CHAPTER 3

ASYMMETRICAL INTERACTIONS AND ASYMMETRICAL CONSTRAINTS

3.0 Introduction

The three approaches to too-many-solutions problems discussed in the previous chapter do not treat them as symptomatic of significant generalizations about input-output mappings rather than outputs. Because this systematic challenge to OT is not recognized, previous attacks on these problems, apart from de Lacy's general proposal, see them as an assortment of individual difficult cases and analyses. Surface-oriented theories such as the P-map, whose predictions are closely tied to the phonetic detail of each particular case, face this difficulty in an especially acute form, unable to provide a unified account for a systematic pattern across various examples with vastly different phonetics. The aim of this chapter is to remedy the situation by offering a general analysis of non-surface-based generalizations that treats them as a systematic phonological problem.

The chapter is organized as follows. I begin by arguing against a potential objection that the generalization about input-output mappings in the domain of prosody-segmental interactions is not a synchronically active one and should not be dealt with by a theory of UG. Concluding that it is a significant phonological generalization, I then discuss the place of non-surface generalizations in a surface-oriented theory. I then propose a new class of markedness constraints that refer directly to input-output mappings and penalize candidates that involve typologically undesirable processes. I then discuss a new proposal for dealing with sonority-driven stress, and conclude the chapter with a brief discussion of some applications of my theory, setting the stage for a detailed investigation of syncope and epenthesis in chapter 4.

3.1 Is the generalization synchronic?

The shortcomings of the previous approaches discussed in the last chapter, I suggest, are due to a general property of OT: its radical claim that all significant phonological generalizations are located in the character of output forms. The data presented in this dissertation suggest that there are phonological patterns that are not attributable to purely surface markedness pressures. I take it as an argument against the radicalism of OT and in favor of a theory of phonology that affords both kinds of generalizations their place. However, such a claim does not automatically follow from the observed existence of typologically observable procedural generalizations, because any phonological generalization can potentially be synchronically inert but arise through external factors such as diachronic tendencies. If this is the case, a synchronic theory of linguistic competence is not responsible for dealing with such a generalization. Thus, before I turn to developing the mechanism of OT constraint interpretation that permits the theory to account for non-surface-based generalizations, it is necessary to show that the relevant claims about typology constitute real synchronic generalizations that the theory should be able to address.

Theoretical linguistics has long entertained the idea that not all observed generalizations can be due to innate UG-based properties of grammar (Blevins 2004). It is not the case that every typological observation about language reflects innate principles; gaps in the typology do not automatically call for UG-based explanation. There are systematic classes of patterns in languages that are absent for other reasons: because they are difficult to learn or to process, or because there is no conceivable diachronic path that leads to them. Just how much of grammar can be explained externally is, of course, subject to vigorous debate in linguistics. I will steer clear of this controversy and adopt the view that is mainstream at least in current phonological theory – that AT LEAST SOME properties of grammars have an external explanation, and that there are independent ways of telling whether a particular generalization is a true universal or an accidental fact that can be explained without invoking innate properties of grammar. This dichotomy between two kinds of generalizations is most clearly laid out in Kiparsky 2004, who proposes several ways of deciding whether a given systematic observation about grammars is a UG-based 'universal' or a 'typological generalization' that can be explained without reference to UG.

If it can be shown that generalizations such as the asymmetries in prosody-segmental interactions are not true universals but arise due to external factors like diachrony, they would fall outside of the responsibility of a synchronic theory of phonology like OT. The question, thus, is whether the relevant facts fall in the category of 'universals' or 'typological generalizations' in the sense of Kiparsky 2004.

Applying several of Kiparsky's criteria to prosody-segmental interactions shows that the generalization cannot be dismissed as spurious. The claim made here is that it is real in the sense that it must be expressed in a theory of the knowledge of language.

Kiparsky argues that true universals are IRREVERSIBLE in the sense that language change cannot subvert them. For example, the sonority hierarchy can be argued to be part of UG on the grounds that sonority-based generalizations are not reversed by potential sound changes. There is a common tendency in languages for the high-sonority vowel *a* to attract stress more strongly than other vowels. This fact follows from the assumption that the sonority hierarchy plays a role in metrical prominence. There is a common sound change by which *a* raises to *ə*, a vowel that is lowest on the sonority scale. If this sound change occurs in a language where stress is attracted to every *a*, then the resulting language would have a rule by which it is the schwa – the least sonorous vowel – that is the most prone to be stressed. The claim is more subtle than a simple assertion that such a sequence of events cannot occur. Rather, Kiparsky argues, whenever it does occur, the result will be, from a synchronic point of view, an arbitrary stress system with stress encoded in the lexicon, rather than a system based on the productive generalization that the vowel *ə* is obligatorily stressed. In other words, the sonority hierarchy is a true universal enshrined in UG because it constrains language change.

