness.
Chapter 4: The problem of constraint-ranking as explanation

A common claim in current Optimality Theory research is that all cross-linguistic variation is a result of language-particular rankings of output constraints. In this chapter, I argue that not all phonological patterning is explained in an interesting way with constraint ranking alone. Instead, considerations of input segmental representations and segmental contrast play a major part in the phonological alternations observed in a language. In encoding segmental contrasts structurally, markedness relations may be represented in terms of structural complexity and some of the burden of explanation is shifted from the ranking of different markedness and Faithfulness constraints to the segmental representations themselves.

From the discussion in Chapter 3 we know that markedness relations between segment types is subject to a certain degree of variability. In §4.1 I argue that this type of variability is not predicted by the fixed ranking of the markedness constraint hierarchy.

Another set of problems arise where we find invariability in markedness relations where a factorial typology makes such universals unexpected (§4.2). Finally, accounts based on constraint ranking fail to capture important generalizations that are non-local to the particular constraints involved. For example, a purely constraint-based account fails to capture the relationship between segmental patterns and inventory shape (§4.3).

4.1 Markedness variability and fixed rankings

As discussed in Chapter 2, most work on segmental markedness in Optimality Theory assumes that, within the set of featural markedness constraints, there are sub-hierarchies whose ranking is universally fixed (e.g. *Lab, Dors>*Cor). This fixed ranking is the explanation offered for cross-
linguistic generalizations regarding the consistent markedness relationships holding between different segment types. However, as pointed out in the previous chapter, it appears that markedness relations between segment types are to some extent universal and to some extent language-particular.

In this section, I argue that the fixed constraint hierarchies proposed by Prince and Smolensky (1993) and Lombardi (1995, 1998) account well for the fixed nature of segmental markedness relations, but fail to allow for the limited range of variation in markedness that is found across languages.

In contrast to this, a representational view of segmental markedness accounts for the variability that is found in markedness patterning. Importantly, the representational view that I have proposed provides a principled explanation for why the attested variability is not random but is limited and predictable.

P&S (1993) and Smolensky (1993) propose the universal ranking *dors, *lab >> *cor, encoding the fact that coronals are less marked than labials and dorsals with respect to assimilation, transparency, and epenthesis. However, as discussed in §3.4, coronals are not always the least marked segment types in a given language. Instead, glottals /ʔ/ and /h/ often appear to be the most unmarked segment type cross-linguistically.

I have presented an account of this variability which relies on the ranking of a single constraint, SEGPLACE, to determine whether coronal or a laryngeal will be the unmarked segment type in a given language (Causley (1998), Rice and Causley (1998)). The general preference for minimal structure is maintained, and the set of possible unmarked segment types is strictly limited.

As an alternative answer to this problem, Lombardi proposes an extended markedness constraint hierarchy. Following McCarthy (1994),
Lombardi assumes a common place specification, [pharyngeal], for guttural consonants (uvulars, glottals, and pharyngeals). The extended markedness constraint hierarchy she proposes includes a position for *phar at the bottom (right end), making pharyngeal consonants (and therefore glottal consonants) the least marked segment type.

(1)  *dors, lab >> *cor >> *phar

While the positioning of *phar at the bottom of the hierarchy accounts for the unmarked status of /ʔ/ and /h/, it raises the question of why the pharyngeal segments /h/ and /ʕ/ are some of the most marked consonants. Lombardi (1995) assumes the representations in (2) where /ʔ/ and /h/ are [+glottal] and /h/ and /ʕ/ are [-glottal]. The marked status of 'true pharyngeals' cross-linguistically is due to the cooccurrence constraint *[pharyngeal][−glottal], which is ranked independently of the hierarchy in (1).

(2)  ʔ, h  h, ʕ

Phar  Phar

[+glottal]  [−glottal]

Assuming a fixed hierarchy of substantive constraints comes with certain difficulties, despite Lombardi's suggested amendment. Let us examine Lombardi's proposal for an extended hierarchy, incorporating *[phar] as the lowest-ranked markedness constraint, as in (1). Since the inventory of phonemes in a language falls out of the ranking of Faithfulness relative to the markedness hierarchy, we expect pharyngeals to be implied in any phonemic inventory. The fact that the pharyngeals /h/ and /ʕ/ are in fact rare in inventories and behave as marked segments is accounted for by the cooccurrence constraint *[pharyngeal][−glottal]. Notice that, because /h/ and /ʕ/ are more rare and more marked than labials and dorsals, this
cooccurrence constraint must be ranked (universally) above *labial and
*dorsal, and therefore functioning as a markedness constraint within the
hierarchy. Thus, the extended hierarchy actually is as in (3).

(3) *[pharyngeal][-glottal] >> *lab, *dors >> *cor >> *phar

Thus, adding *[phar] to the hierarchy forces us to include a cooccurrence
restriction in the markedness hierarchy, classifying some pharyngeals as the
most marked segment type and some as the least marked segment type.¹

With [phar] as the least marked segment type, why do coronals ever
behave as unmarked? With respect to epenthesis, Lombardi argues that many
cases of coronal epenthesis involve issues of sonorancy: since neither /ʔ/
nor /h/ are sonorants, coronals are the least marked possible sonorant.
Therefore in cases where coda epenthesis inserts an n, l, or r, it is a reflection
of the preference for sonorous codas; coronal is the least marked place a
sonorant can have.

Yet, there are cases where sonority does not seem to be the explanation
for coronal epenthesis. For example, sonority should not be at issue in cases of
onset epenthesis, where sonorous segments may in fact be dispreferred.
Lombardi argues that all cases of onset epenthesis inserting coronal sonorants

¹This approach also sacrifices one of the advantages of the markedness hierarchy. Ideally,
cooccurrence constraints should fall out of the interaction of separate fixed hierarchies for
different feature classes. Thus, the cooccurrence of a particular feature F₁ with a marked
feature F₂ is worse than the cooccurrence of the same feature F₁ with a feature less marked than
F₂. For example, McCarthy (1994) and Ni Chiosián and Padgett (1997) have analysed coronal
transparency in certain spreading processes as coronal participation in spreading. Coronal
transparency refers to the spreading of a feature across a coronal segment while other segment
types (e.g. labials) block the spreading of that feature. In their analyses, the spreading feature
does spread to coronal segments (regardless of whether the feature is phonetically realized on
the coronal). The reason coronals are often the only segment type to participate in this way is
because they are unmarked. Thus, if a segment is going to have two place features associated
with it, it is better to have at least one unmarked place rather than two marked place features.

However, the cooccurrence restriction *[phar][-glottal] involves the least marked place
(pharyngeal). Of all the place types the feature [-glottal] could cooccur with, we might expect
pharyngeal to be the best, since according to Lombardi’s suggestion, it is the least marked place
feature.
and some cases of coronal obstruent epenthesis involve specific morphological contexts. Thus, it is a morphological requirement, not a question of markedness, that determines the nature of the epenthetic consonant.

This answer is unsatisfactory since the fact that coronal epenthesis may be restricted to specific morphological contexts does not explain why the choice for epenthetic consonant is coronal and not a more marked segment type such as labial. A morphological explanation which is independent of markedness considerations implies that any segment type may be selected for epenthesis if the morphology requires it.

Lombardi (1997) also addresses cases where coronal epenthesis seems to be driven by a phonological requirement that is seemingly independent of the morphology. One of the most well-known of such cases is in Axininca Campa where epenthetic /t/ is inserted between a vowel final stem and a vowel initial suffix (4).

(4) /i-N-koma-i/  iŋkomati  'he will paddle'
    /i-N-koma-aa-i/  iŋkomataati  'he will paddle again'

The explanation offered for why coronal /t/ is chosen over the less marked /ʔ/ is that Axininca Campa lacks /ʔ/ in its inventory. This lack is due to a high-ranking *ʔ constraint.

This suggestion is very problematic. According to Prince and Smolensky (1993), and Smolensky (1993) inventories fall out from the relative ranking of markedness constraints and Faithfulness. This is how we get the implicational relationships between segment types in an inventory. A constraint such as *ʔ appears to be a type of markedness constraint itself, although it rules out a particular segment rather than segment type. This constraint is ranked independently of the universal markedness hierarchy,
thus a particular language (such as Axininca Campa) may rank it very highly. Similarly, if there is a universal constraint *?, we might also expect constraints for each particular segment (e.g. *p, *s, *t).

An obvious problem arises with this approach: allowing additional markedness constraints such as these to be freely re-ranked with the markedness constraint hierarchy means that any type of inventory may be possible, and the universal ranking of the markedness constraints becomes meaningless. For example, languages are free to rank their *t, *n, *s constraints very highly, resulting in an inventory lacking coronal consonants. Alternatively, a language may choose to rank the *Place constraints in such a way as to completely reverse universal markedness relations. Thus in accepting such an approach, we abandon universality of implicational relationships and allow for unattested variability in inventories.

Aside from the coronal epenthesis cases Lombardi addresses which are (albeit unsatisfactorily) explained away as sonority-driven or morphologically specified, we still need an explanation for the unmarked status of coronals in other respects. For example, why is it so rare to find languages lacking coronals but less rare to find languages lacking /ʔ/ and /h/? Why should coronals be transparent to certain harmony processes (cf. Paradis and Prunet 1991, McCarthy 1995) to the exclusion of labials and dorsals? It is true that the constraint against coronal is lower ranked than the constraints against labial and dorsal. Yet, if coronal is simply another feature in the hierarchy, we might expect not only the "second least marked" segment to pattern with the unmarked class but also the third, fourth, etc.

The extended hierarchy fails to capture an important markedness generalization: both coronals and [+glottal] pharyngeals (/ʔ/ and /h/) are accorded special status with respect to markedness, each under different
conditions in different languages. A fixed hierarchy view of markedness offers no explanation as to why it is only these two place specifications that are special, and not some other arbitrary set of place specifications. What is needed is a way of accounting for the universal aspects of markedness while allowing for some principled variation between languages.

Again, under the representational view of markedness argued for here, allowing a limited degree variation is unproblematic. A structural view of segmental markedness allows us to define an unmarked set of segments (/t/ and /ʔ/) in a principled and constrained way, as discussed in Chapter 3. Unlike the extended constraint hierarchy view, this approach expects only two segment types to be unmarked and not some random class of segments. In addition, unlike the extended hierarchy view, the presence vs. absence of glottal stop in an inventory is not tied into the implicational relationships that govern the distribution of other places of articulation. Instead, placeless segments are ruled out by a constraint that is unrelated to *STRUCTURE (i.e. SECPLACE). In Chapter 3 I demonstrated how these two constraints together select which segment type will appear in TETU environments.

4.2 Unexpected Invarialility: the problem of unexplained fixed rankings

While the fixed ranking of the markedness constraint hierarchy fails to allow enough variation across languages, the possibility for free-ranking of other constraint types predicts too much cross-linguistic variation. In this section, I return to the cross-linguistic generalizations in assimilation patterns that were discussed in §3.2. Despite the fact that the relative ranking of independent constraints should be free, an output-constraint-only approach
to these assimilation patterns would have to recognize that the ranking of certain featural Faithfulness constraints is fixed across languages.

In contrast, I proposed in Chapter 3 that we account for fixed properties of segment types by assuming that their behaviour stems from their input representations. Following much work in segmental underspecification theories (e.g. Steriade (1987), Avery and Rice (1989), papers in Paradis and Prunet (1991), ) I argue that a straightforward account of these patterns comes from the absence vs. presence of input features, coupled with a general notion of Faithfulness, governing input-output identity.

This section touches on several questions relating to the issue of the role of output constraint ranking and input representations. First, I address the issue of Factorial Typologies and the generation of possible grammars. Next, I review generalizations associated with consonant place assimilation and outline the problem posed for a theory which predicts that such a generalization should not exist. I argue that a purely output-constraint based analysis of these patterns suffers from weaknesses that are avoided in an approach that makes use of minimally specified segmental representations in the input.

4.2.1 Factorial Typologies and free re-ranking

Aside from the markedness sub-hierarchies in which the relative ranking of the constraints are claimed to be universal, constraint ranking is language particular. One of the interesting predictions of this type of theory is that major differences between languages can be the result of a simple re-ranking of a few constraints. In principle, any constraint ranking is a potential grammar:

\[^2\text{Also, there is an apparent fixed ranking }\text{FAITH}_{\text{Root}}>>\text{FAITH}_{\text{Affix}},\text{ described in Beckman (1997).}\]
In Optimality Theory, a grammar of a language is a particular ranking of the constraints supplied by Universal Grammar. Permutation is therefore a crucial test of any proposed sub-theory of constraints: are all of the rankings of the constraints attested grammars, or at least possible ones?

McCarthy and Prince 1995: 327

Thus, given three different constraints A, B, and C, we have six different possible rankings. This set of possible rankings represents a factorial typology.

Much OT research has involved demonstrating for a particular ranking relationship between a pair of proposed constraints in one language that the reverse ranking is attested in another language. This ensures that, although a new constraint is proposed for inclusion in the universal constraint set, free re-ranking does not generate impossible grammars. However, where it appears that only one particular ranking of a set of constraints is found cross-linguistically, and a factorial typology generates unattested grammars, the ranking must be universally fixed. This leaves us with a theory involving both universal constraints and universal rankings, an apparent contradiction to the statements of McCarthy and Prince (1995) regarding constraint permutation.

The idea of language-particular ranking of constraints is that observed universal preferences should be reflected in the requirements of a universal constraint or constraint type rather than a particular ranking of constraints: since ranking varies across languages, the only cross-linguistic consistency (aside, perhaps from whatever properties are associated with GEN) is the constraints.

---

3Of course, respect for this principle alone is reason enough to abandon the fixed hierarchy view of segmental markedness: instead of encoding this universal tendency as a universal constraint, this hierarchy allows the ranking to be the universal component.
Related to this notion of universal preferences are the markedness-linked asymmetries in the phonological activity of different features. Recall from §3.2 that phonologically "active" features are those features that spread from one segment to another in assimilation and harmony processes, and those features that are maintained under coalescence of two segments. Phonologically "inactive" features are those features which fail to spread and are often lost in coalescence. We have seen that, in assimilation, the activity of a feature is often related to its markedness status: marked features tend to be "active" while unmarked features are "inactive" (Kiparsky 1985).

In general terms, this means that in consonant assimilation and coalescence patterns, the feature coronal behaves as inactive while the place features such as labials and dorsals are active. This is consistent with the generalization linking phonological activity of a feature to markedness: unmarked features are inactive while marked features are active.

The following section discusses the treatment of this generalization within a purely output-constraint based approach. I conclude this section by restating my position that since minimal input representations are not inconsistent with the framework and provide, I argue, a better account of the assimilation facts, their adoption is an attractive course to pursue.

4.2.2 Markedness and Phonological activity in Optimality Theory

Working within Optimality Theory, Smolensky (1993) addresses the generalization that there is a correlation between markedness of elements and their visibility in the phonology. The cross-linguistic pattern is that unmarked elements tend to be less visible and less active than marked elements. Analyses employing underspecification such as the one presented in Avery and Rice (1989) (see also papers in Paradis and Prunet (1991)) relate the invisibility of unmarked elements to the fact that these elements are
structurally distinct from marked elements in that they are underspecified. A rule cannot refer to a feature of a segment that is unspecified for that feature.

Smolensky (1993) argues that it is not the absence of unmarked elements that entails their inactivity in the phonology, but rather this inertness falls out from the unmarkedness of these elements. Since the notion of unmarkedness in OT translates into being more harmonic relative to something else, the generalization to be explained in OT terms is that more harmonic elements behave differently than less harmonic elements with respect to traditional “phonological processes”.

Smolensky discusses several scenarios with which underspecification and markedness have been associated. These include consonant epenthesis, redundancy effects and assimilation facts. The epenthesis facts fall out from the markedness constraints. The redundancy effects were discussed in terms of Itô, Mester, and Padgett (1995) in §3.3. The assimilation facts, however, Smolensky finds puzzling. As discussed in §3.2, given a sequence [Coronal] + [Labial], assimilation results in a string of two labial consonants and never two coronal consonants. This pattern, as Smolensky points out, seems to generate anti-harmonic forms involving marked segments instead of unmarked segments but he offers no solution within OT for this problem.

However, Kiparsky (1994) does address the issue of assimilation in Optimality terms. In his analysis, he attempts to account for the asymmetry in the assimilatory behaviour of marked vs. unmarked elements. Like Smolensky (1993), Prince and Smolensky (1993), and McCarthy (1994), Kiparsky assumes that it is the very fact that certain segments are unmarked

---

4Harmony refers to the degree of well-formedness of a form: more well-formed is more harmonic.

5At least, the choice of coronal over labial and dorsal in epenthesis falls out from the markedness hierarchy. However, as we saw in §4.1, this account faces difficulty in accounting for languages which choose glottals as epenthetic segments.
which motivates their unique patterning in the phonology. Specifically, he seeks to show how constraints on output forms capture the generalization that unmarked segments assimilate to marked segments, but not vice versa. As a rule, he argues, both the marked and the unmarked value of a feature may spread in a particular language, but if only one feature spreads, it is the marked one.

Kiparsky's suggestion for accounting for these facts within OT involves limiting the constraint vocabulary such that constraints cannot refer to unmarked feature values. Therefore, since coronal is the unmarked feature value for consonantal place, there is no constraint that refers specifically to coronal. Instead, he calls for two types of related constraints: specific constraints which refer to the marked member of a category and more general constraints that refer to the entire category itself. He argues that with this reworking of the constraint set, we are able to account for the assimilation asymmetries discussed above.

First, he shows that the markedness constraints as they stand cannot account for the assimilation facts. Kiparsky posits a constraint SPREAD which, as a general OCP constraint, rules out contiguous specifications for a feature type (e.g. two place features Coronal-Labial). This constraint should interact with other constraints to force the spreading of the marked feature. The markedness constraints *Lab and *Cor, however, wrongly predict preferential spreading of coronal, as shown in the tableau in (5). Recall that, since labial is more marked than coronal, *Labial is ranked above *Coronal.
The constraint \textit{SPREAD} is ranked over the two other constraints since it is clearly fulfilled in the output form in which both consonants share a place of articulation.\footnote{Note that I am abstracting away from directionality effects. The constraints and ranking given in this tableau will predict the same assimilation patterns regardless of whether the order of input features is coronal-labial or vice versa. Other constraints will come into play to limit directions of assimilation.} The problem is that given the constraint ranking based on the universal markedness hierarchy, the less-marked coronal feature should always be the one that spreads. The (c) candidate should be preferred over the (b) candidate since parsing a coronal feature should be better than parsing a labial feature. In reality, it is the (b) candidate that wins, forcing Kiparsky to conclude that these are not the relevant constraints to determine the pattern of assimilation. Instead, he suggests that the appropriate constraints are \textsc{Max}\footnote{In actual fact, Kiparksy's analysis pre-dates the adoption of Correspondence Theory within OT, therefore the Faithfulness constraint he uses here is \textsc{Parse}. However, for the sake of consistency throughout this paper, I substitute the Correspondence Theory constraint \textsc{Max} in discussing his proposal.} constraints from the faithfulness family of constraints. Using these constraints and referring only to marked features (i.e. non-coronal), we can

\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Candidates} & \textbf{SPREAD} & \textbf{*Lab} & \textbf{*Cor} \\
\hline
a. Rt Rt & *! & * & * \\
| cor lab & & & \\
\hline
b. Rt Rt & & ! & * \\
| \lab & & & \\
\hline
c. Rt Rt & & & * \\
| \cor & & & \\
\hline
\end{tabular}
establish a ranking which selects the correct output form. This ranking is
given in the tableau in (6).

(6)
Input: /...Rt  Rt.../

<table>
<thead>
<tr>
<th>Candidates</th>
<th>SPREAD</th>
<th>MAXlab</th>
<th>MAXplace</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Rt  Rt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. R   R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. R   R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first candidate is ruled out by SPREAD, since both place features are
preserved. The second and third candidates differ only in which feature is are
not preserved. Since MAXlab is higher ranked than the more general
constraint MAXplace, the second candidate in which coronal is unparsed
wins.

While Kiparsky's analysis does seem to account for the assimilation
facts without reference to underspecification, we should closely examine what
is encoded in the constraints that he proposes. His constraint vocabulary is
limited to allow two types of Faithfulness constraints: specific constraints
referring only to marked features, and more general constraints referring to a
class of features as a whole. The lack of constraints referring specifically to
unmarked elements has the effect of making unmarked elements invisible
except as a member of a feature class. This invisibility of unmarked elements
relative to marked elements has serious implications for Faithfulness. For
example, the MAX family of constraints ensures that input features are preserved in the output. However, given Kiparsky's constraint vocabulary, the MAX constraints are only able to refer to marked elements or the class as a whole. This means that the MAX family will have a bias for preserving marked elements. Why should markedness be encoded in Faithfulness in this way? What is it about marked elements that allows them special status with respect to MAX? Kiparsky seems to be building underspecification into his constraints: by stipulating that Faithfulness cannot see unmarked elements, he claims that those features have no status as members of the input. If MAX ignores unmarked features, how do we know those features are there in the first place?

My next question is whether an Optimality Theoretic analysis can account for assimilation without building underspecification into the constraints. That is, how can we get the same facts as Kiparsky's analysis without stipulating limitations on the constraint vocabulary? The crucial generalization that Kiparsky's constraints capture is a bias for faithfulness to marked elements. This is the generalization that puzzles Smolensky (1994): why do we see the preservation of marked features when their loss or replacement by unmarked features would result in a more harmonic form?