This criterion of IRREVERSIBILITY indicates that the asymmetry between prosody-segmental interactions discussed above in this dissertation is a true universal rather than a typological generalization. The argument is similar in structure to Kiparsky's argument for the irreversibility of the sonority hierarchy. There is a reasonably likely path of change that can lead to a system that violates the generalization about prosody-segmental interactions, but this path does not seem to be taken by languages in a way that leads to a synchronic system contradicting the proposed generalization. As explained in detail in Chapter 1, there is a common tendency for certain consonantal features to be sensitive to stress. This tendency leads to a synchronic state where the location of the consonantal feature can be predicted in terms of the location of stress. From this starting point, the change that is required for a language to develop a system where stress is attracted to that consonantal property is a RULE REVERSAL reanalysis of the system. If the learners misinterpret the coincidence of stress and some other feature like aspiration as a result of stress attraction rather than aspiration attraction, the result would be a system with aspiration-driven stress. However, speakers appear not to mislearn stress systems in such a way.

Let me illustrate this point with Karo, the language discussed in Chapter 1 as a potential counterexample to the generalization. Recall that Gabas's (1998) analysis of the stress system, outlined in (1) below, made use of stress repulsion from a syllable with a voiced or lenited segment (18)c.

(1) a. Assign stress to the syllable with H tone;
b. Else to the syllable with the nasalized vowel;
c. Else to the penultimate syllable if the final syllable begins with [b], [g], or [ɾ];
d. Else to the final syllable.

Although stress is normally final, it shifts to the penult just when the final syllable's onset is a voiced stop or a [ɾ], as illustrated by the two representative items below.

(2)	a. /pako/	→	[pakó]	'pacu'
	b. /parat/	→	[páɾat]	'curimba'

However, I argued in Chapter 1 that the analysis that requires stress repulsion from syllables with voiced onsets is mistaken. A better analysis makes use of lenition in onsets of unstressed syllables. The main argument was that such a process is independently necessary to account for the alternations at the prefix-stem boundary. On my analysis, forms like those in (2) have unpredictable stress, penultimate or final, and lenition applies to onsets of unstressed syllables to derive flaps and voiced segments from voiceless ones in the ordinary way.

This example illustrates that the relationship between stress and consonant quality in Karo is potentially ambiguous, and thus might have caused the speakers to mislearn it as lenition-driven stress rather than as stress-driven lenition. And yet, the rule reversal path was not taken. This appears to hold in general: even when linguists might be tempted to posit a reverse interaction between segments and stress, speakers do not do so, because a synchronic principle of UG blocks a likely diachronic path. I take this as an argument in favor of the reality of the generalizations I argued for in Chapter 1.

The argument, once again, must be more subtle than simply saying that the particular diachronic path does not occur. Rather, the diachronic sequence of events is precluded by UG from creating a SYNCHRONIC state that flouts the proposed generalization. A rule reversal would, by definition, create a synchronic system that violates the claim of unidirectionality in prosody-segmental interactions. There are other conceivable paths by which a system like that in Karo can fall apart; the claim is that none of them lead to a situation where stress is driven by flapping.

The second relevant criterion for distinguishing universals from typological generalizations is CONVERGENCE. If a generalization arises as an accidental effect of diachronic tendency, it has only one diachronic source. True universals, on the other hand, should be observed to emerge from a variety of converging diachronic pathways. This criterion also suggests that the generalization about stress-segmental asymmetries is a true synchronic universal. The synchronic situation where stress conditions the location of a consonantal property can arise through at least two ways diachronically. First, consonantal features such as aspiration can develop in prosodically strong positions, as happened in English. Second, prosodically strong positions can protect aspiration from deleting (Vijayakrishnan 1999).

The third argument in favor of the stress-segmental asymmetries being true universals is the fact that they emerge outside of the grammar proper. I refer the reader to Section 3.5.2 below, where psycholinguistic evidence for the special status of prosody in grammars is discussed.

I take the arguments above to indicate that phonological theory cannot be released from the responsibility to deal with the process-based generalization about the interaction of prosodic structure and segments.

3.2 Locus of the generalization

Assuming that there is a generalization to explain, let me now detail the place of asymmetric phonological interactions in an output-oriented theory. This discussion will set the stage for the formal analysis in subsequent sections.

A generalization about processes is a claim that a given marked structure is repaired in fewer ways than the freely interacting, output-oriented constraints that OT would make one expect. In the context of discussing a too-many-solution problem in the domain of word-final devoicing, Steriade points out that the problem can be viewed as the mirror-image of the conspiracy argument first suggested by Kisseberth (1970), and later given prominence in the OT literature. Just as, from the point of view of a derivational theory, rules "conspire" to create a particular output structure, from the point of view of OT constraints "conspire" to produce only some input-output mappings but not others. Steriade's suggestion, reproduced in the quote below, will be further developed in Chapter 4, in the discussion of the contexts of vowel syncope.

> "Kisseberth's (1970) insight that conspiracies arise when the sound system aims at a specific target structure via multiple means can lead one to ask the same question, in the context of rule-based phonology: if the rule of final devoicing aims to eliminate final voiced obstruents, why aren't there rules of final obstruent nasalization, deletion, metathesis or post-voiced obstruent epenthesis?" (Steriade 2001: 6).

As discussed above in Chapter 1, the too-many-solutions problem in the case of prosody-segmental interactions arises because misalignment of certain prosodic and segmental categories, such as stress and aspiration, is repaired only by modifying the segments but never by modifying the prosody. Following Steriade's suggestion, then, from the point of view of OT this can be thought of as a conspiracy: OT constraints conspire to allow one but not the other among the potential repairs for the marked structure.

Steriade's approach to the too-many-solutions problem (Section 2.3) was to elaborate the faithfulness apparatus by introducing fixed rankings, grounded in the theory of similarity. The standard OT setup of input-output faithfulness constraints freely interacting with each other and with the markedness constraints is incapable of dealing with the process-based conspiracies. "The P-map's broadest claim", she writes, "is that the range of systematic, cross-linguistically invariant differences [in perceptual similarity that give rise to the too-many-solutions problem go] beyond the capabilities of current theories of correspondence" (2001: 6). We can add to this outlook on faithfulness a similar pessimistic view of standard OT markedness.

In general, in order to determine whether a candidate violates an OT markedness constraint, one is not required to look at any information outside of the candidate itself. Conventional markedness constraints are statements that hold of particular output forms, regardless of their derivational provenance, and of any other property of the grammar in question. The basis of the innovation proposed in this dissertation is the argument that procedural generalizations, if they are to be expressed in markedness constraints, require those constraints to access information outside of the surface form.

To take a specific example, given the situation where stress and aspiration are misaligned, an output constraint stating simply that aspiration should be located on the same syllable as stress is not enough. In order for a markedness constraint to determine whether a given surface form is the result of a licit repair strategy, the constraint must have access not only to the structure of that surface form, but also to the derivational provenance of the relevant properties in it. This means that, when faced with an output where stress and aspiration cooccur on the same syllable, such as a form like *pit^h^á*, the constraint must be able to distinguish whether both properties have been inherited from the input, or whether one has been attracted to other. Obviously, this information is not available solely from the structure of the candidate itself, which contains no clues about whether or not stress has been attracted to aspiration. That information can be obtained only by looking at the rest of the grammar, i.e. at the ranking of the other constraints governing the distribution of stress. Specifically, in order to see if stress has been attracted to aspiration in a form like *pit^h^á*, we need to know how stress would have behaved had aspiration not been present. In other words, if an OT markedness constraint is to be endowed with the power to penalize unwanted repairs, its violation profile must have access to the rest of the grammar.

Let me illustrate this crucial point in more detail. If there is a markedness constraint that is violated just in case an undesirable process like aspiration-driven stress has taken place, identical candidates in different grammars would have different violation profiles of this constraint. Suppose there are two languages, *A* and *B*, which have different stress systems. Language *A* stresses the final syllable by default, while language *B* stresses the initial syllable. An input without aspiration like /pita/ would surface with final and initial stress in the two languages, respectively. Now suppose an input with aspiration, /pit^{h}a/, surfaces in both languages with final stress, *pit^{h}á*. In this case stress can be said to be attracted to aspiration in language *B*, because the presence of aspiration makes a difference for the location of stress. On the other hand, language *A* has no stress-driven aspiration, because the location of stress is the same in the forms with and without aspiration.