An initially attractive answer to this question is in terms of exploded Faithfulness constraints: if the constraint MAXfeature is broken down into a set of distinct constraints governing different features, this opens up the possibility for other constraints to interact separately with individual features. This is the approach taken by McCarthy (1993) in his analysis of certain vowel alternations in Arabian Bedouin Arabic. In this account, the ranking of

---

8Although the analysis changes somewhat in McCarthy (1994), he maintains the same exploded PARSEfeature constraint set to account for these facts.
Maxpharyngeal >> Maxhigh means that where a grammar highly-ranks a
constraint which restricts the parsing of V-place features, it is a worse offence
to fail to parse [pharyngeal] than to parse [high]. Thus high vowels are the
targets for a syncope process which does not target low vowels. This type of
asymmetrical treatment of V-place features bears a striking resemblance to the
assimilation patterns discussed above: assimilation targets coronal
consonants while it does not affect non-coronals. Returning to the coronal
case, and following the example of McCarthy (1993), we might assume a set of
constraints governing the preservation of consonantal place features such as
in (7) and (8).

(7)  maxcor
    The feature coronal if present in the input must be parsed in
    the output.

(8)  max-corr
    Non-coronal input place feature specifications must be parsed in
    the output.

The ranking of max-corr >> maxcor will mean that given a constraint forcing
assimilation of two adjacent consonants (such as Kiparsky's spread
constraint), a coronal will always assimilate to a non-coronal: since parsing a
non-coronal is more important that parsing a coronal feature, it will always
be the non-coronal feature that is shared and the coronal feature which goes
unparsed. Below I present a series of tableaus evaluating candidates for
different place assimilation environments. In (9), I show a candidate set for an
underlying coronal + labial sequence, in which assimilation will give a labial-
labial output.

---

9 The term [-coronal] is a shorthand for the set of non-coronal place specifications. Rather than
getting involved in a discussion of which place features to use, I use this as a cover specification
which includes labial and dorsal consonant places.
Since \textit{SPREAD} is highly ranked, the first candidate with two place specifications is ruled out. The other two candidates both have linked structures, but the third candidate fails to parse the labial feature and is ruled out by \textit{MAX\textsubscript{cor}}. The second candidate is successful since it violates only the lower-ranked \textit{MAX\textsubscript{cor}}. This tableau demonstrates the ranking \textit{SPREAD} >> \textit{MAX\textsubscript{cor}} and \textit{MAX\textsubscript{cor}} >> \textit{MAX\textsubscript{cor}}.

In (10) I give a tableau evaluating a candidate set for the input sequence \textit{dorsal} + \textit{labial}. In this example the assimilation fails, giving the output \textit{dorsal} + \textit{labial}.
This tableau establishes the ranking of $\text{MAX-cor} >> \text{SPREAD}$, since
SPREAD is violated in the successful output candidate (a). Notice that the other two candidates satisfy SPREAD but in doing so violate the constraint $\text{MAX-cor}$.

Thus, the correct assimilation patterns can be accounted for by exploding the $\text{MAX-feature}$ constraint into a series of specific constraints.

However, the asymmetrical treatment of unmarked features by the Faithfulness constraints still remains to be explained. Allowing separate $\text{MAX}$ constraints works, but the ranking will always be $\text{MAX-marked} >> \text{MAX-unmarked}$, since this is the cross-linguistic pattern of assimilation. Since these constraints cannot be re-ranked with respect to one another, it would be best for this fixed ranking to fall out from something other than stipulation.

Other fixed hierarchies in OT such as the markedness constraints $\text{*F}$, are based on a scale of widely motivated markedness generalizations. The ranking $\text{MAX-marked} >> \text{MAX-unmarked}$, however, is a ranking that is the opposite of these harmonic constraints. Thus not only is there no explanation for this
fixed ranking, this fixed hierarchy of feature-specific Faithfulness constraints is inexplicably the mirror image of the markedness hierarchy: the markedness hierarchy says that marked specifications are dispreferred, while the Faithfulness hierarchy says that their maintenance is particularly important.

As discussed in §4.2.1, constraint ranking is a property of an individual language, while constraints themselves are universal. Given that the tendency for marked elements to be "active" is universal, perhaps some effort should be put into finding an explanation for this pattern that lies outside constraint ranking.

Recall that in the account proposed in §3.4, the inactivity of unmarked features is linked to the absence of a specification for that feature underlyingly. The unmarked specification coronal is inactive because it is absent in the input representation. The interaction of representations with Faithfulness has the asymmetries fall out automatically. Assimilation of a more marked segment to a less marked one entails loss of input structure (MAX violations) while assimilation of less-marked to more marked involves no MAX violations. Since labials have more input structure than coronals, assimilation of a labial to a coronal is worse than assimilation of a coronal to a labial because there is more structure to map (be faithful to) in labials. Thus, coronal place of articulation will assimilate to a Dorsal or Labial consonant, even though the reverse pattern produces a more harmonic, less marked form. These assimilation patterns simply represent cases where Faithfulness works to keep the maximum input structure in the output, despite the violations of *STRUCTURE constraints.

The next section deals with the relationship between inventory and segmental patterning, and the problems associated with trying to capture this relationship with constraint ranking alone.
4.3 Constraint ranking and lost generalizations

As argued in §4.1, the variability in the relative markedness of different consonant place specifications is difficult to reconcile with a fixed constraint hierarchy approach to markedness. Parallel to the variability in consonant place markedness, markedness relations in vowel place specifications demonstrate a certain degree of variability cross-linguistically, and therefore pose a similar challenge to a fixed hierarchy view of markedness.

A further complication for this approach arises with the fact that much of this variability can be linked to the presence vs. absence of contrasts in an inventory. This generalization, I argue, is inexplicable if the phonology falls out entirely from the ranking of output constraints alone. This is because the constraints responsible for the markedness patterning of segments in particular processes are completely independent of the constraints determining the segmental inventory. Thus, a correlation between inventory and segmental patterning is unexpected and not easily captured.

The next two chapters probe more deeply into the question I lay out here: what is the relationship between inventory and segmental markedness? As the view espoused here is that markedness is evaluated in terms of representational complexity, this question becomes one of the relationship between inventory and segmental representations.

This section is structured as follows. First, I review from §3.4 the range of markedness variation that is seen in vowel place markedness, defining the set of possible unmarked vowel types. I show that some of this variation is linked to the presence or absence of place contrasts in the inventory. Then I present an account based on output constraints alone and demonstrate that such an account is bound to fail in capturing these inventory-based generalizations.
4.3.1 Variability in Markedness II: V-Place

In §3.4, we saw that languages differ with respect to which vowel place specification is the least marked. However, we also saw that the range of choices for unmarked place is not unrestricted. For example, languages do not appear to choose front rounded vowels as least marked. Instead, the choice seems to be limited to front unrounded, back, or central vowels. Furthermore, the choice does not appear to be completely free within this set, but is constrained by the shape of the inventory. In the languages discussed in §3.4, only those which lack a central vowel have a front or back vowel patterning as unmarked. When a language has a front-central-back contrast, the central vowel is chosen as the least marked (e.g. Rose (1993), Rice and Causley (1998)). Thus, the inventory is somehow linked to the markedness status of different vowel specifications. The next section explores how this link is captured under different views of markedness.

4.3.2 Markedness variability and inventory

Relating the markedness of a specification to the presence of contrasts in an inventory presents a different problem depending on the view of markedness that is assumed.

If markedness is structurally encoded, as argued here, then the differences in one language's treatment of /i/ as unmarked and another's treatment of it as bearing a marked specification has to do with differences in the representation of /i/. Marked specifications are more complex than unmarked specifications, therefore the complexity of /i/ is greater in one language than in another. Furthermore, if the markedness of /i/ is related to the inventory of the language, then the complexity of the representation is related to the inventory.
In §3.5 I outlined exactly how input representations are built in response to inventory shape as well as considerations of complexity. All aspects of markedness behaviour are derivable from these input segmental representations. As we saw in §3.4, since the markedness is represented in terms of structural complexity, a constraint such as *STRUCTURE will work to ensure the least complex segment type is inserted in epenthesis. The same considerations of complexity will also ensure that the least complex segment type is chosen as the unmarked fixed segment in reduplication.

As seen in §3.2, the relationship between phonological activity and markedness is easily captured under a representational view of markedness: marked segment types have specifications to spread and to be faithful to in the output. Unmarked segment types, on the other hand, lack specifications to spread, and have less structure to be faithful to in the output.

In contrast to this view is the view that input representations are unimportant in determining output candidates. Instead, all patterns arise out of the interaction of output constraints. In this section, I examine the generalizations regarding inventory shape and variability in the markedness of segment types from the perspective of a purely output constraint-based phonology.

4.3.2.1 Variability in emergence of the unmarked environments

As we saw in §3.4, languages differ in the type of vowel that appears in default contexts such as epenthesis and reduplication. Some languages prefer front vowels, others central vowels, and still others, back vowels. As discussed in §4.1, the emergence of the unmarked is usually taken to fall out from a fixed hierarchy of substantive markedness constraints. Clearly, given
the variability that is attested in the choice of unmarked vowels, a fixed markedness constraint hierarchy alone cannot be responsible.

One possibility is to allow the markedness constraints governing different vowel types to be freely re-ranked and to abandon the fixed hierarchy view. Thus, the relative ranking of the markedness constraints *u, *i, and *i will allow a language to choose one type of vowel to appear in default contexts, while a different ranking in another language will choose another. For example, ranking *u and *i over *i (as in (11)) means that /i/ will be chosen as epenthetic, as in Yawelmani. Ranking *u and *i over *i means that a language such as Malayalam will choose /i/ in default contexts ((12)), while, in other languages such as Fula, the ranking *i and *i over *u selects /u/ as a default ((13)).

(11) Yawelmani

<table>
<thead>
<tr>
<th></th>
<th>*u, *i</th>
<th>*i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i]</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[i]</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>[u]</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

Another possibility is to posit additional markedness constraints which are independent of the fixed hierarchy, parallel to the suggestions made by Lombardi (1997) for C-place markedness variability. However, as argued in §4.1, this type of proposal in which any segmental markedness constraint may be ranking independently of the fixed hierarchy counteracts the effect of the fixed hierarchy, and amounts to a free re-ranking.

11Representing the vowel-place markedness constraints as constraints on particular vowels is simply a convenient short-hand for the different vowel types (i.e. front, back, and central). The constraints should be interpreted as barring any vowel of a particular place (i.e. front, back, and central).
Thus, re-ranking of markedness constraints on vowel types allows for the predicted range of variation in the choice of epenthetic or default vowels. However, in the next sections I demonstrate that while the ranking of the markedness constraints governs the markedness status of vowel types in epenthesis and reduplication, it is irrelevant in determining the phonological activity of a vowel place specification in assimilation and harmony, despite the fact that this is another important dimension of segmental markedness.

4.3.2.2 Variability in target and trigger behaviour

Patterns of spreading (as in vowel harmony) and assimilation typically are taken to involve the interaction of two types of constraints: constraint(s) requiring the spreading of a particular feature, and some Faithfulness constraint(s) prohibiting the loss of input features.

Usually, where assimilation/spreading does occur in an output, it involves satisfaction of the constraint motivating the spreading at the

---

12Additional constraints, such as constraints on linked structures, etc., may also come into play in determining the optimal output, but these types of constraints are irrelevant to the present discussion of the relationship between segmental markedness and spreading.
expense of the Faithfulness constraint, since the spreading may have involved the loss of an input feature. This is represented schematically in (14).

(14) Input / X Y/  
     |   |  
     x y

The satisfaction of the higher-ranked constraint $\text{SPREAD}^{14}$ means (assuming that the features $x$ and $y$ are not compatible, being, for example, from the same feature class) that either $x$ or $y$ must spread, and displace the feature of the target ($y$ or $x$, respectively). The displacement of this input feature incurs a Faithfulness ($\text{MAXF}$) violation.

In the schematic case in (14), either the spreading of $x$ or the spreading of $y$ will satisfy the general $\text{SPREAD}$ constraint. However, as we know from the discussion of assimilation above, the spreading is not as symmetrical as this, and often only the spreading of one of $x$ and $y$ will be optimal. Also, there are languages where it appears that both $x$ and $y$ will spread, but only at the expense of a third element, $z$, and not at the expense of one another.

These patterns can still be accounted for by the output constraints $\text{SPREAD}$ and $\text{MAX}$, if these constraints are each broken down into a set of

$^{13}$There are cases of assimilation/spreading where an underspecified input representation remains intact in the output, for example where the target of the spreading gains a secondary articulation, but maintains its primary (input) specification, which may be underspecified.

$^{14}$The specific constraint motivating the assimilation/spreading could be an Alignment constraint (McCarthy and Prince (1994)), or a constraint requiring a feature be doubly-linked or licensed (c.f. Ito, Mester, and Padgett(1995)).
constraints governing individual features, and ranked separately. Then, a constraint $\text{SPREAD}_x$ can ensure the spreading of $x$ alone, as in (15).

(15) Input / $X$  $Y$/

\[
\begin{array}{c|c|c|c}
X & Y & \text{SPREAD}_x & \text{SPREAD}_y & \text{MAXF} \\
\hline
\ \ & \ \ & 1 & > & < \\
\hline
X & Y & \ & 1 & > & < \\
\ \ & \ \ & \ & x \ & \ & \\
X & Y & \ & 1 & > & < \\
\ \ & \ \ & \ & y \ & \ & \\
X & Y & \ & 1 & > & < \\
\ \ & \ \ & \ & x \ & \ & \\
\hline
x & y & \ \ & \ & \ & \\
\end{array}
\]

The ranking of $\text{SPREAD}_x >\text{SPREAD}_y$ means that the feature $x$ will spread in favour of $y$. The ranking of the $\text{SPREAD}_x$ constraint above the Faithfulness constraint ensures that spreading will occur in the output, and a non-spreading candidate such as the third one, will be ruled out.

An alternative to relying on feature-specific spreading constraints is to rely on feature-specific Faithfulness constraints (i.e. $\text{MAXF}$) to ensure that one feature spreads and the other does not. This is shown in the tableau in (16).

(16) Input / $X$  $Y$/

\[
\begin{array}{c|c|c|c}
X & Y & \text{SPREAD} & \text{MAX}_x & \text{MAX}_y \\
\hline
\ \ & \ \ & \ & 1 & > \\
\hline
X & Y & \ & 1 & > \\
\ \ & \ \ & \ & x \ & \ & \\
X & Y & \ & 1 & > \\
\ \ & \ \ & \ & y \ & \ & \\
X & Y & \ & 1 & > \\
\ \ & \ \ & \ & x \ & \ & \\
\hline
\end{array}
\]
The general SPREAD constraint ensures that the optimal candidate will definitely involve some sort of spreading, and rules out the last candidate, which does not involve spreading. The choice between spreading the feature $x$ or the feature $y$ is made by the feature-specific Faithfulness constraints. $\text{MAX}_x$ will prefer that $x$ spreads, since spreading of $y$ would involve the loss of $x$. $\text{MAX}_y$ prefers the opposite. The ranking of $\text{MAX}_x$ over $\text{MAX}_y$ ensures that the spreading feature is always $x$.

The constraints in the tableaux in (15) and (16) predict that spreading should occur regardless of the relative order of input elements. However, the spreading processes we see are often not bidirectional. Instead, the relative ordering of two elements is an important factor in determining the output. For example, in the Chumburung vowel harmony described in §3.4, the spreading only occurs from right to left. Recall that, in this process, the front vowel assimilates to a back vowel (repeated below as (17a)); a back vowel does not assimilate to a front vowel (17d).

(17)  (from Snider 1989)
a. $\text{ibv} \text{r}_1 + \text{kudu}$ $\text{iburu kudu}$ ten voices
d. $\text{jon}\_\text{v} + \text{we}$ $\text{jon}\_\text{v} + \text{we}$ dog’s mucus

In this type of situation, the ranking of the spreading and Faithfulness constraints must be such that the place specification of the back vowel (peripheral) is spread to the front vowel, and the input specification of the front vowel is left without an output correspondent. This spreading is unidirectional, so the SPREAD constraint must be directionally parameterized (e.g. $\text{SPREAD}_{\text{left}}$). Since spreading does not occur when the front vowel is on the right, we must assume that the $\text{MAX}_{\text{per}}$ constraint is higher-ranked than the conflicting SPREAD constraint. The tableaux in (18) and (19) demonstrate how
this ranking selects a spreading output candidate in (17a), and a non-
spreading candidate in (17d).

(18) Input / i  u /
    |   |
    [cor] [per]

\[
\begin{array}{|c|c|c|}
\hline
 & \text{Max}_{\text{per}} & \text{Spread}_{\text{left}} & \text{Max}_{\text{cor}} \\
\hline
\text{a} & u  u  \\
\text{\textbackslash}   &   &   & * \\
\text{\textbackslash} & [\text{cor}] & [\text{per}] &   \\
\hline
\text{i} & i  i  \\
\text{\textbackslash} &   &   & *! \\
\text{\textbackslash} & [\text{cor}] &   &   \\
\hline
\text{b} & u  i  \\
\text{\textbackslash} &   &   & * \\
\text{\textbackslash} & [\text{per}] & [\text{cor}] &   \\
\hline
\end{array}
\]

The first candidate above spreads the place feature of the back vowel leftward
to the first input vowel. This satisfies Spread\text{left}, but since the coronal
specification of the first vowel is displaced, Max\text{cor} is violated. This candidate
is preferred despite the Max\text{cor} violation because in the second candidate, the
feature peripheral fails to spread leftward, violating the high-ranked
Spread\text{left}.

(19) Input / u  i /
    |   |
    [per] [cor]

\[
\begin{array}{|c|c|c|}
\hline
 & \text{Max}_{\text{per}} & \text{Spread}_{\text{left}} & \text{Max}_{\text{cor}} \\
\hline
\text{i} & i  i  \\
\text{\textbackslash} &   &   & *! \\
\text{\textbackslash} & [\text{cor}] &   &   \\
\hline
\text{b} & u  i  \\
\text{\textbackslash} &   &   & * \\
\text{\textbackslash} & [\text{per}] & [\text{cor}] &   \\
\hline
\end{array}
\]

The first candidate in the above tableau spreads the feature coronal leftward,
satisfying Spread\text{left}. However, this spreading results in a displacement of the
peripheral feature of the first input vowel, violating the higher-ranked
constraint, Max\text{per}. Since the second candidate fails to spread, and satisfies
Max\text{per}, this candidate is selected as optimal.
To account for languages where front vowel specifications are active while back vowel specifications are inactive, the ranking of MAX\textsubscript{per} and MAX\textsubscript{cor} must be reversed. In this way, the satisfaction of the spreading constraint will be satisfied when the loss of a peripheral specification is at stake, but not when the loss of a coronal specification is at stake. Thus, only coronal will be active in spreading.

In languages where both coronal and peripheral are active (such as in Turkish), the Faithfulness constraints must be both ranked above the spreading constraint, so that spreading will occur everywhere except where it results in the loss of either a coronal or peripheral specification. Recall that, in the spreading patterns in the languages we have seen, this bi-featural spreading occurs where there are central vowels in the inventory. Central vowels serve as the targets for the spreading of coronal and peripheral.

(20) Input /u i/

\begin{center}
\begin{tabular}{|c|c|c|c|}
\hline
 & \text{MAX\textsubscript{per}} & \text{MAX\textsubscript{cor}} & \text{SPREAD} \\
\hline
\text{core} & u & / & ,\text{per} \\
\hline
\text{per} & u & / & ,\text{per} \\
\hline
\end{tabular}
\end{center}

With an input sequence involving a central vowel and a back/peripheral vowel, spreading of peripheral can occur, satisfying SPREAD, without the loss of any input place features, satisfying the MAXF constraints. Failure to spread, as in the second candidate, does not do any better with Faithfulness, but serves only to violate the SPREAD constraint.
An input sequence involving a front vowel and a central vowel is very similar to the case illustrated in (21) with a back-central input. A non-spreading candidate, such as the second in (22) violates SPREAD without motivation by a higher-ranked constraint since Faithfulness is satisfied in the spreading candidate.

With an input involving a front and back vowel, the importance of Faithfulness becomes apparent. Any spreading involves the loss of input features, violating one of the two high-ranking featural Faithfulness constraints. The only solution is to fail to spread, as in the last candidate, and satisfy Faithfulness.

<table>
<thead>
<tr>
<th>Maxper</th>
<th>Maxcor</th>
<th>Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Table" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(21) Input / i i/  

| ![Table](image.png) |

(22) Input / u i/  

| ![Table](image.png) |
4.3.2.3 Ranking of Faithfulness and relationship to inventory

As the previous section demonstrates, it is possible to construct a constraint-based account of the variability in the markedness patterning of the different segment types. The crucial constraints in determining choice of epenthetic segments are the featural markedness constraints: the re-ranking of the different constraints on vowel types varies the preferred vowel type in different languages. In spreading processes, the ranking relationship between a constraint such as the SPREAD constraint and the featural Faithfulness constraints MAXper and MAXcor, as well as between the ranking of MAXper and MAXcor. As long as SPREAD dominates one of the featural Faithfulness constraints, spreading of some feature will occur. Thus, where MAXper is ranked high, spreading of peripheral occurs, to the detriment of coronal specifications. The reverse ranking results in the reverse pattern, with coronal spreading to the detriment of peripheral. Where both MAXcor and MAXper are ranked above SPREAD, neither feature will be lost in the interest of spreading. This means that central vowels will be the target for spreading while coronal and peripheral vowels will not.