(3) Language *A* Language *B*

/pita/	→	*pitá*	/pita/	→	*píta*
/pit^{h}a/	→	*pit^{h}á*	/pit^{h}a/	→	*pit^{h}á*
(no attraction)			(aspiration-driven stress)		

This example illustrates that any markedness constraint that penalizes unwanted repairs such as stress shifts driven by aspiration must have access to information outside of the candidate under evaluation. The identical form *pit^{h}á* would incur a violation of such a constraint in language *B* but not in language *A*; what determines this assignment of a violation is the general character of the stress system in the language, which depends on the ranking of the remaining constraints.

This need to access the rest of the grammar in order to compute the violations of a constraint is a general property of process-based generalizations. Any theory that attempts to capture such generalizations by means of markedness constraints must endow those constraints with the ability to see outside of the candidate under evaluation. I will now propose a formal mechanism of constraint evaluation that allows markedness constraints access to such information, and thereby affords generalizations about processes their place in the theory.

3.3 Procedural constraints and the typology of repairs

Given the pessimism about the ability of standard OT constraints to directly express non-output generalizations, we might seek alternative routes to standard OT theorizing. The path I will take up in this dissertation attempts to preserve as many properties of OT as possible, but at the same time to endow the theory with the capacity to handle generalizations about properties that are not exclusively located in output structures. In this chapter I will develop one proposal that introduces a new type of constraint into the theory. These new constraints come with a mechanism that interprets their violation profile in a novel way, by accessing information outside of the candidate under evaluation. This move is intended as an exploration of what it would take for a theory to both preserve the parallel structure of evaluation and deal seriously with non-surface generalizations by elaborating markedness constraints.

A few words of caution are in order before I proceed with the analysis. A central claim of this dissertation is that there are certain phonological generalizations which are not in the output structures but in the input-output mappings. In order to afford such generalizations their place in the theory, I will take the most straightforward approach possible: I will introduce a new class of markedness constraints that directly refer to phonological processes and assign extra violation marks to undesirable ones. This is not the only possible strategy. Another approach would be to modify not the markedness but the faithfulness system. The P-map theory, for example, is, from the formal point of view, a theory of faithfulness constraints.

I will argue below that current views of input-output faithfulness cannot deal with the kind of asymmetrical interactions that have been the focus of my discussion. It appears that a very serious modification of either, or both, of the current theories of faithfulness and markedness is in order if procedural generalizations are to be accounted for. I consider the work in this dissertation to be an exploration of what it would take for a 'pure markedness' theory to achieve the stated goals; the possibility of a tenable 'pure faithfulness' approach remains open.

A procedural generalization of the type explored in this dissertation amounts to a prohibition of a certain way of getting from an input to an output. In OT terms, it means that there must be a mechanism for making certain candidates incapable of winning, i.e. making them perpetual losers. I will take the simplest possible approach: I will introduce constraints that directly penalize those candidates that involve undesired input-output mappings. The formal mechanism introduced in the remainder of the chapter will provide an algorithm for finding such candidates, and for ensuring that the violation pattern of the new constraints entails those candidates' perpetual loserdom. The strategy will be to attack the problem head on. The new constraints will state directly the directionality of interaction between two phonological categories, and assign a violation mark to each candidate that involves an undesirable interaction. The formal machinery developed below will provide a mechanism for identifying such candidates from among the candidate set.

Thus, the discussion begins by making more strict the notion of what a 'process' is in the context of OT. I will then formalize the intuitive understanding that constraints cause processes to apply by providing a way to determine which constraint causes which process. This discussion will serve to single out candidates that violate a given generalization about input-output mappings. Once these formal underpinnings are in place, I will introduce a new mechanism of assigning violation marks to candidates, penalizing those that involve undesired processes, and demonstrate how my proposal can account for the typological generalizations under discussion.

3.3.1 The notion of 'process' in OT

In rule-based phonology, the notion of 'process' is self-explanatory. Rewrite rules take one form as an input and return another form as an output. Each rule that applies in the course of a derivation corresponds to a process. The change that the rule inflicts upon a form is transparent in its statement. While the phonological grammar is thought of as a complex machine that turns inputs into outputs, the path between the underlying form and the surface form can be straightforwardly decomposed into elementary steps, each of which results from the application of a rule.