However, despite the fact that we have established a link between segment type patterning and the shape of the inventory, none of these rankings reflect these generalizations. Instead, the markedness characteristics of a particular segment type are determined by entirely separate constraints. Thus, the fact that in Yawelmani, the vowel /i/ is unmarked with respect to epenthesis and behaves as unmarked with respect to phonological activity fall out of different constraints. Furthermore, both of these characteristics are unrelated to the presence or absence of contrasts in the inventory. It is also purely coincidental in this account that the Turkish place specifications for
both front and back vowels are active, and that the language contrasts front, back, and central vowel places.

Since the constraints governing segmental markedness in its various guises are operating independently of one another, and their ranking relative is independent of one another, it is difficult to see how any generalization can be made relating (i) the markedness of a particular vowel throughout a language; or (ii) the relationship between the markedness behaviour of a particular vowel and the shape of the inventory. This is dealt with in more detail in the next chapter which explores the issue of inventory in OT and the relationship between constraints determining the inventory and the vowel patterning associated with that inventory.

4.4 Conclusion

This chapter has explored the question of explanation within Optimality Theory. Most current OT research focusses on output constraint ranking as the locus of explanation, while at the same time de-emphasizing the role of input representations in the grammar.

The arguments in this chapter bear on the ability of output constraints to capture important markedness generalizations. The markedness constraint hierarchy is too rigid to allow for the range of variation that is found in C-place markedness. However, allowing free re-ranking of the featural markedness constraints or allowing the inclusion of additional, independently-ranked markedness constraints overgenerates, and predicts far more than the attested variation which is in reality quite restricted in nature.

Phonological activity of features is also linked to markedness: marked features tend to be active, while unmarked features are inactive. In assimilation and coalescence this means that the marked feature is
maintained at the expense of the unmarked feature. Since this results in an anti-harmonic output, we know that the markedness constraint hierarchy cannot be what is driving the activity/inactivity of the different feature types. Instead, feature-specific markedness constraints are ranked in such a way as to preserve marked elements over unmarked elements. Although this type of analysis produces the correct patterns, it leaves unanswered the question of why markedness is linked to Faithfulness in this way.

A representational view of markedness allows for both types of markedness-related generalizations to be captured. The restricted variability in C-Place markedness results from two different constraints on representations, one of which is the general markedness constraint *STRUCTURE. Thus, markedness is seen as structural complexity, and is penalized by *STRUCTURE. The complexity of a representation, and therefore the markedness of a representation, is also linked to phonological activity. Marked elements are active because they have features to spread to other elements, and they are preserved under Faithfulness since their loss entails loss of input structure. The loss of unmarked elements entails a less serious violation of Faithfulness since they have less structure in the input.

Finally, we saw that variability in markedness of V-Place specifications can be linked to the number of contrasts in an inventory. A simple output constraint based approach to these patterns can account for the variability through re-ranking. However, it is unable to relate the re-ranking of constraints to a difference in inventory shape.

The next chapters involve close examinations of vowel coalescence patterns and how the correct input representations, built in response to considerations of contrast and complexity, can provide an explanation for the markedness variability in V-Place and its relationship to the inventory.
5. Coalescence, Markedness, and the Relevance of Inventory

Throughout the previous chapters, I have argued for the role of structural representations in determining phonological patterns and in the evaluation of the markedness status of different segment types. In this chapter, I illustrate through the examination of vowel place specifications in vowel coalescence patterns that segmental representations are constructed in response to considerations of contrast and complexity at the level of the input. This position stands in direct opposition to the current view assumed in most current Optimality Theory work which holds input representations to be universal and therefore powerless to affect the phonology in any language-particular manner. Contra Richness of the Base (Prince and Smolensky (1993), Smolensky (1996), see Chapter 4 for discussion) approaches to input inventories as universal, I argue that some notion of language-particular inventory must exist in the input, and the input inventory influences language-particular segmental representations.

The chapter is structured as follows. First, I demonstrate how the current proposal allows an explanation of a range of vowel coalescence patterns in different languages, while allowing certain inventory-based generalizations to be captured. In contrast to this approach, I show how a full-specification account along the lines discussed in Chapter 4 for the same facts ends up missing generalizations relating inventory shape to the coalescence patterns found in a language. Then I describe some of the difficulties that arise with determining inventories through output-constraint ranking, and conclude that languages must be able to differ in terms of their input inventories.
5.1 Coalescence and Input structure

I have previously shown how phonological activity, an indicator of featural markedness, can be linked to input representations. I have outlined how considerations of contrast and structure minimization work to derive segmental representations in the input. In this section, I test out these claims through examination of vocalic hiatus resolution patterns in different languages.

For each pattern, I will contrast two alternatives. One alternative, argued for in this thesis, relies on the representations of input segments: it is the structure of the segments involved which determines the output. The other alternative, espoused in the general OT literature, relies on feature-specific Faithfulness constraints. The relative ranking of these constraints drives the maintenance of certain features (and thus certain vowels) over others. I will show that only the first alternative allows us to capture the apparent relationship between the outcome of hiatus resolution and the number of contrasts in an inventory.

5.1.1 Preliminary Assumptions: A unified approach to Vowel Coalescence/Elision/Assimilation

Vocalic hiatus resolution gives rise to several different processes. The relevant processes here are vowel coalescence, assimilation, and elision. As Casali ((1996), (1997)) points out, in most languages, the outcome of vocalic hiatus resolution is determined by position (e.g., preserve V₁ over V₂, or assimilate an affix vowel to a stem vowel.). However, where positional factors are not at issue, the patterns depend on the nature of the segments involved. That is, when the coalescence of two particular vowels produces the same output regardless of their input order, the output must be
determined on the basis of the vowels themselves. When two vowels $X$ and $Y$ come together in any order and $Y$ is always deleted, it must be something about the nature of $X$ that keeps it from being deleted and something about the nature of $Y$ that prevents it from surfacing at the expense of $X$. It is these types of cases that are treated here. I summarize the relevant patterns schematically in (1).

<table>
<thead>
<tr>
<th>(1)</th>
<th>Assimilation</th>
<th>$X + Y, \ Y + X \rightarrow XX$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deletion/Elision</td>
<td>$X + Y, \ Y + X \rightarrow X$</td>
</tr>
<tr>
<td></td>
<td>Coalescence</td>
<td>$Rt + Rt, \ Rt + Rt \rightarrow Rt$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$X \ y \ y \ x \ x \ y$</td>
</tr>
</tbody>
</table>

These processes can be viewed as different instantiations of one process: feature spreading. In assimilation, all the relevant features of one vowel spread to an adjacent vowel, resulting in a single long vowel. Vowel elision may also be seen as assimilation, with deletion of one of the input morae. Coalescence involves the spreading of one or more features of one vowel, and deletion of the trigger vowel.

Importantly, all of these processes share a basic similarity which is that the outcome of these processes is determined by the features of the input segments. While they differ in certain respects, their differences are not relevant to the points to be made here. Thus, they are treated together here as one process that spreads and preserves certain features at the expense of others.

As in the previous chapter, phonological activity such as the spreading of a feature will be taken as evidence of its markedness. Thus unmarked
features are expected to be inactive and fail to spread and/or be preserved at
the expense of another more marked feature. In the following sections, I
examine feature-spreading processes in vocalic hiatus resolution which
demonstrate the role of input representations in the phonology. Further, I
will show that the interaction of inventory shape and considerations of
structural complexity determine the range of patterns that we find.

5.1.2 Front Vowels are Unmarked/ Back Vowels are marked

In §3.4, we saw that front vowels often behave as unmarked vowels. In
vowel coalescence patterns in many languages, they also appear to be
unmarked, and their place specification is not maintained in the output. Two
languages which clearly demonstrate the phonological inactivity of front
vowels in coalescence are Dogrib (Athapaskan), and Afar (Cushitic).

(2) Dogrib Vowel Inventory (Causley 1995)

<table>
<thead>
<tr>
<th>Oral</th>
<th>Nasal</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>ê</td>
</tr>
<tr>
<td></td>
<td>ð</td>
</tr>
<tr>
<td>a</td>
<td>ä</td>
</tr>
</tbody>
</table>

In Dogrib, a vowel-vowel sequence arising from affixation often
undergoes coalescence. Both the front vowels /i/ and /e/ will coalesce with
the back vowel /o/ for an output back vowel, regardless of the relative order
of the underlying vowels. In (a)-(d), the first vowel in the sequence is /o/ and
in (e) the second vowel in the sequence is /o/: all cases result in the output
/oo/. Since positional factors are not determining the quality of the output
vowel, it must be either the segmental content of the input segments or the
ranking of feature-specific Faithfulness constraints which are determining the
output.
a. go+idiy k'ala go'ddi you (sg) are still alive
b. go+ekw'od go'kw'od our jaw
c. ?or+i ?oo spruce bough
d. ?o+e ?o snare (cf. xöechi ‘snare support stick’)
e. ye+odzi yoodzi corner

Similar coalescence patterns are found in a language with a five-vowel inventory such as Afar (Cushitic). Like Dogrib, in Afar coalescence front vowels assimilate to non-front vowels. Rose (1993) gives the following vowel inventory and coalescence data for Afar:

(4) Afar Vowel inventory

<table>
<thead>
<tr>
<th>i</th>
<th>u</th>
<th>i:</th>
<th>u:</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>o</td>
<td>e:</td>
<td>o:</td>
</tr>
</tbody>
</table>

The data in (5) demonstrate that /e/ assimilates to /o/ regardless of their relative ordering. In (6), the vowel /i/ assimilates to both back vowels /o/ and /u/ when it precedes them. Notice that /u/ does not itself assimilate to the front vowel /i/ when it is the first vowel in the sequence.

(5) o+e, e+o→ oo (VV → V in closed syllables)

a. da'ro e'xe da'ro oxe I gave grain
b. diidaa'le oob'be diidaa'loob'be I heard a bee

(6) i-u→uu, i-o→oo, * i-u → ü

a. tamaa'ri urt'e tamaa'rutte student got well
b. 'rabbir-ow 'rabbow master
c. a'nu irgi'ce a'nu irgi'ce I cut

In both of these languages, the place features of the back vowels are active, being maintained in coalescence, while the place features of the front vowels are inactive, being lost under coalescence.
Under the contrast-based representational view developed in chapter 3, since there is only a front-back distinction to be made in the inventories of these languages, only one place needs to bear a specification. Since this is the minimum amount of place structure required to make the contrast, only one of either front or back bears a feature. The vowel coalescence/deletion patterns in these languages suggests that it is the back vowels that bear a place specification: the peripheral specification is always present on the output vowel of front V-back V coalescence. The phonological inactivity of the place feature of the front vowels falls out from the fact that these vowels have no place features in the input. This means that Faithfulness has no obligation to preserve front (i.e. coronal) place features, and the output vowel will clearly be peripheral. Thus, the input place representations are as in (7), below.

(7)  
<table>
<thead>
<tr>
<th>Front vowel</th>
<th>Back vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
<td>Place</td>
</tr>
<tr>
<td></td>
<td>Peripheral</td>
</tr>
</tbody>
</table>

This approach to the input representations allows an easy account of these coalescence facts in OT. If the back vowels have a specification for place and the front vowels do not, combining these features results in a back vowel. This is shown in the tableau in (8).
Since the back vowel has a feature in the input, **MAXF** will require that that feature will have an output correspondent as in the first candidate.

Under a non-representational full-specification account, the Faithfulness constraints must be called on to decide between keeping the features of the front vowel or the back vowel.

The ranking of **MAXper** over **MAXcor** gets the desired result: the output vowel will always carry the peripheral specification over the coronal one. It is thus possible to account for this type of vowel assimilation using feature-
specific Faithfulness constraints. However, as we will see in subsequent sections, this type of account will not allow us to capture generalizations about the relationship between inventory shape and the type of vowel assimilation patterns found in a language.

5.1.3 Back Vowels are Unmarked/Placeless; Front vowels are marked

As we saw in §3.4, back vowels in some languages pattern as though they are unmarked with respect to epenthesis, reduplication, and vowel harmony. In Japanese colloquial speech, back vowels behave as unmarked in assimilation.

The Japanese inventory is given in (10). In Japanese colloquial speech, a back vowel /o/ and a front vowel /i/ may coalesce to give a front vowel /e/. This is demonstrated in the forms in (11); data is from Newman (1997) and appears in Rice and Causley (1998).

(10) Japanese vowel inventory

```
i           u
   e          o
   a
```

(11) basic form          slang variant

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>osoi</td>
<td>ose:</td>
<td>slow, late</td>
</tr>
<tr>
<td>toroi</td>
<td>tore:</td>
<td>slow, stupid</td>
</tr>
<tr>
<td>amai</td>
<td>ame:</td>
<td>sweet</td>
</tr>
</tbody>
</table>

In this language, the front vowel appears to be the marked and active vowel type. The back vowel is inactive, and its place features are not maintained under coalescence. As in Dogrib and Afar, the inventories of these languages are such that only one of the two places (front and back) need receive a place specification to distinguish them. In contrast to Dogrib and
Afar, it appears as though the vowel type which receives a feature specification in these languages is the front vowel. Thus, the input vowel place representations posited for this type of language are as in (12). Recall that these are the same representations posited in §3.4 for other languages with unmarked back vowels such as Chamorro and Fula.

(12) Front vowel Back vowel
    Place Place
    |     |
    coronal coronal

The coalescence/elision patterns demonstrated in these languages fall out from the representations in (12) and the general Faithfulness constraint, MAXF, governing the preservation of input features. This is illustrated in the tableau in (13).

(13) Input /V₁ + V₂/
    [coronal]

<table>
<thead>
<tr>
<th>Candidates</th>
<th>MAXF</th>
<th>*STRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. V₁ V₂</td>
<td></td>
<td>*[cor]</td>
</tr>
<tr>
<td>\ /</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[cor]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. V₁ V₂</td>
<td>*! &lt;[cor]&gt;</td>
<td>*[per]</td>
</tr>
<tr>
<td>\ /</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[per]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first candidate preserves the coronal input specification, satisfying the Faithfulness constraint MAXF. Notice that both candidates here violate *STRUCTURE to the same degree. However, the second candidate incurs an unnecessary, and fatal, Faithfulness violation in not having an output correspondent for the input coronal specification.

Again, in an account that assumes full specification in the input, the ranking of feature-specific Faithfulness constraints determine the output. In
the case of an active coronal specification, these constraints need to be ranked $\text{MAX}_{\text{cor}} >> \text{MAX}_{\text{per}}$, as in the tableau in (14).

(14) Input $/V_1 + V_2/$

<table>
<thead>
<tr>
<th>Candidates</th>
<th>$\text{MAX}_{\text{cor}}$</th>
<th>$\text{MAX}_{\text{per}}$</th>
</tr>
</thead>
</table>
| 1. $V_1 \ V_2$
   \ /  
   [cor] | * | |
| 2. $V_1 \ V_2$
   \ /  
   [per] | *! | |

This particular ranking of the Faithfulness constraints means that coronal input specifications will be kept over peripheral ones. Therefore, the inactivity of the peripheral feature in vowel coalescence is linked to the lower ranking of the $\text{MAX}_{\text{per}}$ constraint.

5.1.4 Central Vowels are Unmarked/Placeless: Front and Back have place

Recall that in §3.4 we saw that where a language had a front-central-back contrast in the inventory, the central vowel was the inactive/unmarked vowel type. This also seems to be true in cases of vowel coalescence: when central vowels coalesce with front or back vowels, the output vowel is always front or back, respectively. In this section, we look at coalescence processes in two languages with central vowels, Chaha and Korean. These patterns illustrate the unmarked status of central vowels relative to front and back vowels.

The following inventory (15) is given for Chaha by Rose (1993).
(15) Chaha vowel inventory

\[
\begin{array}{ccc}
  & i & u \\
  e & \varepsilon & o \\
\end{array}
\]

The phonological activity of the different vowel places in vowel coalescence patterns is consistent with the findings of the chapter 4.

In (16)(a)-(d), the central vowels never contribute a place specification to the output vowel of the coalescence (data from Rose 1993)².

(16) a. dəwə-axe dəwaxə your relatives  
   b. yə-ef yef let him cover with a lid  
   c. yə-od yəd let him tell  
   d. aɔ’ə-u aɔ’o it is wood

In (17) we see that front vowels pattern with back vowels in resisting coalescence with a vowel differing in backness. Thus, instead of a front vowel coalescing with a back vowel, glide insertion applies to resolve the hiatus.

(17) a. mezə-axe mezəwaxe your five cents  
   b. k'əte-axe k'əteyaxe your footprint  
   c. c'əki-axe c'əkiyaxe your wooden pick  
   d. Turi-u ?Turiwu/Turiyu he is an expert  
   e. gəβəre-u gəβərewu/?gəβəreyu he is a farmer

Thus, the front and back vowels in Chaha behave as active or marked vowel types in coalescence, while central vowels are inactive or unmarked.

The central vowels in the Chaha system present an additional place contrast and therefore necessitate additional complexity in segmental representation to distinguish central from front and back vowels. Because there are three types of vowels to be distinguished, at least two of the vowel

²The high central vowel is epenthetic and therefore does not occur in a context where coalescence/vowel deletion might occur.
types must receive a specification. Thus, the minimal representations of the three vowel places are as in (18).

(18)  

<table>
<thead>
<tr>
<th>Place</th>
<th>Place</th>
<th>Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>coronal</td>
<td></td>
<td>Peripheral</td>
</tr>
</tbody>
</table>

If these are the input place representations, then an input sequence of a front vowel and a central vowel involves only a single place specification, as in (19).

(19)  

<table>
<thead>
<tr>
<th>Input /V₁ + V₂/</th>
</tr>
</thead>
<tbody>
<tr>
<td>[cor]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Candidates</th>
<th>MAXF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. V₁ V₂</td>
<td></td>
</tr>
<tr>
<td>\ /</td>
<td></td>
</tr>
<tr>
<td>[cor]</td>
<td></td>
</tr>
<tr>
<td>2. V₁ V₂</td>
<td>*&lt;[cor]&gt;</td>
</tr>
<tr>
<td>\ /</td>
<td></td>
</tr>
<tr>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

Since the front vowel has a coronal specification in the input and the central vowel has no place specification in the input (or in the output), the best output candidate will satisfy Faithfulness in preserving the coronal feature.

In the case of inputs with a front and back vowel, two place specifications are involved: coronal and peripheral. Thus, Faithfulness has the job of preserving two features in the output, as in (20).
The first two candidates successfully resolve hiatus, but only at the expense of losing two of the input features. This violates MAXF, the non-specific featural Faithfulness constraint, and therefore both are ruled out. The third candidate preserves both features in the resolution of hiatus through the insertion of a glide. Presumably, this incurs a violation of a constraint against glide-insertion, but this constraint is not ranked highly enough to rule this candidate out.

Notice that the ranking between the Faithfulness constraint and a constraint such as *GLIDE-INS does not have to change in Chaha from what it is in Dogrib, Afar, or Japanese. The preservation of input features is always highly ranked in all of these languages. Instead, it is within the input specifications themselves that the differences lie, and these differences are directly linked to the contrasts in the inventory.

---

3I have not tried to formalize exactly what type of a constraint it is that penalizes glide insertion. Under the view of markedness I am arguing for here, it is surely a constraint penalizing excess structure. However, I leave explicit characterization aside, as it is tangential to the present point.
Under the alternative view in which input representations are always identical and always fully specified, the patterning of the different vowel types is attributed entirely to constraint ranking. Since, in the case of a front vowel-back vowel input sequence, hiatus is resolved through glide insertion instead of coalescence, we know that Faithfulness to the place specifications of the front and back vowels must be very important. In particular, the Faithfulness constraints must be higher-ranked than in the languages seen in the previous two sections where either the front or back vowel place specification was lost. This ranking will block any coalescence/elision between front and back vowels, as demonstrated in the tableau in (22). The fact that central vowels readily undergo coalescence with either front or back vowels (21) presumably falls out from the fact that central vowels have no place features to be faithful to.

(21) Input /V₁ + V₂/

<table>
<thead>
<tr>
<th>Candidates</th>
<th>MAXper</th>
<th>MAXcor</th>
<th>*GLIDE-INS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. V₁ V₂</td>
<td>!&lt;per&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. V₁ V₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. V₁ G V₂</td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

In the first candidate, coalescence results in a central vowel in the output, and this incurs a fatal MAXper violation. The second candidate is the winner in incurring no Faithfulness violations while resolving hiatus and at the same time satisfying *GLIDE-INS. The last candidate, on the other hand, has a glide
in the output to resolve the hiatus. For this candidate evaluation, MAX\textsubscript{cor} and *GLIDE-INS do not need to be ranked with respect to one another: the correct candidate will be chosen regardless of their relative ranking. However, we can see from the following tableau that, for Chaha, MAX\textsubscript{cor} (in addition to MAX\textsubscript{per}) must be higher-ranked in this account.