In OT, there is no primitive concept that is a direct counterpart to the stepwise derivations of the earlier theories.[12] There is only one well-defined process in OT, the input-output mapping, which is achieved in a single step. There is no notion of "modifying" an underlying form incrementally to arrive at a surface form. While in rule-based theory the question of whether a particular process has or has not applied in the

[12] McCarthy's theory of candidate chains (McCarthy 2006), an important recent development in OT, is a notable exception. Most relevant for the discussion here is that theory's requirement that every output be reachable from the input via a sequency of discrete markedness-improving steps. I leave for future research the question of how structure-building operations such as stress assignment, and prosodic

course of a given derivation can be straightforwardly answered, it is not directly answerable in an OT derivation. Let me return to the hypothetical example of the two languages in (3). The intuition is clear: stress is attracted to aspiration in language *B* but not in language *A*. In derivational terms, it is clear what would be responsible for this attraction: at some point in the derivational history of *pit*^{h}*á* in language *B*, there must be a rule that affects metrical structure in a way that is sensitive to the aspiration of some segment in the form. On the other hand, the metrical structure assignment rules of language *A* make no reference to such segmental information. Once again, there is no straightforward translation of this explanation into OT terms, because OT does not have any counterpart of stress-assignment rules. The task of this section is, then, to formalize the intuition about the different interactions between stress and segments in language *A* and language *B*.

Let me begin defining the notion of 'process' in OT with a concrete example of the Latin stress rule, gradually making the discussion more general and eventually arriving at an abstract definition. The Latin stress rule assigns stress to the penultimate syllable if it is heavy and to the antepenult otherwise. Descriptively, a moraic trochee is built as close to the right edge of the word as possible, without incorporating the final syllable, unless leaving out the final syllable would leave too little material to make a non-degenerate foot. This statement in terms of conflicting pressures translates into an OT analysis of this system involves the following constraints and ranking.

(4) a.	AL-FT-R	'All feet right'
b.	FT-BIN	'Feet are binary at some level of analysis'
c.	*μμμ	'Feet are not trimoraic'
d.	TROCHEE	'Feet are trochaic'
e.	NON-FIN	'The final syllable is not footed'
f.	CULMIN	'There is at least one foot in every PrWd'

structure in general, fit into the theory, and whether it can be used to constrain the candidate space sufficiently to rule out the unwanted interactions between prosody and segments.

FT-BIN, CULMIN, TROCHEE
|
*μμμ
|
NON-FIN
|
AL-FT-R

All of these constraints must rank above the prosodic faithfulness constraint MAX-HEAD, because stress in Latin is (almost) entirely predictable, so no matter whether some other syllable is marked as the head in the input, the output corresponds to the Latin stress rule.

Now suppose we add a constraint to the system that calls on stressed syllables to have aspirated onsets, ASPIRATE/σ́. What does it mean formally for this constraint to force or not to force a stress shift?

If, in Standard OT, such a constraint is ranked above the stress constraints in (4), and provided that the aspiration faithfulness constraint is also high-ranked, then stress will surface on any syllable whose onset is aspirated in the input. For example, the input /p^hilippus/ would surface with initial stress, **p^hílippus*, rather than with the output of Latin stress rule, *p^hilíppus*. Intuitively, the constraint ASPIRATE/σ́ has caused a stress shift because adding it to the grammar has changed the output pattern. If the constraint were not present, stress would have surfaced in its 'normal' position.

I take this intuition as the basis for the formal notion of stress shift. To determine whether a constraint *C* affects stress placement, we will compare the location of stress in the actual output to its location in the alternative grammar with the constraint *C* removed from the ranking. A stress shift will then be said to occur whenever stress is in a different place in the outputs of the two grammars.

In the remainder of this section, I will built up the notion of 'process' incrementally in four steps, as follows. I will begin by defining a correspondence relation between candidates that share an input, making precise the notion of being 'in a different place' in two outputs of two grammars (3.3.1.1). This will lead to a definition of STRESS SHIFT

(3.3.1.2). Having illustrated this particular case of the more general notion, I will then define the concept of DESIGNATED STATE of a phonological object (3.3.1.3), which will allow for a general definition of PROCESS (3.3.1.4).

OUTLINE FOR THE REST OF THE SECTION

- (a) Define correspondence between candidates that share an input
- (b) Define STRESS SHIFT
- (c) Define DESIGNATED STATE
- (d) Define PROCESS

3.3.1.1 Correspondence between candidates that share an input

Once again, the intuition to be formalized is that a constraint *C* causes a stress shift for a particular input if removing *C* from the ranking results in a different stress pattern.

We need to first pin down the concept of 'being in the same place' when we are talking about two candidates in two different grammars. The material in the two candidates must stand in some correspondence relation. Because the candidates share an input, and because each of them stands in a correspondence relation with that input, we can define cross-candidate correspondence transitively, via the input. I rely here on the concept of T-CORRESPONDENCE from McCarthy 2003, where it was defined for two candidates in the same grammar. I reproduce McCarthy's definition below.

(5) **T-CORRESPONDENCE** (Definition) (McCarthy 2003:8)
Let *cand1* and *cand2* be two candidates from input *inp*. Let *s1* be a segment (or other corresponding element) in *cand1* and *s2* be a segment in *cand2*. Then *s1* t-corresponds to *s2* iff *s1* corresponds to some segment *s-inp* in *inp* and *s2* also corresponds to *s-inp*.

This idea can be extended in the obvious way to two candidates from two different grammars, as long as they share an input: two pieces of phonological structure in two candidates stand in a 'cross-grammar' t-correspondence relation as long as they share a correspondent in the input.

This definition is straightforward enough when applied to those aspects of structure that have an input corresopndent. But, of course, epenthetic segments correspond to nothing in the input and hence, by the definition in (5), do not t-correspond to anything. Also outside of the scope of this definition are elements like syllable structure, which may not have input correspondents (Prince and Smolensky 1993, McCarthy and Prince 1997; McCarthy 1999).[13]

At the same time, there is a clear intuitive sense in which syllables in two output forms can correspond: the first syllable of *pʰílippus* from the example above is in some sense 'the same' as the first syllable of *pʰilíppus*, and ditto for the second and third syllables of those forms. This 'sameness', however, does not follow from the definition in (5), because, arguably, those syllables have no correspondents in the input and thus they do not t-correspond to anything. I will work around this problem by stipulating that two syllables t-correspond to each other if the segments that fill their nuclei t-correspond.

(6) T-CORRESPONDENCE (Extension)
Let *cand1* and *cand2* be two candidates from input *inp*. Let σ_1 be a syllable in *cand1* and σ_2 be a syllable in *cand2*. Then σ_1 t-corresponds to σ_2 iff the segments filling the nuclei of σ_1 and σ_2 t-correspond according to the definition in (5).

By this definition, clearly, the first syllables of *pʰílippus* and *pʰilíppus* from the same example stand in a t-correspondence relation, because the vowels in their nuclei come from the same input.

This definition entails that vowel epenthesis and deletion, as well as other syllable-adding operations, do not 'disrupt' t-correspondence between syllables: in the candidate with initial epenthesis *əpʰílippus*, the second syllable *pʰí* t-corresponds to the initial syllable of *pʰílippus* because their nuclei share an input and thus t-correspond. To see

[13] A strong argument can be made that there are in fact faithfulness constraints referring to syllable structure (see Kiparsky to appear). If this is so, then syllable structure does fall under the scope of (5), and requires no special treatment.

how vowel syncope affects t-correspondence, consider the following two candidates sharing an input (7). Assume that the vowels named V_i in the two candidates t-correspond. Then, by the definition in (6), the syllables σ_1 and σ_3 in the two candidates t-correspond, while σ_2 in (7)a t-corresponds to nothing in (7)b, because its nucleus V_2 has no t-correspondent there.

(7) a. $CV_1\ CV_2\ CV_3$ b. $CV_1C\ CV_3$

$\sigma_1\ \sigma_2\ \sigma_3$ $\sigma_1\ \sigma_3$

3.3.1.2 Stress shift

Recall that we are trying to pin down the intuition that an OT constraint causes a process affecting a piece of phonological structure if removing the constraint from the ranking produces a DIFFERENT result with respect to that structure. The notion of t-correspondence allows us to talk about phonological objects being 'same' and 'different' in two different candidates from two different grammars, so long as those candidates share an input. The words 'same/different' will be used here as shorthand equivalents of 'standing/not standing in t-correspondence'. In particular, we can say that the stress in the form *pʰílippus* is not in the 'same' place as the stress in *pʰilíppus*, because the stress-bearing syllables *pʰi* and *li* do not t-correspond.

This leads me to the notion of STRESS SHIFT, defined for some input and some constraint.

(8) **STRESS SHIFT** (Definition)
Given a grammar *G*, a constraint *C*, and an input /i/, *C* is said to force a STRESS SHIFT for /i/ if the location of stress is different in the optimal candidate for /i/ in *G* and the optimal candidate for /i/ in *G*' that is identical to *G* except that *C* has been removed.