(22) **Input** /V\textsubscript{1} + V\textsubscript{2}/

<table>
<thead>
<tr>
<th>Candidates</th>
<th>MAX\textsubscript{per}</th>
<th>MAX\textsubscript{cor}</th>
<th>*GLIDE-INS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. V\textsubscript{1} V\textsubscript{2}</td>
<td>(_____)</td>
<td>*!</td>
<td>(____)</td>
</tr>
<tr>
<td>(_)</td>
<td>(_)</td>
<td>(_)</td>
<td>(_)</td>
</tr>
<tr>
<td>[per]</td>
<td>[per]</td>
<td>[per]</td>
<td>[per]</td>
</tr>
<tr>
<td>2. V\textsubscript{1} V\textsubscript{2}</td>
<td>(_____)</td>
<td>*!</td>
<td>(____)</td>
</tr>
<tr>
<td>(____)</td>
<td>(___)</td>
<td>(___)</td>
<td>(___)</td>
</tr>
<tr>
<td>[cor]</td>
<td>[cor]</td>
<td>[cor]</td>
<td>[cor]</td>
</tr>
<tr>
<td>3. V\textsubscript{1} y V\textsubscript{2}</td>
<td>(_____)</td>
<td>(_____)</td>
<td>*</td>
</tr>
<tr>
<td>(_____)</td>
<td>(_____)</td>
<td>(_____)</td>
<td>(_____)</td>
</tr>
<tr>
<td>[cor]</td>
<td>[cor]</td>
<td>[cor]</td>
<td>[cor]</td>
</tr>
<tr>
<td>[per]</td>
<td>[per]</td>
<td>[per]</td>
<td>[per]</td>
</tr>
</tbody>
</table>

Coalescence does not take place between front and back vowels because both the Faithfulness constraints requiring the preservation of input features are highly-ranked. Since the input vowels both have place specifications in the input, neither of the candidates demonstrating coalescence (the second and third) will be chosen as optimal. Instead, glide insertion occurs in the output form, violating *GLIDE-INS.

While the high-ranking of both feature-specific Faithfulness constraints gets the correct pattern for this language, a crucial generalization is lost in this type of analysis: with the presence of central vowels in the inventory, Faithfulness to the features of both the front and the back vowels becomes important. Recall that in the patterns demonstrated by languages lacking central vowels, one of the Faithfulness constraints had to be low-
ranked enough to compel coalescence and loss of an input feature. In inventories with central vowels, the Faithfulness constraints must be promoted (other constraints demoted). What is the relationship between these ranking generalizations and the shape of the vowel inventory? In the next section I address this question and demonstrate that the correlation between vowel inventory and coalescence patterns is only expected under a contrast-based representational view of markedness.

The Korean vowel inventory, like Chaha, includes central vowels (23). The central vowels in this inventory also behave as though they are unmarked, being inactive in coalescence patterns.

(23) Korean vowel inventory (Sohn 1987)

```
i  i  u
e  a  o
æ  a
```

In the following data (from Rose (1993)), front vowels behave as marked segments relative to the central vowels. In (a)-(d), coalescence of a central vowel with a front vowel or glide gives a front output vowel. In (e), a central vowel coalesces with a back vowel to give a back output vowel.

(24) a. is-imyən iimyən to connect
    b. p‘yəm p‘əm cheek
    c. pe-ə pee to cut
    d. kæ-ə kæə to fold
    e. cu-ə co o to give

Unlike Chaha, the coalescence of a back and a front vowel or glide gives a non-structure preserving front rounded vowel. Again, this is not unexpected if the front vowels in this system bear their own place specification in the input.
(25)  

a. kwemul       kōmul       monster  
b. wisəŋ       üsəŋ       hypocrisy  
c. p'yoçok       p'öçok       sharp  

Parallel to the Chaha system, the central vowels in the Korean system present an additional place contrast and therefore necessitate additional complexity in segmental representation to distinguish central from front and back vowels. Therefore, the place representations for the different vowel types are identical to those given for Chaha in (26), repeated below.

(26)  

<table>
<thead>
<tr>
<th>Front vowel</th>
<th>Central vowel</th>
<th>Back vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
<td>Place</td>
<td>Place</td>
</tr>
<tr>
<td>coronal</td>
<td></td>
<td>peripheral</td>
</tr>
</tbody>
</table>

A structural account of the Korean patterns, with representations based on contrast, is straightforward: if a front segment bearing coronal is compelled to combine with a segment bearing peripheral, faithfulness to both of these features in the output gives a coronal-peripheral segment (provided that the complex output vowel is permitted).

While an account parallel to the structure-based account is available in a full-specification view of the input, the same problems arise of unexplained correlations between inventory and the ranking of apparently unrelated constraints.

5.1.4 Summary

In this section, I have shown how different vowel hiatus patterns can be made to fall out from input representations which are driven by considerations of minimizing complexity, while maintaining contrasts. In the next section, I explore the issue of inventory shape within Optimality Theory.
5.2 Where is the inventory?

The arguments of this thesis are based on a certain view of the segmental inventory of a language, namely, that we can speak meaningfully about a particular segment being within the set of permissible units in a language, and other segments being impermissible in a language. However, what has yet to be made explicit is whether inventories are a property of the input, or of the output, or both.

I will argue that we must have some notion of segmental inventories in the input. Since contrasts play an important role in the construction of input representations, I hold that we need some notion of phonemic contrast at the level of the input. This is in direct contrast to the view standardly assumed in current research in Optimality Theory, where inventory is explicitly held to be an output property, derived from output constraints (Prince and Smolensky (1993)). I suggest that, if the inventory is an effect of output constraints alone, then inventory-related generalizations must fall out from the same set of constraints that determine the output inventory. In §5.2.3, I demonstrate that this position cannot be maintained without loss of significant generalizations.

5.2.1 Richness of the Base and the universal input set

As discussed in §2.2, one of the most widely held assumptions in OT research is that all differences between languages arise from the ranking of output constraints.

It is contradictory, under this assumption, to talk about language-particular underlying or input inventories, or language particular input segmental representations. One of the fundamental claims of Prince and Smolensky (1993) is that segmental inventories fall out entirely from
constraint ranking: segment types which are not present in an output
inventory are a result of ranking of markedness over faithfulness. The set of
possible inputs to the grammar is universal: by *Richness of the Base* (ROB),
there are no constraints on input form. ROB is defined in (27).

(27) *Richness of the Base* (P&S 1993:191)

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.

Given *Richness of the Base* (e.g. Smolensky (1996)), there is no difference in
the nature of an input in language A and language B: all differences arise as a
result of the language particular ranking of constraints. Inputs, like
candidates, are universal.

The output inventory is determined by the relative ranking of
Faithfulness and structural/markedness constraints. As shown in (28), if
Faithfulness to a particular element \( \alpha \) is ranked above the constraint
prohibiting \( \alpha \), that element will be compelled by Faithfulness to appear in the
output if present in the input. If however, the constraint prohibiting \( \alpha \) is
ranked above Faithfulness, that element will not appear in the output and
therefore \( \alpha \) will not be part of the output inventory.

(28) Deriving the inventory:

| \( \text{FAITH}_\alpha \gg \ast \alpha \) | \( \alpha \) is found in the output inventory |
| \( \ast \alpha \gg \text{FAITH}_\alpha \) | \( \alpha \) is not found in the output inventory (given ROB, it could be in the input, but will never surface) |

In this way, the ranking of output constraints results in an output
inventory which is a subset of the possible inventory of sounds given the
universal set of inputs. Therefore, language-particular inventories only exist in the output.

5.2.2. The relevance of input inventories

This view makes serious predictions about how the inventory is determined in a language. We have seen in previous sections that the presence or absence of certain vowels in an inventory can determine the patterns of phonological activity that we find. Assuming that the inventory of a language is entirely determined by output constraints means that the patterns we find are not about the number and type of vowels found in a language; this is simply a reflection of the output constraint ranking. Instead, we need to look to the constraints that derive that output inventory, and hope to find explanations for the generalizations there. As mentioned, the constraints which are responsible for giving the output inventory are the markedness and Faithfulness constraints that govern features and their combination in segments.

In our discussion of vowel place patterning, we looked at two different types of inventories: ones with a front vowel and back vowel only, and ones with a front, central, and back vowel. Deriving these systems is a matter of finding the appropriate ranking of markedness and Faithfulness constraints. The markedness constraints on vowel place I will abbreviate as *u, *i, and *i.4

In languages with all three vowel types the ranking of the markedness constraints relative to Faithfulness must be FAITH >> *u, *i, *i. This means

---

4For the purposes of the present discussion, I do not assume these constraints to exist in a fixed sub-hierarchy, although many OT researchers have. The problem with proposing such a ranking is treated in Flemming (1995), Padgett (1997), Causley (1997b), and Rice and Causley (1998). Briefly, the difficulty lies in the positioning of /i/ relative to the other vowel types. In terms of implicational markedness, /i/ is more marked than either /i/ or /u/, and therefore *i should be ranked higher than *i and *u. However, in terms of phonological patterning and emergence of the unmarked situations, *i is often less marked than /i/ or /u/, thus implying the reverse ranking.
that the need to be faithful to inputs outweighs considerations of vowel-place markedness, and the output inventory contains all three vowel types. In languages where the inventory of vowel place types is limited to front and back vowels, the ranking must be $^i > FAITH > u, i$. This means that the markedness constraint barring /i/ outranks considerations of Faithfulness, so it is not found in the output of the language.

Now we return to the question of inventory-linked vowel place patterning. Recall that in systems with two vowel places (front and back), one vowel behaves as unmarked in a range of phonological processes. These patterns may be attributed to the effects of feature-specific Faithfulness constraints. In coalescence and assimilation patterns, where one feature (e.g. coronal) is lost instead of another feature (peripheral), we attribute this to the constraint ranking in (29). At points A or B is a constraint compelling the loss of a feature, such as a constraint motivating coalescence or feature spread.

(29) \[ FAITH_{per} >> FAITH_{cor} \]
\[ \uparrow \quad \uparrow \]
A \quad B

In systems with all three vowel place types, the central vowel behaves as the unmarked vowel type, while front and back vowels behave as marked. In assimilation and/or coalescence, this means that the features of both the front and back vowels are active in spreading, and/or maintained under coalescence, while central vowels are not. This change in the activity of vowel place types is attributed to the ranking in (30), where point A indicates the position of a coalescence/feature spreading constraint which motivates the feature-loss relative to the feature-specific Faithfulness constraints.

---

5As seen in §5.1, coalescence is usually motivated by a few constraints working in concert. For example, NOHIAIATUS will compel coalescence if coupled with the requirements of additional constraints which prohibit resolution by non-coalescence means (e.g. glide insertion). What is important here is that they occur in positions A or B.
If the inventory is a artifact of output constraint ranking, and the patterning of the vowel types is related to the inventory, then there should be a definite relationship between the inventory-determining constraint ranking, and the vowel-patterning constraint ranking. The inventory determining rankings are summarized in (31a), below. In (31b), the constraint rankings deriving the vowel patterning are summarized.

<table>
<thead>
<tr>
<th>(31a)</th>
<th>Inventory Type</th>
<th>Constraint Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>i - u</td>
<td>*i &gt;&gt; FAITH &gt;&gt; *u, *i.</td>
<td></td>
</tr>
<tr>
<td>i - i - u</td>
<td>FAITH &gt;&gt; *u, *i, *i.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(31b)</th>
<th>Inventory Type</th>
<th>Constraint Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>i - u</td>
<td>FAITHper &gt;&gt; FAITHcor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td></td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F-loss in these positions</td>
<td></td>
</tr>
<tr>
<td>i - i - u</td>
<td>FAITHper, FAITHcor &gt;&gt; F-loss</td>
<td></td>
</tr>
<tr>
<td></td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

There seems to be no obvious relationship between these two partial hierarchies. It is certainly not clear that there should be a relationship between the ranking of such disparate constraints as NOHIATUS or other constraints compelling coalescence or glide insertion, and the segmental markedness constraints. Similarly, in the previous chapter, we saw the same inexplicable ranking generalizations with the relative ranking of the spreading constraint SPREAD and the feature specific Faithfulness constraints. In two-place inventories, the constraint SPREAD dominated either FAITHper
or \textit{FAITH}_{cor}; in three-place inventories, both Faithfulness constraints dominated \textit{SPREAD}. Again, there is no reason why one ranking should be associated with a particular inventory type, since the relevant constraints are not the sole determiners of the vowel inventory. Thus, the link between the (output) inventory and the vowel place patterning generalizations is completely unexpected.

There is another problem with the Richness of the Base view of inputs that arises with certain cases of Faithfulness-derived segment types. Given ROB, the input inventory is the set of all possible sounds; the satisfaction of ranked output constraints will serve to reduce this universal inventory to give a language-particular inventory which is a subset of input possibilities. As stated in Smolensky (1996: 3):

\begin{quote}
The lexicon of a language is a sample from its inventory: all systematic properties of the lexicon thus arise indirectly from the grammar, which delimits the inventory from which the lexicon is drawn.
\end{quote}

A particular hierarchy of output constraints serves, in a sense, as a language-particular filter, characterizing the set of segments to be found in the output. However, there are many cases where the output (phonemic) inventory is in fact itself a subset of the inventory allowed by the constraint ranking.

For example, if a particular segment \( \alpha \) is absent in an output inventory, it is because the segmental markedness constraint \( ^*\alpha \) dominates the relevant Faithfulness constraint. Segmental markedness constraints can rule out entire classes of segments (e.g. \( ^*[\text{pharyngeal}] \)) or they may rule out the cooccurrence of particular feature-types. Thus, in a language \( L \), \( \ddot{u} \) is not present in a given output inventory because \( ^*\ddot{u} \) (a short form for the cooccurrence restriction \( ^*[\text{coronal-peripheral}] \)) is ranked above the Faithfulness constraints \( \text{MAX}_{\text{coronal}} \) and \( \text{MAX}_{\text{peripheral}} \), as in the tableau in (32).
This tableau looks complicated, but the interactions are quite simple. At the very right-end of the tableau are the markedness constraints ruling out single occurrences of coronal and peripheral (abbreviated as *i, *u). These are dominated by the Faithfulness constraints requiring that input instances of coronal and peripheral receive output correspondents. The relative ranking of these two sets of constraints means that, since /i/ and /u/ are part of the universal input set, the output inventory will contain both /i/ and /u/. However, the cooccurrence constraint *[per,cor] is ranked above the featural Faithfulness constraints. This means that a segment containing both a coronal and a peripheral specification will not be permitted in the output, despite the fact that failing to be faithful to either one of these features will result in a MAXF violation. Thus, a language with such a ranking will not have output occurrences of [ü]. If the ranking is such that the featural Faithfulness constraints outrank the constraint *[per, cor], then /ü/ will be a part of the output inventory of the language. The relationship between the constraint ranking and the output inventory is summarized below.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Output inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. *[per,cor]&gt;&gt;FAITHplace&gt;&gt;*Per,*Cor</td>
<td>i, u</td>
</tr>
<tr>
<td>b. FAITHplace&gt;&gt;*[per,cor],*Per,*Cor</td>
<td>i,u,ü</td>
</tr>
</tbody>
</table>

In this way, the constraint ranking determines the output inventory. However, under this view of the input a difficulty arises in cases where non-phonemic segments are derived through output constraints, and in particular, Faithfulness constraints. To take a case we have already seen, the
Korean coalescence patterns demonstrate that the markedness constraint *[coronal-peripheral] is violated in order to preserve the features coronal and peripheral and ü surfaces as the output vowel. On the basis of these patterns, we may infer that Korean must have the ranking given in (b) above: the vowel [ü] is part of the output inventory. However, the distribution of front-rounded vowels in Korean is entirely predictable: they occur in situations of coalescence but not elsewhere. How is this explained? If front-rounded vowels are allowed in the output, why should they not occur freely in morphemes?

Under a traditional definition of phoneme, there is a distinction to be made between input and output inventories: ü may be part of the output inventory but, not being contrastive in the language, it is not part of the input inventory. According to P&S (1993) the ranking of output constraints is meant to determine the inventories of languages and no difference in principle should exist between the input and output inventories. Thus in Korean, there is no principled reason why front rounded vowels are non-phonemic: they could as easily be included in inputs as /i/ or /u/. Instead, they only occur where "predictable", i.e. when compelled by the need to simultaneously satisfy Faithfulness and NOHIATUS.

This example is one of many that can be imagined where "derived" segments are present in the output but never function contrastively in the lexicon. Allophonic variation is a perfect place to look for more cases of inexplicably non-contrastive output segments. What is curious is that, since these non-phonemic segments are not ruled out by the output constraints, what prevents the learner from selecting (from the universal set of inputs) input forms that contain these segments? How do we explain, for example,
why Korean⁶ fails to select /CüC/ as an underlying stem form? Since such a form is no doubt contained within the universal constraint set, there is no reason for it to not be part of the input lexicon of Korean, and therefore no reason (by the output constraints) for it not to appear as a faithful output representation.

5.2.3 Phonemic inventories characterize the input

The view espoused in this thesis is that input inventories play an important role in the phonology, and are constructed on a language-particular basis. The learner, who constructs input representations based on these contrasts, must ascertain phonemic contrasts. Included in the input inventory are all segments which can be deemed contrastive.

Thus, the learner acquiring Korean will notice that [ü] occurs in predictable environments: it appears only in cases of coalescence where Faithfulness to /i/ and /u/ are at stake. If [ü] always occurs predictably, then it does not need to be distinguished from other vowels in the input, and therefore it is not part of the contrastive inventory in the input. Where other segments occur in non-predictable environments, they are presumed to be part of the input inventory, and therefore need to be distinguished representationally from other contrastive segments. This is done according to the principles outlined in §3.5.

⁶McCarthy (1996) discusses a parallel case in Rotuman: coalescence results in output vowels not present in the phonemic inventory. Such patterns pose similar problems for the position that inventories derive from output constraints alone.
5.3 Conclusion

This chapter has tested the proposal for determining segmental representations based on considerations of contrast and minimization of representational complexity. I have shown that such a proposal allows an explanation of a range of vowel coalescence patterns in different languages, while allowing certain inventory-based generalizations to be captured. In contrast to this approach, I have demonstrated that a full-specification account of the same facts ends up missing generalizations relating inventory shape to the coalescence patterns found in a language. Thus, we must have some notion of language-particular in the input, and this input inventory must be allowed to influence language-particular segmental representations. Finally, I have described some of the difficulties that arise with determining inventories through output-constraint ranking, and concluded that languages must be able to differ in terms of their input inventories.
6. Height Coalescence

In his survey of vocalic hiatus resolution in dozens of languages, Casali (1996) finds an interesting correlation between inventory shape and height coalescence patterns in languages. Parallel to the correlation between place contrasts and coalescence patterns seen in Chapter 5, Casali finds that the coalescence patterns demonstrated in a language are related to the number of height distinctions in the vowel inventory.

In this section, I describe the relevant coalescence patterns and the types of inventories associated with them. Then I present Casali's (1995) proposal to account for this relationship. I argue that although his proposal is designed to account for the relationship between inventory type and coalescence pattern, it fails to account for certain natural class behaviour relating to inventory shape. Instead, I argue, his proposed constraints and rankings amount to a simple description of the generalizations, and leave unanswered the basic question of why inventory type and segmental behaviour are so closely connected.

Next, I demonstrate that inventory-driven segmental representations predict not only that links between inventory and coalescence patterns should arise, but also make predictions about what type of patterns will be associated with which types of inventories. Making use of some suggestions for inventory-driven height representations from Goad (1993), and Rice and Avery (1993), I argue that these inventory shapes (and therefore, vowel height representations) can be made to fall out from interacting considerations of complexity.

Finally, I argue that the height system of a language cannot be derived through output-constraint ranking, and instead languages must be able to differ in terms of their input inventories.
6.1 The Patterns

Casali contrasts two types of height coalescence patterns which he calls "\(e\)-coalescence" and "\(\varepsilon\)-coalescence". Most typically, in languages with "\(e\)-coalescence," an /a + i/ sequence is resolved as \(e\), and/or an /a + u/ sequence is resolved as \(o\). In the languages surveyed (which include such languages as Gichode, Krachi, Nawuri, and Chumburung) some or all of the following patterns are also found.\(^1\)

\[
\begin{align*}
\text{(1)} & \quad a + i > e & a + u > o \\
& \quad e + i > e & e + u > (y)o \\
& \quad o + i > (w)e & o + u > o
\end{align*}
\]

Descriptively speaking, the featural output of this type of coalescence is to some extent positionally determined so that the output of the coalescence preserves the "non-high" feature\(^2\) of the first vowel and the place feature of the second vowel.

Languages with "\(\varepsilon\)-coalescence" demonstrate a different set of patterns with the non-high vowels (2), and commonly display some or all of the patterns involving high vowels (3). The list of languages in the survey which demonstrate \(\varepsilon\)-coalescence includes: Anuf\(\bar{\text{u}}\), Bolia, Central Kambari, Dangme, Edo (Bini), Efik, Ewe, Nkengo, Owon Afa (Casali 1996:93). From the vowel coalescence patterns in Kisi described in Childs (1996) we could also add Kisi to the list of three-height systems with \(\varepsilon\)-coalescence.

\[
\begin{align*}
\text{(2)} & \quad a + e > \varepsilon & a + u > \partial \\
& \quad \varepsilon + e > \varepsilon & \varepsilon + o > (y)\partial \\
& \quad \partial + e > (w)e & \partial + o > \partial \\
\text{(3)} & \quad a + i > e & a + u > \partial \\
& \quad \varepsilon + i > \varepsilon & \varepsilon + u > (y)\partial \\
& \quad \partial + i > (w)e & \partial + u > \partial
\end{align*}
\]

\(^1\)Glides may or may not be part of the output. Languages with larger vowel inventories have more possible combinations. This is discussed below.

\(^2\)I will be discussing the precise feature in subsequent sections.
Again, there are positional factors which determine the output of the coalescence: the backness/roundness of the output vowel is determined by the place feature of the second input vowel. Also, like the patterns in (1), these patterns could be described as the preservation of the [-high] feature of the first vowel.

Interestingly, Casali notes, whether a language has $e$-coalescence or $\epsilon$-coalescence is predictable from its vowel inventory. The discussion in the next two sections demonstrate that languages with three height distinctions in the front and back vowels (e.g. $i/e/\epsilon$ and $i/1/\epsilon$) display $\epsilon$-coalescence while languages with four height distinctions (e.g. $i/1/e/\epsilon$) demonstrate $e$-coalescence.

6.1.1 Three height systems and $\epsilon$-coalescence

Casali's survey reveals that $\epsilon$-coalescence is found in languages demonstrating non-positionally determined height coalescence (c.f. §5.1.1) whose height systems are as in (4). Casali identifies the system in (4a) as a Yoruba-type language, while the system in (4b) is a Kikuyu-type language.

\[
\begin{align*}
4a. & \quad i & ( & u & b. & \quad i & ( & u \\
& \quad e & ( & o & \quad & i & ( & u \\
& \quad \epsilon & \quad \epsilon & \quad \epsilon & \quad \epsilon \\
& \quad a & \quad a
\end{align*}
\]

To illustrate the patterns associated with $\epsilon$-coalescence in a Yoruba-type language, data from Anufa is given in (6). Casali gives the following vowel inventory for Anufa. In describing these patterns in pre-theoreticial terms, I assume full specification of vowels in binary-valued features, as shown in (5).

---

The presence or absence of central vowels is irrelevant in determining height coalescence patterns. The parentheses in the inventories indicate that central vowels may or may not be present in the inventory.
The issue of precise featural characterization of the vowels is the subject of §6.4-6.6.

(5)

<table>
<thead>
<tr>
<th>[+high]</th>
<th>[-back]</th>
<th>u</th>
<th>[+back]</th>
<th>[+]atr</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>e</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[-high]</td>
<td>e</td>
<td>a</td>
<td>[-low]</td>
<td></td>
</tr>
</tbody>
</table>

(6) Anufc (Casali 1996:120, citing Adjekum, Holman, and Holman 1993)

<table>
<thead>
<tr>
<th>Underlying</th>
<th>Surface</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ye-i</td>
<td>ye:</td>
<td>raise it</td>
</tr>
<tr>
<td>b. jo-u</td>
<td>jo:</td>
<td>cool you</td>
</tr>
<tr>
<td>c. bo-u</td>
<td>bo:</td>
<td>beat you</td>
</tr>
<tr>
<td>d. bu-i</td>
<td>bwi:</td>
<td>break it</td>
</tr>
<tr>
<td>e. so-i</td>
<td>swe:</td>
<td>carry it</td>
</tr>
<tr>
<td>f. bo-i</td>
<td>bwe:</td>
<td>beat it</td>
</tr>
<tr>
<td>g. fa-i</td>
<td>fe:</td>
<td>take it</td>
</tr>
<tr>
<td>h. fa-u</td>
<td>fo:</td>
<td>take you</td>
</tr>
<tr>
<td>i. n-de-u</td>
<td>ndo:</td>
<td>I will take it from you</td>
</tr>
</tbody>
</table>

Notice that in cases where the first vowel is /a/, /ɛ/, or /ɔ/ the output vowel is either /ɛ/ or /ɔ/, depending on whether the second vowel is front or back. The non-high specification of the first vowel is maintained, while the place features of the second vowel are maintained, as seen in (i) where the mid-front vowel/high back vowel sequence results in a mid-back output vowel. Also, the [-atr] value of the first vowel is maintained so that if the first vowel is [-atr] as in (a) and (b), the output vowel is [-atr]. Two generalizations are important for this discussion. First, it is the [-atr] value and not the [+]atr value that is maintained or “active” in the coalescence. Second, the low vowel /a/ patterns as a [-atr] vowel along with /ɛ/ and /ɔ/.
Another language, Owon Afa, provides word-juncture coalescence examples with a full range of vowels. The inventory of Owon Afa is the same as the one given for Anufọ, in (5).

(7) Owon Afa (Casali 1996: 120, citing Awobuluyi 1972)

<table>
<thead>
<tr>
<th>Underlying</th>
<th>Surface⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ri eme</td>
<td>reme</td>
</tr>
<tr>
<td>b. ewo oba</td>
<td>ewo:ba</td>
</tr>
<tr>
<td>c. te esi</td>
<td>te:si</td>
</tr>
<tr>
<td>d. bu iwe</td>
<td>bi:we</td>
</tr>
<tr>
<td>e. ke ibo</td>
<td>kebo</td>
</tr>
<tr>
<td>f. ewo iwe</td>
<td>ewe:we</td>
</tr>
<tr>
<td>g. dɔ iwe</td>
<td>dɛwe</td>
</tr>
<tr>
<td>h. da iwe</td>
<td>dɛwe</td>
</tr>
<tr>
<td>i. da opu</td>
<td>dɔ:pu</td>
</tr>
<tr>
<td>j. da ehwe</td>
<td>dɛhwe</td>
</tr>
<tr>
<td>k. da uju</td>
<td>dɔju</td>
</tr>
<tr>
<td>l. ri aka</td>
<td>ra:ka</td>
</tr>
</tbody>
</table>

In this language, if either of the input vowels is non-high the output vowel will be non-high. Therefore in (7a), a high vowel + a mid vowel results in a mid vowel, and in (7k) a low vowel + a high vowel results in a mid vowel. If either of the input vowels are [-atr], the output vowel will be [-atr]. Thus, in (7b) an [+ atr] vowel followed by a [-atr] vowel results in a [-atr] output, and in (7c) a [-atr] vowel followed by a [+atr] vowel results in a [-atr] output. Place is determined by the second input vowel. As in Anufọ, the persevering or active [atr] value is [-atr] and not [+atr]. Also, the vowels /a, e, and ɔ/ behave as a natural class of [-atr] non-high vowels.

⁴Note that in Owon Afa, length is only preserved in some cases of coalescence. Casali does not describe what determines the preservation or loss of vowel length.
6.1.2 Four height systems and e-coalescence

In languages with a height system as in (8), e-coalescence is the typical pattern in non-positional height coalescence.\(^5\) Again, I assume for the present discussion that featural specifications are binary and fully specified, as in (8).

(8)\[
\begin{array}{c|c|c|c|c|c}
\text{[+high]} & \text{[-back]} & \text{[+back]} & \text{[+atr]} & \text{[-atr]} \\
\hline
\text{i} & \text{u} & \text{u} & \text{[+low]} \\
\text{e} & \text{o} & \text{[+low]} \\
\text{e} & \text{a} & \text{[+low]} \\
\\end{array}
\]

The differences between e-coalescence and e-coalescence patterns are illustrated with the data from Gichode in (15). The inventory Casali gives for Gichode is the same as the one above in (8).

(9) Gichode (Casali 1996:135, citing fieldnotes of Keith Snider)\(^6\)

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Actual segments</th>
<th>Underlying</th>
<th>Surface</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. high + mid → mid</td>
<td>i + o → o</td>
<td>okuli otu</td>
<td>okulotu</td>
<td>husband’s heart</td>
</tr>
<tr>
<td></td>
<td>i + i → o</td>
<td>fuli ölə</td>
<td>fulölə</td>
<td>deer’s sore</td>
</tr>
<tr>
<td></td>
<td>u + o → o</td>
<td>dobobu ọsịnj</td>
<td>dobọsịnj</td>
<td>farmer’s vein</td>
</tr>
<tr>
<td>b. mid + high → mid</td>
<td>o + i → o</td>
<td>ọgono ịfịn</td>
<td>ọgọne ịfịn</td>
<td>dog’s veins</td>
</tr>
<tr>
<td></td>
<td>o + i → e</td>
<td>ọgono ịb</td>
<td>ọgọne ịb</td>
<td>dog’s sores</td>
</tr>
<tr>
<td></td>
<td>e + i → e</td>
<td>atanatịse ịfịn</td>
<td>atanatịse ịfịn</td>
<td>female twin’s veins</td>
</tr>
<tr>
<td>c. front + back → back</td>
<td>e + o → o</td>
<td>gibide otu</td>
<td>gibidotu</td>
<td>slave’s heart</td>
</tr>
<tr>
<td></td>
<td>i → o → o</td>
<td>okuli otu</td>
<td>okulotu</td>
<td>husband’s heart</td>
</tr>
<tr>
<td>d. back + front → front</td>
<td>o + i → e</td>
<td>ọgono ịfịn</td>
<td>ọgọne ịfịn</td>
<td>dog’s veins</td>
</tr>
<tr>
<td>OR → glide + front</td>
<td>u + i → wi</td>
<td>dobobu ịfịn</td>
<td>dobọwịfịn</td>
<td>farmer’s veins</td>
</tr>
<tr>
<td></td>
<td>u + i → wi</td>
<td>obilimbu ịdọ</td>
<td>obilimbwịdọ</td>
<td>young woman’s yams</td>
</tr>
</tbody>
</table>

\(^5\)The languages in Casali’s survey which fit this description include Chumburung, Gichode, Krachi, Nawuri, S. Sotho (Casali 1996:94).

\(^6\)Casali notes that there are phonetic details which are irrelevant to the coalescence patterns which have been left out.
Let us examine the coalescence patterns in the examples in (9a)-(9b).

From the data in (9a) we notice that a sequence of a high vowel + mid vowel results in a mid vowel. Also, in the example in (9b) a mid vowel + high vowel sequence results in a mid vowel. Thus, the height of the output vowel is determined by the non-high vowel.

In comparing (9c) and (9d), we find that the place specification of the output vowel is determined by the second vowel. Thus, in the forms in (9c) a front-back input sequence results in a back vowel, and in (9d) a back-front input sequence results in a front vowel.

Finally, the most important thing to notice is the [atr] specification of the output vowels. If either of the two input vowels is [+atr], the output vowel is [+atr]. Thus, in the examples in (9e) the first input vowel is [+atr] and the second is [-atr], and the output vowel is [+atr]. In (9f), the first input vowel is [-atr] and the second vowel is [+atr], giving a [+atr] output. The relative order of the [atr] specifications is therefore irrelevant: [+atr] is always preserved over [-atr]. This is unlike the patterns seen with the e-coalescence languages, where [-atr] was the "active" value and [-atr] determined the [atr]
specification of the output. Here, [+atr] is the specification that is maintained and therefore active. A [-atr] output vowel only surfaces in cases where both input vowels are [-atr], as in the forms in (9h).

Now consider the following additional Gichode coalescence data involving low vowels.

(10)

- a. \( a + [+atr] \)  
  \( a + i \rightarrow e \)  
  \( a + o \rightarrow o \)  
  \( \text{diga id}\)\(\text{g}\) \(\text{diga otu} \) \(\text{dige}\)\(\text{g}\)  
  young man’s yams  
  young man’s heart

- b. \([-atr] + a \)  
  \( i + a \rightarrow a \)  
  \( \text{okuli ansid}\)\(\text{c} \) \(\text{okulansid}\)\(\text{c} \)  
  husband’s face

- c. \( a + [-atr] \)  
  \( a + o \rightarrow o \)  
  \( \text{oka cb} \) \(\text{ok}\)\(\text{b} \)  
  wife’s sore

- d. \([-atr] + a \)  
  \( e + a \rightarrow a \)  
  \( \text{atanatfise ansid}\)\(\text{c} \) \(\text{atanatfisansid}\)\(\text{c} \)  
  female twin’s face

- e. \( a + \text{mid} \)  
  \( a + o \rightarrow o \)  
  \( \text{diga otu} \) \(\text{digo tu} \)  
  young man’s heart

- f. \( \text{mid} + a \)  
  \( e + a \rightarrow a \)  
  \( \text{atanatfise ansid}\)\(\text{c} \) \(\text{atanatfisansid}\)\(\text{c} \)  
  female twin’s face

- g. \( a + \text{high} \)  
  \( a + i \rightarrow e \)  
  \( \text{diga id}\)\(\text{g} \) \(\text{dige}\)\(\text{g} \)  
  young man’s yams

- h. \( \text{high} + a \)  
  \( i + a \rightarrow a \)  
  \( \text{okuli ansid}\)\(\text{c} \) \(\text{okulansid}\)\(\text{c} \)  
  husband’s face

  \( u + a \rightarrow wa \)  
  \( \text{obilimbu atc} \) \(\text{obilimbwatac} \)  
  young woman’s things

As in the coalescence patterns discussed in (9), positional factors largely determine the place of the output vowel. With the low vowel /a/, it also appears that positional factors determine the height of the output vowel. Thus, in all cases where /a/ is the second vowel, the output vowel is [a]. Notice that, in contrast to the Anufc and Owon Afa patterns, /a/ does not form a natural class with /ε/ and /ɔ/ with respect to determining the features of the output vowel. In (10a)-(10d), /a/ does not seem to have any effect on the [atr] value of the output vowel: /a/ preceded by a [-atr] vowel results in [a], as in (d). /a/ followed by the [-atr] vowel /ɔ/ gives a [-atr] vowel as in (10c). This means that [a] cannot be [+atr], since an input sequence involving a [+atr] vowel always gives a [+atr] output, as we saw above in (9). However, when
/a/ is preceded by a [+atr] vowel in the input as in (10b), the output vowel is
/a/, no different from the output of an [-atr] vowel + /a/ sequence as in (c). So
/a/ does not seem either to be affected by the presence of [atr] in the input, nor
does it affect the [atr] outcome of an input sequence. Thus, /a/ appears to be
neutral with respect to [atr] in these systems.

Compare this with the parallel Owon Afa example from (7h) and (7i),
repeated below.

(7)  h.  /da iwe/  [dewe]         buy book
    i.  /da opu/  [dɔ:pu]         buy dog

These data suggest that in contrast to the Gichode system, in this type of
system /a/ is [-atr] in the input.

To summarize briefly at this point, the type of height coalescence
patterns seen in a particular language can be predicted from the inventory. In
three height systems such as Anufọ and Owon Afa, the [-atr] value of the
non-high vowels is preserved over [+atr], while in four-height systems, [+atr]
is preserved over [-atr]. Also, in three height systems, /a/ patterns with the
[-atr] non-high vowels /ε/ and /ɔ/ in determining the quality of the output
vowel. In four-height systems, /ε/ and /ɔ/ alone pattern together without
/a/. These generalizations are summarized in the table in (11).

<table>
<thead>
<tr>
<th>Inventory type</th>
<th>[atr] Patterns</th>
<th>Natural class</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 height</td>
<td>i</td>
<td>![Pattern 3 height]</td>
</tr>
<tr>
<td>i</td>
<td>![Pattern 3 height]</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>![Pattern 3 height]</td>
<td></td>
</tr>
<tr>
<td>ε</td>
<td>![Pattern 3 height]</td>
<td></td>
</tr>
<tr>
<td>4 height</td>
<td>i</td>
<td>![Pattern 4 height]</td>
</tr>
<tr>
<td>i</td>
<td>![Pattern 4 height]</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>![Pattern 4 height]</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>![Pattern 4 height]</td>
<td></td>
</tr>
<tr>
<td>ε</td>
<td>![Pattern 4 height]</td>
<td></td>
</tr>
</tbody>
</table>

Casali makes the related observation that in treatments of many
languages with three-height systems, it is the [-atr] value that is posited as the
lexically marked feature due to its behaviour as the phonologically active feature in processes such as spreading and assimilation. In four-height systems, however, it is the [+atr] value which is preserved under coalescence and has otherwise been analysed as the lexically marked feature due its phonological activity in featural processes.

In the next sections, I describe a few ways that these patterns may be accounted for. The first type of approach relies on output constraints alone to motivate the different patterns of coalescence. The second type of approach looks to the structure of input representations to explain the different patterns.

6.2 An output constraint-based account

Parallel to the discussion of place features in Chapter 5, we are faced with the question of how to capture the apparent relationship between the number of height contrasts in a language and the patterning of height/atr specifications in a system. As with the place coalescence facts, there are a few possible ways of accounting for the height coalescence patterns. First, if it is assumed that phonological patterning is a result of output constraints alone, and that segmental representations are based entirely on the phonetic outputs, we must rely on feature-specific Faithfulness constraints to derive the correct patterns in a language.

---

7Casali (1996) lists the following languages and corresponding references where [-atr] is posited as the lexically marked feature:
- Edo (Wescott 1962)
- Ewe (Westermann 1930)
- Komo (Thomas 1942)
- Moba (Russell 1985)
- Wolof (Archangeli and Pulleyblank 1994)
- Yakata (Motingéa 1993)
- Yoruba (Archangeli and Pulleyblank 1994)
Under this sort of account, a language in which the [-atr] value is "active" (three-height system in (11)) would have a ranking of MAX-atr >> MAX+atr. This would ensure that the [-atr] value in an input is preserved over a [+atr] specification, as in the tableau in (12). In languages where the coalescence patterns demonstrate an active [+atr] value (four-height system in (11)), the ranking is reversed, and [+atr] is always preserved, as in the tableau in (13).

\(12\) INPUT \( e+\varepsilon \)

\[
\begin{array}{cccc}
\text{MAX-attr} & \text{MAX+attr} \\
1. e & *! & \\
2. \varepsilon & * & \\
\end{array}
\]

\(13\) INPUT \( e+\varepsilon \)

\[
\begin{array}{cccc}
\text{MAX+attr} & \text{MAX-attr} \\
1. e & * & \\
2. \varepsilon & *! & \\
\end{array}
\]

Of course, the problem with such an approach is that it offers no explanation as to why MAX-attr should be higher ranked in an inventory with three heights, while MAX+atr should be ranked higher in a four-height inventory. As with the place coalescence patterns, it is clear that what is needed is some notion of input specification that is dependent on the inventory. This is pursued in the next section.
6.3 An alternative approach to height specifications

As demonstrated in the previous sections, there is an important generalization to be made regarding the relationship between the number of height contrasts in an inventory and the behaviour of vowels in the system. Recall from chapter 5 that a similar relationship is observed between the number of place contrasts in a vowel inventory and the patterning of vowels in the system. I have argued that one promising way to explain these generalizations is to attribute the different behaviours of a particular vowel to differences in its input representations: a particular vowel place may have different representations depending on the shape of the inventory. A similar view of height representations is likewise proposed here.

In this section I argue that the relationship between height coalescence patterns and height contrasts is easily captured following the proposals laid out in the Chapter 3 regarding the building of segmental representations in response to pressure of contrast and considerations of complexity. In the discussion of height features, I further develop the anti-complexity constraint *STRUCTURE to separately govern different types of structural complexity.

The section is organized as follows. First, I will argue for adopting the assumption of Clements (1989), Rice and Avery (1993) and Goad (1993), that in a system such as the one in (14a), the distinction between /e/ and /ɛ/ is one of height and not [atr]: the vowel /ɛ/ is specified as [low]. In a system such as the one in (14b), on the other hand, the distinction between both the high vowels (i/i) and the mid vowels (e/ɛ) is one of [atr]. Since the feature [atr] is unary rather than binary-valued, there is no expected phonological "activity" on the part of a [-atr] feature. In a three-height system as in (14a), the feature which we have been calling [-atr] in our description of the patterns is actually the feature [low] under this view. The activity of [atr] in the four-height system
versus its inactivity in a three-height system is thus explained: only four-height systems bear [atr] specifications.

(14)

a. 

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>()</td>
<td>u</td>
</tr>
<tr>
<td>e</td>
<td>()</td>
<td>o</td>
</tr>
<tr>
<td>ε</td>
<td>a</td>
<td>o</td>
</tr>
</tbody>
</table>

b. 

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>()</td>
<td>u</td>
</tr>
<tr>
<td>I</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>()</td>
<td>o</td>
</tr>
<tr>
<td>ε</td>
<td>a</td>
<td>[low]</td>
</tr>
</tbody>
</table>

In the following sections, I present arguments in support of the specifications in (14).

6.3.1 [low] in three-height systems

Importantly, the behaviour of the vowels {ɛ, a, o} in the two types of systems supports such specifications. In systems as in (14a), the vowels /ɛ/ and /ɔ/ are specified as [low], along with the vowel /a/. Recall from the coalescence discussion in §6.1 that it is in these very systems that /ɛ/, /ɔ/, and /a/ pattern as a natural class. I repeat some of the data from Owon Afa below in (15) for the reader's convenience.

(15) Owon Afa (Casali 1996: 120, citing Awobuluyi 1972)

a. /da ehwe/ [dehwe] buy book
b. /te esi/ [te:si] press on the ground
c. /da opu/ [dɔ:pu] buy dog
d. /ewo ɔba/ [ewo:ba] king's money

Recall from the previous discussion of these facts the observation that it is the [-atr] value and not the [+atr] value that is maintained or "active" in the coalescence. Also, the low vowel /a/ patterns as a [-atr] vowel along with /ɛ/ and /ɔ/. Given the view that the vowels /ɛ/ and /ɔ/ are specified as [low]
in this system, it is not surprising that they pattern as a natural class with the low vowel /a/. Further, it becomes unnecessary to explain the [-atr] patterning of /a/: the 'active' feature here is not [-atr] but [low].

In addition to coalescence data, there is further support for the specification of /ɛ/ and /ɔ/ as [low] in languages with inventories like (14a). This support comes in the form of phonological processes grouping /ɛ/ and /ɔ/ with the low vowel /a/ as a natural class, and the phonological activity demonstrated by the feature [low] on these vowels.

One process Goad describes (citing Clements 1974) is a vowel height assimilation in Adangbe Ewe. The vowel inventory for this language is given in (16). In this process, the post-clitic /e/ agrees in height with an adjacent stem-final vowel (pp. 41-42).

(16)

\[
\begin{array}{cccc}
\text{i} & \text{u} \\
\text{e} & \overset{\circ}{\text{o}} \\
\overset{\varepsilon}{\text{a}} & \overset{\circ}{\text{o}}
\end{array}
\]

Notice that in (17), the clitic assimilates to the height of the high vowels of the stems, and itself becomes a high vowel.