In this sense, the constraint ASPIRATE/$\acute{\sigma}$ forces a stress shift for the input /p^{h}ilippus/, because the output in the grammar with this constraint is *p^{h}ílippus*, and the output in the grammar without this constraint is *p^{h}ilíppus*, and the stress-bearing syllables in the two forms are different, as explained above. The constraint *[+voi], on the other hand, does not force a stress shift, because its presence or absence in the system has no effect on the location of stress. Likewise, ASPIRATE/$\acute{\sigma}$ does not force a stress shift for the input /p^{h}aretra/ 'quiver', because the aspiration is on the same syllable that receives stress by the Latin stress rule (*p^{h}á.re.tra*), so the output in the grammar with ASPIRATE/$\acute{\sigma}$ is identical to the output in the grammar without ASPIRATE/$\acute{\sigma}$.

More generally, the constraint ASPIRATE/$\acute{\sigma}$, if high-ranked, forces a stress shift in an easily definable set of cases: those where the location of the aspiration is not on the syllable that is the head according to the Latin stress rule.

The statement in (8) defines the process of stress shift caused by a constraint, a particular case of the more general notion of 'process', to which I turn in the next two sections. Note that it is hopeless to define 'process' in terms of the relation between inputs and outputs, because of the principle of Richness of the Base. An input may have the stress on any syllable, e.g. on the last syllable in /p^{h}aretrá/, but in the output stress will still surface on the antepenult, whether or not the constraint ASPIRATE/$\acute{\sigma}$ is present in the system. We would not want to say that a stress shift has occurred in this case. What matters is the location of stress in the output relative to its DEFAULT location, not relative to some arbitrary input.

3.3.1.3 Designated state

Let me now generalize this notion of 'process' beyond stress. Informally, a markedness constraint forces some process if this constraint contributes something to the input-output mapping, i.e. if the result would have been different had the constraint not been present.

In the stress shift example, I have referred informally to the 'default' place of stress – its location in the output for a grammar with some constraint taken out. This 'default location' determines whether a constraint has forced a stress shift or not. In order to generalize the definition of 'process' beyond stress, it is first necessary to be more precise about this notion of 'default', and to extend it to phonological categories other than stress.

Any given phonological object, be it a feature, a prosodic constituent, a tone, and so forth, can be affected by a markedness constraint. An object has been affected by a constraint if that object would in some way behave differently if the constraint were not there – i.e. if its t-correspondent in the grammar with the constraint removed is not identical to it. "Differently" here can mean a number of things, depending on the nature of the object in question: it could refer to the value of a feature, or its location, or the association lines of a tone, and so forth.

In order to determine whether a markedness constraint *C* forces a process that affects a phonological object *p*, we need to compare *p* in the grammar with *C* to *p* in the grammar without *C*. I first define this baseline of comparison, the DESIGNATED STATE (or LOCUS) of a phonological object.

(9) **THE DESIGNATED STATE (LOCUS)** (Definition)
Let there be a grammar *G*, a constraint *C*, an input /i/and its output [o] in *G*, and a phonological object *p* in [o]. Then the DESIGNATED STATE (or DESIGNATED LOCUS) of *p* for /i/ and *C* in *G* is *p*'s t-correspondent in the optimal candidate in the grammar *G*' that is identical to *G* except that *C* has been removed.

Because it is often useful to think of the designated state as some priviledged location in a form, I will use the term 'designated state' interchangeably with 'designated locus'. There is no theoretical difference betweeen the two; both depend on the definition (9).

Put simply, the designated state of some object for a given constraint is how that object would behave if the constraint were not present. I postpone the discussion of whether the designated state is uniquely determined by the definition given in (9) until Section 3.3.4.

Let me illustrate this notion of designated state with some examples. First, repeating the pseudo-Latin example from above, the designate state of stress for the input /pʰilippus/ and the constraint ASPIRATE/σ́ is the stress in the optimal candidate in the grammar without ASPIRATE/σ́, i.e. the penultimate stress in *pʰilíppus*, the default location according to the Latin stress rule.

In Latin there is also a constraint against long vowels followed by other vowels, *VV.V. In general, long vowels shorten in this environment. The input /docē-ō/ surfaces as [doceō] 'I teach'; this is the output in the grammar *G* where the constraint *VV.V is present. If the constraint *VV.V were not present in the system (grammar *G*'), nothing would prevent /docē-ō/ from surfacing faithfully as [docēō]. The segment *e* in the output of grammar *G* t-corresponds to the segment *ē* in the output of the grammar *G*', because they share the input /ē/. Thus, the designated state of the segment *e* for the input /docē-ō/ and the constraint *VV.V is the long vowel *ē* in [docēō].