(17) High stem-vowel

\[
\begin{align*}
\text{a. } i + e & \rightarrow i i & \overset{\varepsilon}{\text{si}} + \overset{\circ}{\text{e}} & \rightarrow \overset{\varepsilon}{\text{si}}\overset{\circ}{\text{i}} & \text{it's water} \\
\text{b. } u + e & \rightarrow u i & \overset{\varepsilon}{\text{vũ}} + \overset{\circ}{\text{e}} & \rightarrow \overset{\varepsilon}{\text{vũ}}\overset{\circ}{\text{i}} & \text{it's a dog}
\end{align*}
\]

In (18), the clitic vowel surfaces as a mid vowel following mid stem vowels /e/, /o/, and /a/.
(18) Mid stem-vowel

a. \( e + e \rightarrow e e \)  
\( \text{ayé} + é \rightarrow \text{aye é} \)  
*it's a spider*

b. \( ø + e \rightarrow e e^8 \)  
\( \text{anye} + é \rightarrow \text{anye é} \)  
*it's me*

c. \( o + e \rightarrow o e \)  
\( \text{owo} + é \rightarrow \text{owo é} \)  
*it's you*

Finally, in (19), the clitic vowel becomes \([e]\) following the low stem vowel \(/a/\) and the stem vowels \(/\varepsilon/\) and \(/\varepsilon/\). This pattern is easily accounted for if it is assumed that, like \(/a/\), \(/\varepsilon/\) and \(/\varepsilon/\) are specified as \([\text{low}]\).

(19) Low stem-vowel

a. \( e + e \rightarrow e e \)  
\( \text{âve} + é \rightarrow \text{âve é} \)  
*it's a weaver*

b. \( a + e \rightarrow e e \)  
\( \text{âgb} + é \rightarrow \text{âgb é} \)  
*it's a load*

c. \( ò + e \rightarrow ò ò \)  
\( òs + é \rightarrow òs é \)  
*it's you*

Another language discussed by Goad (1993) with a vowel system similar to Adangbe Ewe is Bijago, as given in (20).

(20)

\[
\begin{array}{cccc}
& i & u & \\
& e & o & \\
& ð & a & ò
\end{array}
\]

In Bijago, there are two processes that treat \(/\varepsilon/\) and \(/\varepsilon/\) together with \(/a/\) as low vowels. In one process, all three vowels trigger a lowering of \(/\varepsilon/\) and \(/o/\) to the left of them to \([e]\) and \([\varepsilon]\). Examples for the noun class prefix \(o-\) are given in (21).

(21)

\[
\begin{array}{ll}
\text{ò-te} & \text{father} \\
\text{ò-raas} & \text{captive} \\
\text{cf.} & \text{o-misom} \quad \text{mother} \\
\text{ò-jooko} & \text{person, Bijagó}
\end{array}
\]

---

\(^8\)Note that the central stem-final vowels \(/a/\) and \(/a/\) become front vowels before the post-clitic \(/ê/\).
The second process is a height assimilation applying to certain noun class prefixes of the shape Cu-. The prefix vowel is lowered to [o] before mid vowel stems\(^9\) and [ɔ] before stems containing /ɛ/, /ɔ/ and /a/.

\[(22)\]

<table>
<thead>
<tr>
<th>High Vowel</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ku-gbi</td>
<td>body</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>ku-siŋ k-u-ni-ny</td>
<td>my cows</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mid Vowel</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>c.</td>
<td>ko-pono</td>
<td>mark</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>ku-siŋ ko-so obe</td>
<td>two cows</td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>ku-siŋ ko-deŋbọŋ</td>
<td>big cows</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low Vowel</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>f.</td>
<td>ko-no</td>
<td>liver</td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td>ku-siŋ ko-gaŋ</td>
<td>those cows</td>
<td></td>
</tr>
</tbody>
</table>

Thus, the patterning of the vowels /ɛ/ and /ɔ/ with the low vowel /a/ in these three-height systems supports the claim that /ɛ/ and /ɔ/ are specified as [low]. Also, the activity of "[-atr]" in coalescence patterns in these languages is explained: it is not the feature [-atr] that is being preserved, but the feature [low].

6.3.2 [atr] in a four-height system

Recall from the height coalescence patterns described above that it appears as though in four-height systems, the active value of [atr] in a binary-feature system is [+atr]. Thus, for instance, in the Gichode data, repeated below, the [atr] value of the output of coalescence is determined by the [+atr] vowel, regardless of the order of the input vowels. Only where there are two [-atr] input vowels (as in the forms in (h)) does the output vowel surface as [-atr].

---

\(^9\)Notice that in this process the high vowel assimilates to the mid vowel while in the Adangbe Ewe data given in (17), the mid vowel assimilates to the height of the high vowel. In §6.4 I argue that these differences are a result of different input representations which are driven by competing constraints on structural complexity.
(9) Activity of [+atr] in a four-height system

<table>
<thead>
<tr>
<th></th>
<th>[+atr] + [-atr] → [+atr]</th>
<th>e + i → e</th>
<th>gibege idgan</th>
<th>gibegeidgan</th>
<th>slave's thighs</th>
</tr>
</thead>
<tbody>
<tr>
<td>f.</td>
<td>[-atr] + [+atr] → [+atr]</td>
<td>i + i → i</td>
<td>fuli idgo</td>
<td>fuli idgo</td>
<td>deer's yams</td>
</tr>
<tr>
<td>h.</td>
<td>[-atr] + [-atr] → [-atr]</td>
<td>e + i → e</td>
<td>atanafise ifin</td>
<td>atanafiseifin</td>
<td>female twin's veins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>e + i → e</td>
<td>atanafise oc</td>
<td>atanafiseoc</td>
<td>female twin's sores</td>
</tr>
</tbody>
</table>

These facts fall out neatly if it is assumed that in four-height systems, the vowels /i, i/ and /e, e/ are distinguished by a monovalent [atr] specification, as in the chart below in (23).¹⁰

(23)

```
i    u    [atr]
i    o
e    o    [atr]
e    o
```

When input specifications are maintained under coalescence, the presence of an [atr] vowel in the input will entail an [atr] output. From the discussion of input specifications and phonological activity in §3.2, we may expect that if [atr] is present in input specifications, it could function actively in phonological processes other than coalescence. This is exactly what we find in Chumburung and Turkana, languages with inventories similar to Gichode. In (24) I give the inventory for Chumburung, a Kwa language spoken in Ghana.

¹⁰I return to the question of representing the height distinctions in different systems in §6.4.
Snider (1989) describes a process of [atr] harmony in Chumburung which operates both within words and between phrases. Within a word, all vowels are either [+atr] or all are [-atr]. The vowel /a/ occurs with either set of vowels, displaying an allophonic variant [ɔ] when it occurs to the left of a [+atr] vowel in a word.

(24) Chumburung Vowel Inventory

<table>
<thead>
<tr>
<th>I</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>E</td>
<td>O</td>
</tr>
<tr>
<td>E</td>
<td>O</td>
</tr>
<tr>
<td>A</td>
<td>a</td>
</tr>
</tbody>
</table>

Within phrases, [+atr] spreads both leftward and rightward. Rightward spread affects only a high vowel in the next syllable, while leftward spread affects any vowel in the next syllable to the left, and only subsequent high vowels further left.\(^{12}\)

(25) 

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>jono</td>
<td>dog</td>
</tr>
<tr>
<td>b</td>
<td>keri</td>
<td>side</td>
</tr>
<tr>
<td>c</td>
<td>okpe</td>
<td>witch</td>
</tr>
<tr>
<td>d</td>
<td>ipesi</td>
<td>broom</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>friri</td>
<td>deer</td>
</tr>
<tr>
<td>f</td>
<td>koti</td>
<td>monkey</td>
</tr>
<tr>
<td>g</td>
<td>sa?</td>
<td>woman</td>
</tr>
<tr>
<td>h</td>
<td>ipiri</td>
<td>doors</td>
</tr>
</tbody>
</table>

(26) /a/ left of [+ atr]

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>kpeksi</td>
<td>rat</td>
</tr>
<tr>
<td>b</td>
<td>jende?</td>
<td>onion</td>
</tr>
<tr>
<td>c</td>
<td>kapidini</td>
<td>ring</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>kuruma</td>
<td>donkey</td>
</tr>
<tr>
<td>e</td>
<td>tukpa</td>
<td>travel</td>
</tr>
<tr>
<td>f</td>
<td>sbruwa</td>
<td>eight</td>
</tr>
</tbody>
</table>

(27) Rightward spread

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>jono + wiri</td>
<td>jono wuri</td>
</tr>
<tr>
<td>b</td>
<td>kofi + kibaŋ</td>
<td>kofi kibaŋ</td>
</tr>
<tr>
<td>c</td>
<td>oluŋ + kisibɔ</td>
<td>oluŋ kisibɔ</td>
</tr>
</tbody>
</table>

---

\(^{11}\) In the transcription system used in Snider (1989), the symbol [v] is used to represent the lax counterpart of [u]. I follow this practice in the Chumburung data that follow.

\(^{12}\) Note that there is an additional process of place harmony evidenced in (28). This process is discussed in more detail in \(\S 6.3\), but is irrelevant to the present discussion. In (28b), the vowel [ɔ] appears as the [atr] correspondent to /a/.
(28) Leftward spread

a. firi + keri  \[\text{furi keri}\]  deer’s side
b. ki\[\text{ŋa}\] + jono  \[\text{ki\[\text{ŋa}\] jono}\]  slave’s dog
c. ibiri + kudu  \[\text{iburu kudu}\]  ten voices

Snider notes that [-atr] never spreads to [+atr] vowels, and therefore is inactive.

To summarize briefly at this point, in three height inventories, the vowels /e/ and /\[\text{o}\]/ are [low], and vowels such as /e/ are not expected to pattern as [atr]. In four height systems, /i/, /e/, /u/, and /o/ are [atr] and therefore are expected to pattern accordingly. The following tableaux demonstrate the different behaviour of the same vowels in the two different systems. Notice that since Faithfulness is not feature-specific and works only to preserve input features in general, it is the input representations that determine the output of the coalescence.

(29) Three-height system:  no [atr] contrast

\[\varepsilon = [\text{low}]\]

\[
\text{INPUT } /\varepsilon + \varepsilon / \\
/\varepsilon/ \\
[\text{low}] \\
\]

<table>
<thead>
<tr>
<th>\text{VV}</th>
<th>\text{MAXF}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. e</td>
<td>*!</td>
</tr>
<tr>
<td>2. \varepsilon</td>
<td>&lt;[\text{low}]&gt;</td>
</tr>
</tbody>
</table>

\[\varepsilon \rightarrow \text{low} \]

\[
\begin{array}{c|c|c}
1. & \text{VV} & \text{MAXF} \\
\hline
\text{e} & & *! \\
\hline
\varepsilon & & <[\text{low}]> \\
\end{array}
\]
Four-height system:

\[ e = [\text{atr}] \]

\[
\text{INPUT} /e+\varepsilon/ \\
\left| \\
[\text{atr}] \\
\]

<table>
<thead>
<tr>
<th>( \text{VV} )</th>
<th>MAXF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( \varepsilon )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>![atr]</td>
</tr>
<tr>
<td>2. ( \varepsilon )</td>
<td></td>
</tr>
</tbody>
</table>

Thus, the postulation of \( \varepsilon \) as [low] in a three-height system allows the correct output to be selected by the Faithfulness constraint MAXF. In a four-height system, since the distinction between the mid vowels is one of [atr], the vowel /\varepsilon/ is unmarked while the vowel /e/ is marked as [atr]. The output constraints select the more faithful candidate in the case of coalescence, which is the candidate bearing the [atr] specification.

As Casali (1996) points out, it is not enough to simply state that in three-height systems, there is no [atr] distinction but only a high-mid-low distinction, while in four-height inventories, [atr] functions actively and therefore is included in the system's feature specifications. Instead, an explanation for these generalizations is to be sought. Therefore, the question addressed in the next section is why a particular system should prefer a particular configuration of specifications over another.
6.4 Constraining representations

As outlined in §3.6, the position of this thesis is that representations are determined in response to two conflicting requirements: while representations for different feature specifications need to be made distinct, structural complexity is to be minimized. In this section I will argue for representations which allow different specifications to be distinct, while at the same time conforming to the requirements of structure minimization. I will also refine the requirement of structure minimization through the parameterization of *STRUCTURE to govern different types of complexity. Structural complexity, I suggest, is evaluated in two ways: the degree of vertical complexity, and the degree of horizontal complexity. We see the interaction of these considerations in the range of possible inventory types and the patterning of vowels within those inventories.

Obviously, one major difference between a three-height system and a four-height system is the number of contrasts made in the inventory. The major generalization in terms of the proposed representations is that [atr] only comes into play in the larger systems. [atr] therefore seems to be a distinction that is typically utilized after three height distinctions have already been made.13 In terms of a path of elaboration, then, [atr] seems to be a later distinction than height. In this section, I address the question of whether this generalization can be made to follow from the organization of features and from principles governing representational complexity. I will discuss a proposal of Goad (1993) for representing height/[atr] specifications which appear to lead towards an answer. Drawing on the advantages of this proposal, I suggest a configuration which allows an account of the different

13Although the Kikuyu system, discussed below, seems to suggest that [atr] can be brought in in a two-height system. Below, in §6.4.3, I suggest that the Kikuyu-type system represents one possible phonetic realization for a system which chooses [atr] before the feature [Open].
patterns associated with vowels in different inventories, as well as
generalizations about implicational relationships observed between height
contrasts and [atr].

First I present proposals for height representations from Goad (1993),
and show some of the arguments in favour of her view of vowel height.
Then I demonstrate how Goad's proposal fails to account for the occurrence of
certain inventory types and the non-occurrence of others. Finally, drawing
on some notions of constraining segmental elaborations from Rice and Avery
(1993), I show how the different attested inventory types can be derived from
constraints on different types of structural complexity.

6.4.1 The representation of Height

In the representation of vowel height that Goad (1993) proposes, height
features form their own constituent outside of Place features, as in (31). All
vowels minimally bear the node Voc(a1ic). The Vocalic node has two
dependent features: [open] and [low]. As this view of features assumes
monovalency of feature specifications, [open] replaces [-high] in a binary
specification model.

(31) Voc
     |
   [open]
     |
   [low]

Thus, the representations for different vowel heights (high, mid, low)
are as in (32). Note that [low] is ruled out as a direct dependent of Vocalic.

---

14Later in this section, [atr] is included as a dependent of Vocalic.
15Goad intends these features not as correlates of articulatory gestures but as acoustic features:
"height features do not map onto a single articulatory gestures and are best defined in acoustic
terms" (p. 118).
Goad presents arguments from several different types of processes in support of these height representations. First, she argues that raising and lowering shifts\textsuperscript{16} are easily and elegantly formulated with these representations involving "subset" relations. Other height models face difficulty in accounting for these shifts because they involve, for example, raising from mid to high and raising from low to mid: two different structural changes. With Goad's representations, these shifts may be captured via a unified process: deletion of a terminal dependent of Vocalic (raising), or addition of a terminal feature to Vocalic (lowering).

An additional advantage to such an approach is that analysing raising as the removal of structure is consistent with the fact that it is often considered a type of weakening process, while lowering, analysed as the addition of structure, is a type of strengthening process (Harris (1990)).

Further arguments for the representations in (32) come from different height harmony processes. Several predictions regarding the operation of harmony fall out of the representations in (32). I summarize these in (33), and discuss them each in turn below.\textsuperscript{17}

\textsuperscript{16}Interestingly, Goad notes that the type of raising and lowering shifts she describes occur frequently in what we are calling two and three-height systems, but not in languages with four or more heights.

\textsuperscript{17}Note that there is an assumption here that only terminal features can spread.
(33)  
a. Low Harmony  
i. Mid vowels are targets  
ii. High vowels are either transparent or opaque, but do not participate  
b. Mid Harmony  
i. Mid and high vowels participate  
ii. Low vowels are opaque  
c. High Harmony  
i. Involves delinking or removal of structure\(^{18}\)  
ii. Mid or low vowels lose structure to raise in height.  

For low harmony processes, where presumably the feature [\text{low}] spreads, the representations predict that only mid vowels should be targets for the process (34a.i). High vowels should either behave as neutral or block the spread of [\text{low}], but they should not participate since they lack the Vocalic dependent [\text{open}] for [\text{low}] to spread onto (33a.ii). An examination of low harmonies finds high vowels behaving in just these ways: in Yoruba, high vowels are opaque to the spreading of low, while in Wolof, high vowels are neutral to low harmony (Goad (1993)). The inventories of these languages are given in (34).

\[(34)\]  
a. Yoruba inventory  
\begin{align*}  
i & \ u \\
e & \ e  \\
\varepsilon & \ a  \\
\end{align*}  
b. Wolof inventory  
\begin{align*}  
i & \ u \\
e & \ e \ & \ o  \\
\varepsilon & \ a \ & \ o  \\
\end{align*}  

For mid-harmony processes, the representations predict that only high and mid vowels should participate (33b.i). Low vowels should be opaque

\(^{18}\)Once outright loss of input structure is permitted (outside of neutralization), the output is no longer dependably determined by input structure, and the theory loses a great deal of predictive power. In §6.4.3, I discuss an alternative account of high harmony and some metaphony processes that upholds the view that outputs are predictably determined by input representations.
since any spreading of [open] over them would result in crossed association lines (33b.ii). Goad finds that these predictions are borne out in the mid-harmony systems of Yaka and Chichewa\(^{19}\).

In addition to these arguments, these representations also allow for an elegant account of certain generalizations about height coalescence patterns. As Casali (1996:78) points out, [-high] specifications are always preserved over [+high] under coalescence. Thus, an input consisting of a [-high] vowel and a [+high] vowel will coalesce to a [-high] vowel. Casali's account of this is to posit a universally fixed ranking of feature specific Faithfulness constraints so that MAX[-high] is always ranked over MAX[+high].

This ranking is a simple statement of the facts rather than an explanation. Notice however, that representations such as those proposed by Goad (1993) allow this generalization to fall out as a consequence of the feature specifications: since the specification of [-high] vowels involves the nodes [Voc] and [Open] while [+high] vowels are simply [Voc], maximal preservation of input material entails an output with both [Voc] and [Open]: a [-high] vowel. Thus the representations like those in (32) obviate the need for a stipulated fixed ranking of Faithfulness constraints; only a general notion of Faithfulness to input structure is required.

The representations in (32) fit in perfectly with the project of this thesis: the behaviour of the different segments is largely determined by their representations. Like the place specifications argued for in Avery and Rice (1989) and Rice and Avery (1993), and in the previous chapter, these height representations are tied to the inventory of the language, and the representations are more complex as the number of contrasts increase. In the

\(^{19}\)In Goad's analysis, low vowels are prohibited from being triggers of [open] spread themselves in these systems by a constraint on the process that says that [open] only spreads if it is terminal.
next section, I examine the proposal of Goad (1993) to include \[\text{[atr]}\] as an additional height distinction, and outline the consequences of such a move.

6.4.2 \([\text{atr}]\) as a Height feature

One important part of Goad's proposal is that the feature \([\text{atr}]\) be grouped with height features as a dependent of the Vocalic node. The features \([\text{atr}]\) and \([\text{low}]\) occupy the same position in the geometry as dependents of \([\text{Open}]\) under \([\text{Vocalic}]\).

\[
\begin{array}{c}
\text{Voc} \\
\mid \\
[\text{[open]}] \\
\mid \\
[\text{[low]}/[\text{[atr]}]
\end{array}
\]

Although they occupy the same position, \([\text{low}]\) and \([\text{atr}]\) are not the opposite values of the same feature and they are not subject to the same distributional constraints: where \([\text{low}]\) is not permitted as a direct dependent of \([\text{Vocalic}]\), \([\text{[atr]}]\) directly under \([\text{Vocalic}]\) gives a high [atr] vowel as in the representation in (36).

\[
\begin{array}{cccccc}
\text{High Vowel} & \text{Mid Vowel} & \text{Low Vowel} & \text{High, Atr V} & \text{Mid Atr V} \\
\text{Voc} & \text{Voc} & \text{Voc} & \text{Voc} & \text{Voc} \\
\mid & \mid & \mid & \mid & \mid \\
[\text{[open]}] & [\text{[open]}] & [\text{[open]}] & [\text{[open]}] & [\text{[open]}] \\
\mid & \mid & \mid & \mid & \mid \\
[\text{[low]}] & [\text{[atr]}] & [\text{[atr]}] & [\text{[atr]}] & [\text{[atr]}]
\end{array}
\]

Goad addresses the question of why it is that \(\text{/e/} \) and \(\text{/o/} \) appear to be \([\text{low}]\) unless the high vowels in the system contrast for \([\text{[atr]}]\) (in which case the mid vowels may contrast for \([\text{[atr]}]\), leaving \(\text{/e/} \) and \(\text{/o/} \) marked for \([\text{[atr]}]\) and \(\text{/e/} \) and \(\text{/o/} \) as simple mid vowels.) The answer, she suggests, is in the fact that \([\text{[atr]}]\) is selected as a contrastive feature in the inventory only if the vowel contrasts in the system cannot be captured by \([\text{[low]}]\) alone. Thus, \([\text{[low]}]\) is
utilized in an inventory before [atr]. This leaves the feature specifications in a three-height system as in (32), and in a four-height system as in (36).

There are a few things to note about Goad's proposal to include [atr] as a dependent of Vocalic. First, the arguments she presents for the inclusion of [atr] are all essentially a single argument: [low] and [atr] are in complementary distribution. Thus, [atr] harmony may be blocked by an intervening low vowel in the harmony span because the [low] feature already occupies the same position that [atr] would spread to.

Although having [atr] and [low] occupy the same position in the geometry is one way to ensure they do not co-occur, another fairly straightforward and less controversial approach is to posit a cooccurrence constraint *[atr][low], which would allow for the same distributional facts to find an explanation (see, e.g. Rice and Avery (1991), Archangeli and Pulleyblank (1994) for some discussion of this.).

There are, on the other hand, arguments against [atr] being a dependent of [Open]. First, the position of [atr] under [Open] rules out the possibility of languages utilizing an [utr] contrast for all vowels in its inventory, including low vowels. Yet, it appears that such systems do, in fact, exist (cf. Maddieson (1984): Amo (122), Logbara (215)).

Second, it is an apparent stipulation that Goad suggests that [utr] is only selected after [low]. It does not follow from the geometry, particularly since [utr] can be both a dependent of [Voc] and of [Open].

Finally, since Goad assumes that [utr] is selected only after the height features [Voc], [Open], and [Low] are exhausted, she rules out the possibility of systems such as the Kikuyu-type system (repeated in (37)), where high vowels

---

20Later in this section I will argue for representations which allow for these distributional facts to fall out from more general constraints on representational complexity.
contrast for [atr] but mid vowels do not. As Casali's survey suggests, the
cvowels /ɛ/ and /ɔ/ function as a natural class with /a/. However, the high
vowels require an [atr] feature to distinguish them.\(^{21}\)

\[
\begin{array}{ccc}
\text{i} & \underline{\text{u}} & [\text{atr}] \\
\text{I} & \underline{\text{U}} \\
\varepsilon & \text{a} & \text{O} & [\text{Open}]
\end{array}
\]

Under Goad's assumptions about the exhaustion of height features before the
introduction of an [atr] distinctions, this type of system is difficult to explain.
It seems that these systems do not require three height features to distinguish
their vowels, but two height features and an [atr] distinction in the high
vowels. Thus, the feature [low] need not be introduced before the feature [atr].
This means that an account such as this provides no explanation for why in a
Yoruba-type three-height system, /ɛ/ and /ɔ/ behave as [low] and /ε/ and /o/
do not behave as [atr].

\[^{21}\text{We might expect to find evidence of the [atr] specification on /i/ and /u/ in coalescence}
patterns in these languages, but unfortunately, of the three languages in Casali's survey which
have this type of inventory, /i+I/ and /I+i/ sequences are resolved positionally, and the first
vowel is deleted (Casali 1996:127). However, the vowels /a/, /ɛ/, /ɔ/ behave identically in
coalescence patterns, indicating that they all share a common height specification. In
subsequent discussion, I will argue that these systems actually represent two-height systems,
with an [atr] contrast in the high vowels.\]
6.4.3 [atr] as a Separate Dimension

In seeking an explanation for cross-linguistic generalizations about vowel inventory shapes, Rice and Avery (1993) propose a theory of feature specification which is similar to Goad (1993) in having height specifications which associate the degree of aperture (lowness of the vowel) with the complexity of the representations (38).²²

(38)     | i       | u       | a       |
       Root          Root                    Root |
       | \                  | \                  |
       | Vocalic            | Vocalic            |
       |                   |                   |
       Place            Place-Peripheral

In their view, vowel inventory shapes arise out of considerations of cumulative representational complexity. For example, lower vowels are more complex in having a dependent under Vocalic, therefore if Place structure is to be elaborated in the system, the high vowels are the first to expand in Place structure. This creates a less complex segment overall than elaborating Place structure on a lower vowel (i.e. adding [Place] to /a/). This view allows an explanation for why inventories tend to allow higher vowels to have more place contrasts than lower vowels.²³

The system in (38) is expanded to a five-vowel system through the addition of mid vowels. Rice and Avery (1993: 146) suggest that mid-vowel /e,o/ representations are the same as those for /i,u/ except that the node [Open] is added as a dependent of [Vocalic].

Rice and Avery (1993) assume, like Clements (1989) and Goad (1993), that in systems such as the one in (39a), the distinction between the mid

---

²²I have simplified the representations of Rice and Avery (1993) in (38), retaining only the nodes which have relevance to place and aperture specifications of vowels.

²³For a similar suggestion regarding elaboration of inventories, see Hamilton (1996).
vowels /e,o/ and /ɛ, ɔ/ is one of height and not [atr]. Like Goad (1993), Rice and Avery (1993) assume that [atr] is only distinctive for mid vowels where it is also distinctive for the high vowels (i.e. in systems such as (39b).

(39)

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i</td>
<td>u</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>e</td>
<td>o</td>
<td></td>
<td>[Open]</td>
</tr>
<tr>
<td></td>
<td>ɛ</td>
<td>a</td>
<td>ɔ</td>
<td>[Open-Low]</td>
</tr>
<tr>
<td>b</td>
<td>i</td>
<td>u</td>
<td></td>
<td>[atr]</td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>u</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e</td>
<td>o</td>
<td></td>
<td>[Open] [atr]</td>
</tr>
<tr>
<td></td>
<td>ɛ</td>
<td>a</td>
<td>ɔ</td>
<td>[Open]</td>
</tr>
</tbody>
</table>

However, unlike Goad (1993), in this view [atr] is not treated as a height feature, that is as a dependent of Vocalic. Instead [atr] occupies a separate branch in the structure, horizontal to the Vocalic node. It is this placement of [atr] that allows the addition of [atr] to be a predictable characteristic of four-height inventories.

From the level of detail of Rice and Avery (1993), however, it is difficult to ascertain what feature actually distinguishes mid from low vowels: [Open] is treated as equivalent to [low] (p. 146). In all of their representations of mid vowels /e, ɛ, ɔ/ and the low vowel /a/ there appears to be no height distinction between them: /a/ is set off from the mid vowels in not having Place, but it is not clear how the three different heights are distinguished in this model.

On the other hand, if we assume that there is in fact a feature [low] which, as Goad suggests, is a dependent of [Open], three heights can be distinguished. If only two heights are to be distinguished, then [low] need not play a role. In this type of system, the low vowel is marked with [Open] only.
In addition, the positioning of [atr] on a separate branch from vowel height avoids the pitfalls of Goad's geometry. Thus, the proposed geometry is as in (40).

(40) \[SV^{24}\]
     \[\text{[Vocalic]} \quad \text{[atr]}\]
     \[\text{[Open]}\]
     \[\text{[Low]}\]

Note that this height representation benefits from the same empirical advantages as those proposed in Goad (1993): lower vowels are increasingly more complex in terms of their representation of height. This allows the same elegant analysis of the raising and lowering shifts discussed in Goad (1993), as well as the different height harmonies she discusses. As far as the patterns in height coalescence discussed in Casali (1996), this representation, like Goad's, allows for the preservation of [-high] to fall out as a prediction instead of requiring a fixed ranking of height Faithfulness constraints.

At the same time, the positioning of [atr] on its own branch instead of as a dependent of [Vocalic]/[Open] allows for the possibility for languages to make an [atr] distinction among low vowels as well as non-low vowels.

Recall that one of the important generalizations to be explained is why mid vowels are never specified for [atr] unless the high vowels also are. Instead, in three-height systems, the vowels /ɛ/ and /ɔ/ behave as though they are [low] while /e/ and /o/ are mid. We have said already that there is a tendency for languages to make use of all the height distinctions available,

---

24Recall that the organizing node SV ("Sonorant Voice" in the system of Rice and Avery (1989, 1991, 1993) and "Spontaneous Voice" in Piggott (1992)) is contained in all sonorant segments, including vowels.
and then, if [atr] is introduced to create additional contrasts, the high vowels are the first to be distinguished.

I suggest that if the representations have [atr] on a separate branch of structure (as in (40)) we can account for the distribution of [atr] in terms of representational complexity: the addition of [low] as a dependent of [Open] before the addition of [atr] as an additional branch of structure is the difference between the elaboration of vertical versus horizontal complexity. Vertical complexity involves embedded structure: a single dependent involves one level of vertical structure, and a dependent on that specification adds another level. Horizontal structure is branching structure: a node with a single direct dependent has less horizontal structure than a node with multiple direct dependents.

In terms of the proposed representation in (40), the difference between horizontal and vertical complexity is crucial: making all three height distinctions requires that one set of vowels in the systems (i.e. the low vowels) have a maximally complex Vocalic projection, where Vocalic has two levels of dependents. The addition of an [atr] distinction would require that some vowels in the system have a branching structure under SV. I suggest that these two types of complexity are constrained by two different types of *STRUCTURE constraints, as given in (41).

(41)

a. \(^*\text{STRUCTURE}_{\text{Horizontal}}\)
   No branching representations.

b. \(^*\text{STRUCTURE}_{\text{Vertical}}\)
   A node should have no dependents.

A system which makes use of all of its height contrasts before introducing an [atr] distinction is elaborating vertical structure at the expense of \(^*\text{STRUCTURE}_{\text{Vertical}}\). In these languages, then, we might postulate a ranking
\*\*STRUCTURE_{Horizontal} >> \*STRUCTURE_{Vertical}. Thus, in systems such as the Yoruba-type system (42), no [atr] distinction is introduced. Instead, there are three different heights contrasted.\textsuperscript{25}

(42)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Voc</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>o</td>
<td>Voc — Open</td>
</tr>
<tr>
<td>e</td>
<td>a</td>
<td>O — Voc — Open — Low</td>
</tr>
</tbody>
</table>

In a Kikuyu-type system (43), on the other hand, the ranking appears to be reversed: horizontal complexity is tolerated in the form of an [atr] distinction on the high vowels, and only two heights are distinguished. Therefore, \*\*STRUCTURE_{Horizontal} is violated in favour of minimally violating \*\*STRUCTURE_{Vertical}.\textsuperscript{26}

(43)

<table>
<thead>
<tr>
<th>Atr</th>
<th></th>
<th>Voc</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>u</td>
<td>Voc</td>
</tr>
<tr>
<td>e</td>
<td>a</td>
<td>O — Voc — Open</td>
</tr>
</tbody>
</table>

Thus the ranking deriving this type of system is \*\*STRUCTURE_{Vertical} >> \*\*STRUCTURE_{Horizontal}. The two rankings result in the selection of different systems as illustrated in (44) and (45), below.

\textsuperscript{25}In the following tables, "vertical" structure is simply structure which does not branch, but has only one dependent per node. "Horizontal" structure is branching structure.

\textsuperscript{26}Minimally is the key word since the presence of [Open] on the vowels /ɛ/, /ɔ/, and /ɑ/ violates \*STRUCTURE_{vertical}, although not to the same extent as a further elaboration through the addition of [low].
Recall from Chapter 3 that the output constraints serve to guide the building of representations in the input. The representations are built in response to the number of contrasts in the inventory, as discussed in §3.4. The choice of which features are used to encode these contrasts is decided by the constraints on complexity type. The tableaux in (44) and (45) illustrate the selection of a set of representations based on a previously determined three-way contrast. A segment bearing a single specification incurs a single violation of both *STRUCTURE\text{Horiz}\text{Vert} and

27 Notice that here I am assuming that the learner has already established that the contrasts that they are making are either height or [atr] contrasts.
*STRUCTUREVert. A segment which bears specifications in two different dimensions (i.e. has branching structure) incurs two violations of *STRUCTUREHoriz. A segment bearing a specification which bears a dependent incurs two violations of *STRUCTUREVert.

The evaluation of a candidate system involves two considerations. First, there is a complexity ceiling \(^{28}\) that is determined by the most complex segment in a given system: the most complex segment incurs the greatest number of local violations of a particular *STRUCTURE constraint. Thus, two candidate systems may be compared with respect to their complexity ceilings. All things being equal (i.e. all systems successfully making the same number of contrasts) the system with the lowest complexity ceiling is preferred. Thus, in the candidate systems in \((44)\), the third candidate has the lowest horizontal complexity ceiling, with the most horizontally complex segment incurring a single violation. This same candidate system has a maximally complex segment in terms of vertical structure, incurring three violations of *STRUCTUREVert. However, since *STRUCTUREHoriz outranks *STRUCTUREVert in this language, the third candidate system wins out over the first and second candidates.

A candidate system is also evaluated in terms of the degree of complexity used to make the contrasts: for every additional level of complexity on a particular segment type, a violation of *STRUCTURE is incurred. The next tableau demonstrates the effect of a reverse ranking of the two *STRUCTURE constraints.

---

\(^{28}\)A similar approach is taken in Hamilton (1996).
a. Input: /make three contrasts/

<table>
<thead>
<tr>
<th></th>
<th>*STRUCTURE\text{Vert}</th>
<th>*STRUCTURE\text{Horiz}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>i Voc</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Atr</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>e &lt; Voc-Open</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>e Voc-Open</td>
<td>**</td>
</tr>
<tr>
<td>2</td>
<td>Atr</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>i &lt; Voc</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>/\varepsilon^{29} Voc</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>e Voc-Open</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>i Voc</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>e Voc-Open</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>e Voc-Open-Low</td>
<td>**</td>
</tr>
</tbody>
</table>

The first and second candidate systems fare better than the last system in terms of their complexity ceiling for vertical structure: the third candidate system incurs three violations of $^{*}$\text{STRUCTURE}_{\text{Vert}} while the first and second incur only two violations. Although the third candidate system maximally requires only a single violation of $^{*}$\text{STRUCTURE}_{\text{Horiz}}, it is ruled out. However, two candidate systems still remain. Both of them have the same complexity ceiling involving a maximum of two levels of structure. The choice between the first and second candidate systems is also decided on the basis of structural complexity except that, in this case, the selection of an optimal system is determined by a constraint that is the conjunction of the two $^{*}$\text{STRUCTURE} constraints. In the first candidate system, the mid vowel /\varepsilon/ will return to the issue of representing an i/e contrast in some three height systems as an [atr] contrast. I tentatively propose that such representations may be adopted to account for certain metaphony patterns as seen in many Romance languages.
bears a specification for both [atr] and has two levels of height complexity: [Voc] and a dependent [Open]. The fact that our actual output candidate system (candidate two) involves both segments with horizontal complexity (/i/) and vertical complexity (/ε/), means that it is neither a problem of *STRUCTUREHoriz or *STRUCTUREVert that rules out candidate one. Instead, I suggest, it is the fact that both types of complexity occur on a single segment in the system.

Recall that one of the generalizations to be made regarding the distribution of [atr] in inventories is that high vowels are the first to be elaborated for [atr] in an inventory. That is, high vowels may be specified for [atr] in a system regardless of whether mid vowels are or not (c.f. Kikuyu). However, [atr] is contrastive in the mid vowels only if high vowels also contrast for [atr]. This falls out from the fact that high vowels are the least complex in terms of height:\textsuperscript{30} bearing only a [Vocalic] specification, the addition of [atr] to high vowels creates a less complex segment type than the addition of [atr] to the already more complex mid vowels. The distribution of [atr] can be formally constrained with the notion of Local Conjunction (Smolensky (1995, 1997), Hewitt and Crowhurst (1995), Ito and Mester (1996), and Alderete (1996)) applied to the *STRUCTURE constraints. Local Conjunction allows the recombination of constraints in a particular domain so that violation of the combined constraints is worse than violations of the constraints separately. Thus, constraints \( A \) and \( B \) may be lower-ranked than \( C \) (\( C \gg A,B \)), but combined violations of \( A \) and \( B \) may motivate the violation of \( C \) (\( A&B \gg \)).

\textsuperscript{30}Rice and Avery (1993) make this same point: if any height is to be elaborated with [utr] it will be the high vowels since they are the least complex height.
Ranking of the conjoined *STRUCTURE constraints above the single constraints means that the cumulative violations of *STRUCTURE_{Horizontal} and *STRUCTURE_{Vertical} in a single domain (i.e. on a single segment) are worse than violations of these constraints separately. Therefore a high [atr] vowel is preferred over a mid [atr] vowel. In (47), the inventory with the mid [atr] vowel (candidate 1) incurs worse violations of the conjoined constraint in (46) since it involves horizontal complexity (branching structure) and two levels of vertical complexity (a [Vocalic] node with an [Open] dependent). The inventory with the high [atr] vowel (candidate 2) is preferred over the first candidate since it involves more distributed complexity: one type of vowel (/i/] is horizontally complex (branching structure), and a different type of vowel is vertically complex (/ε /). This distribution allows violations of the conjoined complexity constraint to be avoided.

(46) *STRUCTURE_{Vertical} & *STRUCTURE_{Horizontal}

An individual segment should not have both horizontal and vertical complexity.
Thus, the Kikuyu-type system is selected with this ranking. Note that this type of system is essentially a two-height system. The distinction between the lowest vowels (\{e, a\}) and the rest of the vowels in the system is one of height: the lowest vowels are marked as [open] while the other vowels are unmarked. The distinction between the vowels \{i, u\} and the rest of the vowels in the inventory is one of [atr]: the vowels /i/ and /u/ are marked for [atr] while the rest of the vowels are not. Thus, in this type of system, the vowels which are the most unmarked in terms of height and [atr] specifications are /i/ and /u/ since they bear only an empty Vocalic node. In this type of system, we expect that the high [atr] vowels may play an active role in the phonology, and spread their [atr] specification (see footnote 28).

One set of processes where we do see high vowels behaving as though they have an input specification to spread is in metaphony and high harmony processes. For example, Dyck (1995) describes a process of

\[31\] represent one violation of a conjoined constraint for every pair of simultaneous violations.
metaphony in Pasiego (Spanish) where a stressed mid vowel raises when followed a high vowel (see also McCarthy (1984)).

(48) Pasiego Inventory (Dyck (1995))\textsuperscript{32}
\begin{align*}
&i & u & e & o \\
& a
\end{align*}

(49) Pasiego Metaphony (Dyck (1995: 12), citing Penny (1969))

\begin{tabular}{ll}
Unmetaphonized & Metaphonized \\
a. afilit[é]ros & afilit[í]ru 'needle-cases, needle-case' \\
b. g[ó]rdo & g[ú]rdú 'fat (neuter), fat (masc.)'
\end{tabular}

The forms on the left in (49) demonstrate that a mid vowel in stressed position appears as mid when followed by a mid vowel. When an underlying mid vowel in stressed position is following by a high vowel, as in the forms on the right, the vowel appears as a high vowel. As shown in the forms in (50), only mid vowels participate in this alternation. High vowels remain high whether followed by a mid or low vowel ((50a-b) on the left) or a high vowel ((50a-b) on the right). Low vowels are also unaffected by the height of the following vowel (50c).

(50)
\begin{align*}
a. & luz m[í]yos & il m[í]yu 'mine (pl.), mine (sg.)' \\
b. & bj[ú]da & bj[ú]du 'widow, widower' \\
c. & br[á]dos & br[á]du 'arms, arm'
\end{align*}

In languages with this type of raising process, it appears as though a height feature from a high vowel is spreading to mid vowels.\textsuperscript{33} This poses a

\textsuperscript{32}In addition to these vowels, Dyck’s description involves two diphthongs ([we] and [je]) which I leave out for the purposes of the present discussion.

\textsuperscript{33}The alternative analysis suggested in Goad (1993) is also possible. Under this analysis, high harmony involves the delinking of Vocalic dependents. One of the reasons that this approach will not be pursued here is that the delinking would involve loss of input structure and therefore a violation of Faithfulness. This pattern would therefore be at odds with all of the other spreading and assimilation processes discussed in this work which involve the satisfaction of Faithfulness as the primary determinant of the nature of output segments.
potential problem for the current proposal since there is no feature [high] in the representation of height. However, in the Kikuyu-type system discussed above, the vowels /i/ and /u/ do have a specification to spread, and that is [atr]. If the distinction between high and non-high vowels in Pasiego is taken to be a distinction involving the presence vs. absence of an [atr] specification, then the high vowels /i/ and /u/ could bear a specification for [atr], while other vowels, including the mid vowels /e/ and /o/ bear no such specification. The raising of mid vowels could therefore be analysed, not as the spreading of a height feature from high vowels, but instead as the spreading of [atr]. This is shown schematically in (51).

(51) / é C i /    ⇒    [i C i ]

\[
\begin{array}{cccc}
SV & SV & ⇒ & SV & SV \\
\backslash & / \backslash & / \backslash \\
Voc & atr & Voc & Voc & atr & Voc
\end{array}
\]

Note that an analysis of this sort involves a certain degree of abstractness in the representations, particularly since it involves identical representations for /i/ and /e/ in systems where they behave as the unmarked vowel type. However, recall from the discussion in Chapter 3 that the actual phonetic placement of vowels in an inventory may be driven by considerations of phonetic distance and articulatory effort. This means that the phonetic realization of a particular height representation may also vary, given the relative prioritization of these considerations. In terms of the auditory distance metric of Flemming (1995), the vowel [e] is exactly equidistant from the vowels [i] and [a] on the F1 dimension. Thus, in Pasiego, the realization of this unmarked height representation as [e] achieves the maximum amount of distance from each vowel. In contrast, the Kikuyu-type system involves an additional vowel type (/ê/) and so the possible positions
for the unmarked vowel are more restricted. In Flemming's system, the vowels [e] and [i] allow for the same degree of distance between [i] and [e]. I tentatively assume that the preference for [i] for the realization of the unmarked vowel in Kikuyu is due to relative degree of effort in articulation.

To sum up this subsection, I have shown that the proposed representations coupled with constraints on different types of complexity allow for all the systems that we have discussed. These representations capture the generalizations regarding natural class patterning of segments in various coalescence and spreading processes. I have offered an explanation as to why segment types pattern differently in these processes depending on what type of inventory they appear in. In the next section, I describe the approach taken in Casali (1996) to account for these same generalizations.

6.5 Casali 1996: Inventory-based representations

6.5.1 The representations

Casali (1996) suggests that since the behaviour of segments changes with the shape of the inventory, feature specifications must change depending on the number of height distinctions in an inventory. Thus, similar to the proposal of this thesis, the same phonetic vowel may have a different representation in a different inventory. The feature specifications he proposes for the different height systems are given below.

---

34 That is, [e] is closer to [e] and further from [i]; [i] is closer to [i] and further from [e].
(52)
a. two-height system

<table>
<thead>
<tr>
<th></th>
<th>+high</th>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>-high</td>
<td></td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>low</td>
<td></td>
<td></td>
<td>a</td>
</tr>
</tbody>
</table>

b. three-height system (Yoruba type)

<table>
<thead>
<tr>
<th></th>
<th>+high</th>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>-high</td>
<td></td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>-atr</td>
<td></td>
<td>ε</td>
<td>Ī</td>
</tr>
<tr>
<td>low</td>
<td></td>
<td></td>
<td>a</td>
</tr>
</tbody>
</table>

c. three-height system (Kikuyu type)

<table>
<thead>
<tr>
<th></th>
<th>+high</th>
<th>+atr</th>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>-high</td>
<td></td>
<td></td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ε</td>
<td>Ī</td>
</tr>
<tr>
<td>low</td>
<td></td>
<td></td>
<td></td>
<td>a</td>
</tr>
</tbody>
</table>

d. four-height system:

<table>
<thead>
<tr>
<th></th>
<th>+high</th>
<th>+atr</th>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>-high</td>
<td></td>
<td></td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>-atr</td>
<td></td>
<td></td>
<td>ε</td>
<td>Ī</td>
</tr>
<tr>
<td>low</td>
<td></td>
<td></td>
<td></td>
<td>a</td>
</tr>
</tbody>
</table>

Casali points out that it is not enough to simply state the featural representations of the different segments within a particular system. Instead, an explanation is required for why these representations are just these and not some other possible combination of features. Casali suggests that the representations are determined by a set of constraints on feature specifications and cooccurrences. Together these constraints work to ensure that only the attested systems occur and not any other of the logical possibilities.
6.5.2 Three-height systems

First, Casali suggests that vowel height is composed of two types of features: Primary features ([+high], [-high], and [low]) and Secondary features ([+atr] and [-atr]). A given vowel may bear at most one primary and one secondary feature. One generalization that Casali attempts to capture through constraints is that, within any inventory, all vowels have a primary feature specification while only some vowels bear a secondary feature. The constraint that drives this tendency, he suggests, is a constraint barring secondary specifications, *Secondary, given in (53).

(53) *Secondary (Casali 1996: 144)
Avoid secondary height features.

This constraint is violated by every vowel bearing a secondary height specification in a system. Primary features alone will serve to distinguish high vowels from mid vowels ([+high]/[-high]), and mid vowels from low vowels ([ ]/[low]). Since contrastive height levels must be distinguished, a system requiring more than these three distinctions (i.e. a three or four-height system) will require at least one secondary feature specification. However, since violation is to be minimal, the constraint *Secondary rules out any three-height system making use of more than one secondary specification.

---

These are not in fact the terms Casali uses for the features he proposes. He proposes a set of height features defined in auditory-acoustic rather than articulatory terms. The actual features he uses are given below. Since he states that these features correspond directly to the Standard height features as below, I use the standard features for ease of discussion.

<table>
<thead>
<tr>
<th>Casali's features</th>
<th>standard height feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>[+high]</td>
</tr>
<tr>
<td>L</td>
<td>[-high]</td>
</tr>
<tr>
<td>XL</td>
<td>[low]</td>
</tr>
<tr>
<td>h</td>
<td>[+atr]</td>
</tr>
<tr>
<td>l</td>
<td>[-atr]</td>
</tr>
</tbody>
</table>
Thus, one consequence of the *Secondary constraint is that a system must make use of primary height features before using secondary height features to contrast vowels.\textsuperscript{36}

The *Secondary constraint, then, goes a certain distance in characterizing the set of possible feature specifications for inventories of different sizes. However, for three-height inventories, although there are only two possibilities posited by Casali (cf. (52b) and (52c), the *Secondary constraint allows for four different possibilities, as given in (54).

\begin{tabular}{llll}
(54) & a. & i & [+high], [+atr] & b. & i & [+high] \\
 & I & [+high] & & I & [+high], [-atr] \\
 & e/\varepsilon & [-high] & & e/\varepsilon & [-high] \\

c. & i & [+high] & d. & i & [+high] \\
 & e & [-high], [+atr] & & e & [-high] \\
 & \varepsilon & [-high] & & \varepsilon & [-high], [-atr] \\
\end{tabular}

According to *Secondary's feature minimizing requirements, (a) and (b) exist as equally good candidates for the three-height Kikuyu-type system: both utilize only one secondary feature specification to contrast the different vowels. However, Casali suggests that only (b) is possible in these systems. For three-height Yoruba-type systems, (c) and (d) fare equally well as far as fulfilling the *Secondary requirements, yet only (d) is posited for these type of inventories.

\textsuperscript{36}This is similar to the views expressed in both Goad (1993) and Rice and Avery (1993): [atr] is utilized only when additional height distinctions are required.
A second observation Casali draws from the feature specifications in (54) is that, for the two unattested systems, either the high vowels bear a [-atr] or the low vowel bears a [+atr] specification. Since [+atr] is associated with auditory raising while [-atr] is associated with auditory lowering, Casali suggests that there may be a constraint requiring that a secondary feature specification involve the same "directional tendency" as a segment's primary feature. His formulation of such a constraint is given in (55).

(55) Secondary Reinforcement (Casali 1996:146)
[-high] should be reinforced with [-atr]
[+high] should be reinforced with [+atr]

Satisfaction of these two constraints leaves us with two choices for vowel height representations in three-height systems: (a) the Kikuyu-type system in (a) and the Yoruba-type system in (d).

6.5.3 Four height systems

Unlike the three-height systems discussed above, a four-height system will minimally require two secondary feature specifications to contrast all four vowel heights. *Secondary will ensure that only two secondary specifications are used, and Secondary Reinforcement will ensure that [+high] vowels bear only [+atr] specifications, while [-high] vowels will bear [-atr]. Casali suggests that the only way of specifying vowels in a four-height system is as in (56b) (cf. (52c)). However, given the constraint Secondary Reinforcement, the system in (56a) is incorrectly chosen as optimal over (56b).

(56).

<table>
<thead>
<tr>
<th></th>
<th>a. i</th>
<th>[+high], [+atr]</th>
<th>b. i</th>
<th>[+high], [+atr]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>[+high]</td>
<td>I</td>
<td>[+high]</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td>[-high]</td>
<td>e</td>
<td>[-high], [+atr]</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td>[-high], [-atr]</td>
<td>e</td>
<td>[-high]</td>
</tr>
</tbody>
</table>
Casali suggests that the relevant generalization is that languages prefer to restrict the total inventory of features in a system, and therefore prefer to use the same feature specifications in both the [+high] and [-high] domains. This preference is formalized in a constraint Casali calls Secondary Harmony.

(57) Secondary Harmony (Casali 1996: 150)
    Secondary features used in the [+high] and [-high] regions should be identical.

This constraint, ranked above Secondary Reinforcement, will choose the system in (56b) over (56a) since it makes use of the secondary feature [+atr] only. However, there is third possibility that uses a minimal number of secondary feature types: one that uses only [-atr] as a secondary feature. Notice that the system in (56c) also satisfies Secondary Harmony.

(56)
   c.  i  [+high]
   I  [+high], [-atr]
   e  [-high]
   e  [-high], [-atr]

How can a system using only [-atr] be ruled out? Casali addresses this problem by splitting up Secondary Reinforcement into two separate constraints, as in (58).

(58)  a.  [+atr] Reinforcement    [+high] should be specified as [+atr]
     b.  [-atr] Reinforcement    [-high] should be specified as [-atr]

Since [+atr] is always used in four-height systems instead of [-atr], the relative ranking of these constraints is universally fixed (a) >> (b). Notice also that, since there is only one possibility for a four-height system (i.e. the one in
(56b)), the relative ranking of all four constraints needs to be universally fixed as in (59).


This same constraint ranking also works for the three-height systems.

6.5.4 Problems with Casali's account

Despite the fact that Casali attempts to explain rather than simply describe the differences between different inventory types, there is nonetheless an arbitrariness to his account. In addition, the specifications he proposes fail to capture certain natural class behaviours demonstrated in different inventory types.

Implied in his constraints and the definition of his features is a feature geometry: [atr] specifications are effectively dependents of height specifications by virtue of their not being permitted in a system until all height features are instantiated. Couched in the fixed ranking he posits for the constraints is a descriptive generalization: [-atr] is never an active feature in a four-height system. The constraint ranking ensures this. Since Secondary Harmony is relevant only to four-height systems (given that only four-height systems require two secondary specifications), then only in four-height systems will a decision have to be made whether to have two [-atr] specifications or two [+atr] specifications. The ranking [+atr] Reinforcement >> [-atr] Reinforcement ensures that [+atr] will always be chosen in these systems.37 The postulation of this fixed ranking in effect amounts to simply stating that four-height systems use [+atr] and not [-atr]. In the account I have

---

37 Casali himself notes the arbitrariness of this fixed ranking:

Although splitting Secondary Reinforcement into h-Reinforcement and l-Reinforcement does not seem unreasonable, it is not clear what functional motivation, if any, underlie <sic> the fixed ranking h-Reinforcement >> l-Reinforcement upon which the success of this approach crucially depends.  (Casali 1996: 152)
proposed this generalizations is derived through general principles of contrast and structure minimization.

In addition to the problem of the unexplained fixed ranking of constraints, Casali's proposed feature specifications in (52) fail to account for different natural class behaviours in the different systems. The problem basically stems from the treatment of the low vowel /a/. Under Casali's system, this vowel bears a primary feature [low]. No secondary features will be assigned to low vowels since they are not required for the purposes of contrast. However, recall that coalescence patterns in the three-height Yoruba-type systems indicate that the vowels /ɛ/, /ɔ/, and /a/ form a natural class. All of these vowels seem to pattern as [-atr] vowels. This patterning is unexplainable if there is not feature that is common to both /ɛ/ and /ɔ/, and /a/. In contrast, recall that in four-height systems, /a/ does not form an obvious natural class with the mid vowels /ɛ/ and /ɔ/. In Casali's system of specification, the relationship between between /ɛ/, /ɔ/, and /a/ is no different in a four-height system than in a three-height system: in both cases, there is no feature specification that is common to all three.

In summary, I have argued that the representations proposed in Casali (1996) cannot explain why /ɛ/, /ɔ/, and /a/ pattern as a natural class in a three-height system but not in a four height system. In addition, Casali's constraints offer no principled non-stipulative explanation for why [-atr] is active only in a three-height system.

6.6 Richness of the Base and output Inventories

Recall from the discussion in §5.2.2 that the standard view of segmental inventories in Optimality Theory assumes the principle of Richness of the Base (Prince and Smolensky (1993), Smolensky (1996)).
Richness of the Base states that all languages have an input inventory that includes all possible segments, and that language-particular inventories are a property of the output. The output inventories are determined by the ranking of output constraints. This means that there is no language-specific inventory in the input and therefore no easy way to capture the relationship between different patterns of phonological activity and the contrasts in an inventory.

In addition to this problem is the one posed by certain types of coalescence which create non-structure-preserving output segments. I argued in §5.2.2 that the same configuration of constraints that determine the phonemic inventory of a language also determine the output segments of coalescence. This means that we should not expect the constraints determining the output of coalescence to create any segment types that are not part of the phonemic inventory of the language. However, as shown in §5.2.2, place coalescence in Korean does create a non-phonemic segment type. The fact that this segment type does not occur outside of predictable contexts (i.e. outside of coalescence or assimilation processes) is unexplained.

Height coalescence also provides a testing ground for the view that language-specific inventories are a property of the output. If the inventory of language is determined by the same constraint configuration as the output of height coalescence, then there should be no difference in the segment types permitted in the phonemic inventory and the segment types created by coalescence. Yet it appears that, as with place coalescence, height coalescence can create segments that are not part of the phonemic inventory.

One well-known example of this is found in Sanskrit coalescence. In Sanskrit vowel coalescence, a low vowel coalesces with a following high vowel to create a mid vowel (Whitney 1960). However, in the underlying inventory of Sanskrit there are no mid vowels, instead the inventory is {i, u,
Thus, the coalescence creates an output segment which is not contrastive, but occurs only in predictable contexts (i.e. under coalescence). A second example is found in Wuhan, a Mandarin language. In Wuhan, the addition of a diminutive suffix to a stem results in the coalescence of two underlying vowels to a single vowel (Zhou (1998)). The output vowel resulting from the coalescence takes the height of the first vowel, which is the nucleus of the non-coalesced form, and the place of the second vowel, which shows up as a glide in the non-coalesced form. This process creates surface [ɛ] and [ɔ], which are not found anywhere else in the language but as the output of coalescence.

(60a) Wuhan vowels

\[
\begin{array}{c}
i \\
e \\
a \\
u \\
o \\
\end{array}
\]

(60b) Wuhan coalescence

<table>
<thead>
<tr>
<th>Underlying</th>
<th>Coalesced form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ai + dimin. suffix -r</td>
<td>ɛ + r</td>
</tr>
<tr>
<td>au + dimin. suffix -r</td>
<td>ɔ + r</td>
</tr>
</tbody>
</table>

The constraints involved in determining output inventories are Faithfulness constraints and markedness constraints (P&S (1993)). Given a universal set of inputs, if a Faithfulness constraint requiring the preservation of [α] is ranked higher in a language than the markedness constraint barring [α], then that element should appear as part of the phonemic inventory. If, however, the ranking is reversed, [α] should not be a contrastive element in the language.

From the inventory of Wuhan phonemes, we may conclude that the constraints prohibiting the segments /ɛ/ and /ɔ/ are higher-ranked than the Faithfulness constraint(s) requiring that they be preserved in the output. Recall from §5.2.2. that constraints against individual segments are actually
cooccurrence constraints prohibiting the presence of some set of features on the same segment. Thus the segment /ɛ/ is ruled out by a constraint prohibiting low front vowels, *[Low, Cor]. In order for this constraint to rule out /ɛ/ as a phoneme, it must outrank the featural Faithfulness constraints (i.e. MAXF) which require the preservation of these features in the output. If Faithfulness outranks this cooccurrence constraint, /ɛ/ should be part of the phonemic inventory (as shown in (b) below).

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Output inventory</th>
</tr>
</thead>
</table>
| a. *[Low,Cor]>> FAITHplace>>*Low,*Cor | *low vowels, and front vowels  
*no low front vowels |
| b. FAITHplace>>*[Low,Cor],*Low,*Cor | *low vowels, front vowels, and  
low front vowels |

Thus, Wuhan has the ranking in (a), and /ɛ/ does not function as a phoneme in the language. However, as demonstrated in the data in (60), coalescence derives this segment (and the segment [ɔ]) in the output. As shown many times in previous sections, the output of coalescence is determined by Faithfulness constraints. In order for Faithfulness constraints to derive the output [ɛ], they need to outrank the cooccurrence constraint *[Low, Cor].

Thus, the same ranking paradox described in §5.2.2, arises here. Why does this ranking allow [ɛ] as a coalesced segment but not as a regular segment whose occurrence in a form is non-predictable?

In contrast to this approach, the current proposal views the inventory as a crucial part of the input: input inventories are language particular and not universal. This language-particular system guides the building of

---

38In actual fact, the cooccurrence constraint need only outrank one of the Faithfulness constraints preserving these features (e.g. either MAXCoronal or MAXLow). This means that an input [Low, Coronal] vowel would simply lose the feature whose MAXF constraint was outranked.
language-particular segmental representations, which are informed by output constraints (as described in §3.5).

6.7 Conclusion

This chapter has explored the role of inventory in determining segmental behaviour in height coalescence patterns. I have demonstrated that there is an important relationship between the presence of height and atr distinctions in a system and the phonological processes that we see. I have argued that the patterns fall out from the input representations of the different segment types. Given the divergent behaviour exhibited by a single segment type in different systems, I concluded that the input representations must vary from language to language.

These input representations are built in response to contrast and complexity. Two different types of complexity are defined: vertical and horizontal. These two types of complexity are governed by separate structure minimization constraints, and their relative ranking determines the type of contrasts that are made in an inventory.

I have also made proposals (following work by Goad (1993) and Rice and Avery (1993)) regarding the representation of height and atr distinctions which are based on the type of natural class behaviours that are seen in the different systems.

Deriving language-particular inventories from output constraints misses the link between inventory and segmental patterning. In addition, I have shown that the constraints responsible for determining the phonemic inventory are the same constraints that determine the outcome in (asymmetric) coalescence. This becomes problematic when coalescence inexplicably allows a different set of output segments than the phonemic inventory.
Chapter 7.

In this chapter I take the opportunity to reiterate the main points of the thesis and to summarize my theory of segmental markedness as well as my view of the status of input representations.

7.1 Markedness as Structure

I have argued from the perspectives of theoretical economy and empirical and explanatory adequacy that segmental markedness in Optimality Theory should be evaluated in terms of representational complexity. Segmental representations which are more complex are more marked, while representations that are less complex are less marked (Avery and Rice (1989), Rice and Avery (1993), Rice (1993, 1995), Goad (1993)). This theory subsumes the current view of segmental markedness in OT which evaluates markedness through a fixed hierarchy of substantive constraints.

The theoretical economy arguments are based on the fact that the markedness of representations throughout the phonology is already evaluated implicitly by structure minimization constraints. My proposal is to unify this structure minimization requirement as a single family of markedness constraints *STRUCTURE (McCarthy and Prince (1993), Myers (1994)). *STRUCTURE constraints penalize all structure at different levels of representation including subsegmental structure. This type of constraint obviates the need for the anti-epenthesis Faithfulness constraint DEP since input vs. GEN-supplied structure is already distinguished by MAX.

Having segmental markedness governed by *STRUCTURE instead of a universal hierarchy of markedness constraints makes segmental markedness less of an oddity within OT where universal preferences should be an effect of universal constraints instead of fixed rankings.
This view of segmental markedness also achieves empirical adequacy in accounting for cross-linguistic preferences for unmarked segments in TETU environments, as well as other markedness-related generalizations such as implicational relationships between segment types, and universal tendencies in inventory shapes. Unmarked segments are preferred wherever Faithfulness is not at issue because in such situations, *STRUCTURE determines the output. *STRUCTURE consistently prefers the simplest representation, which in this theory is the least marked form. Implicational relationships and inventory shape generalizations are also linked to the complexity of representations: an inventory allowing $n$ degrees of complexity also allows $n-1$ degrees of complexity in the inventory. As discussed in Chapter 6, the shape of inventories can also be linked to complexity and preferences for different types of structure.

One particularly positive aspect of this view of segmental markedness is seen in my account of the link between markedness and patterns of phonological activity. While the standard OT view of segmental markedness does not relate the activity of a feature to its markedness status, a representational account cannot help but make this connection. Complex specifications are marked because of the amount of structure in their representations. They are active because the presence of structure in their representations means that constraints can require that structure to spread, and Faithfulness can require it to be preserved.

7.2 Inputs are Players in the phonology

A major claim of this thesis is that language-particular inventories and segmental representations must be determined in the input. Despite claims in the OT literature that all cross-linguistic variation can be determined by
output constraints alone, language-particular inputs are required to account for inventory-driven variations in segmental markedness relations.

I have made several arguments that inventory plays a major role in determining segmental behaviour. For example, I have shown that it influences patterns of phonological activity in assimilation, vowel harmony, and coalescence. Segmental behaviour in my theory derives from input representations, and therefore the inventory shapes the segmental representations in the input. Since markedness is determined by the complexity of representations, inventory also influences markedness. This is an important result, since inventory is linked to the selection of unmarked segment types. For example, I have shown that in inventories with a front-central-back place contrast in vowels, the central vowel is unmarked. In inventories with only a front-back contrast, either may be marked.

This variation in segmental markedness is well-defined and predictably limited. In contrast, a view in which segmental inventories are defined in the output cannot make this same link between inventory and segmental markedness.

Furthermore, a view in which markedness is determined by a fixed hierarchy of markedness constraints cannot adequately restrict the range of variation seen in markedness relations. Instead, once variation in markedness constraint ranking is permitted, all and any variation should be possible.

I have also argued that output constraints cannot always accurately define the phonemic inventory of a language. Since the outcome of coalescence is determined by the same ranking of Faithfulness and markedness constraints that determine the inventory, we should not expect coalescence to allow a different set of segments than occur in non-predictable
contexts in the language. However, I have presented cases where this is exactly so.

7.3 (Under)Specification in OT

The theory that I have been arguing for makes use of minimally specified representations: *STRUCTURE allows only the minimum amount of structure necessary to distinguish contrasting elements. This means that some segments are left without specification if they are the least marked segment type.

In addition to the theory-internal motivation for underspecified representations is the motivation that comes from the patterning of unmarked elements in the phonology. The failure of unmarked elements to be preserved in the face of assimilation, coalescence, or harmony means that they are ignored by Faithfulness. Under an unbiased view of Faithfulness where input elements are treated as equal, the habitual failure of Faithfulness to preserve some element in the output is evidence of its absence in the input.
References


Awobuluyi, O. (1972) "The morphophonemics of Owon Afa." Research notes from the Department of Linguistics and Nigerian Languages, University of Ibadan 5.2:25-44.


Causley, T. (1997a) "Featural Correspondence and Correspondence constraints." Talk given at HILP 3. VU: Amsterdam.