To take another example, there is a constraint against final voiced stops *[+voi]#, active in Russian, which has final devoicing, and inactive in English, which does not. The input /rod/ surfaces as *ro*[d] in English, and the similar input /rod/ 'gender' in Russian comes out as *ro*[t]. The feature [voi] has a designated state for the constraint *[+voi]# and the input /rod/ in both languages, which is the correspondent of that feature in the output of /rod/ in the grammars that differ from the actual grammars of English and Russian in that *[+voi]# has been removed from the system. In both cases, the designated state is the feature [+voi] on the final segment of the word, because in the absence of *[+voi]# nothing prevents the final consonant from surfacing faithfully as *ro*[d].

3.3.1.4 Process

The concept of designated state defines a baseline of comparison that allows us to see whether a given constraint has caused some process to apply to some form. It bears emphasizing that this baseline cannot be determined from the input alone, but must be supplied, as formalized in the definition above, from the output in the alternative grammar with the constraint in question removed. The input fails to provide the crucial clues about a form's default behavior due to ROB, as explained above.

Now with the notion of the designated state of a phonological object we can define what a process is in OT.

(10) PROCESS (Definition)
For a grammar *G*, a constraint *C*, an input /i/ and its output [o] in *G*, and a phonological object *p*, the constraint *C* is said to affect *p* if *p* is not identical to its designated state as to location or value. *C* is then said to force a PROCESS affecting *p*.

Under this more general definition, the stress shift forced by the constraint ASPIRATE/σ́ in the pseudo-Latin example above is the process forced by that constraint. An example of a segmental process would be flapping in English, forced by the constraint *[t/d]/WEAK. If this constraint were not present in the system, inputs like /atom/ would surface faithfully without flaps. So, the designated state of the feature [son] of the English word [aɾom] is the *t* in [atom], and thus there is a process of flapping forced by the constraint.

To summarize this section: I suggested a formal way of answering the question of whether a given constraint has forced some process for a given form. This concept rests on a comparison between the actual output with the output in the alternative grammar where the constraint in question is removed. Unlike the eponymous concept in rule-based theories, process in OT is a matter of the whole grammar. In order to determine what processes a constraint forces, we need access to the entire ranking. In the case of stress shifts, for example, we need the entire ranking of the constraints in the language to determine what the designated state of stress is – i.e. its 'default' location. Given this OT-based notion of process, we can begin talking about typological generalizations having to do with input-output mappings an OT markedness constraint can cause. In the next section I introduce a new type of OT constraint whose job is to penalize candidates that involve unwanted processes.

3.3.2 The Implicational Constraint Principle

In this section I introduce the new constraints that will bear the burden of explaining typological generalizations like those discussed earlier in this dissertation. The new constraints will state what is a preferred phonological process rather than what is the preferred output. I will refer to these new constraints as PROCEDURAL markedness constraints. Throughout the discussion, the names of the new type of constraints will be preceded by the sign ✦: procedural markedness constraints will have names like ✦M, while traditional output-oriented constraints will retain names like M.

The task of the new procedural constraints is to penalize unwanted interactions between phonological properties, and thereby to enforce generalizations about input-output mappings. These constraints encode processes, i.e. interactions between two phonological categories. Such constraints will be stated as IMPLICATIONS: they have the general form 'If *P* has property *x*, they *Q* has property *y*'. The formal mechanism I propose will link the asymmetrical statement of such constraints to the asymmetry in interaction between the two properties mentioned, in such a way that the property mentioned in the antecedent can influence the property mentioned in the consequent, but not vice versa. I will propose a way of interpreting such constraints that assigns extra violation marks to those candidates that involve a process affecting the property mentioned in the antecedent of the constraint. As I will show, the effect of my proposal will be that a procedural onstraint 'If *P* has property *x*, they *Q* has property *y*' will have the ability to force a process affecting *Q* but not *P*. This mechanism provides a general way of accounting for procedural generalizations using OT markedness constraints.

If a constraint of the form 'If *P* has property *x*, they *Q* has property *y*' is interpreted as a standard OT constraint, the the direction of implication in the statement makes no predictions about the direction of interaction between the relevant aspects of representation. The constraint is simply a statement about an output structure. In principle, any process that results in the implication being true can be a repair strategy for such a constraint. As I showed in Chapter 1, the constraint ASPIRATE/σ́ 'If a syllable is stressed, its onset is aspirated' can be satisfied by aspiration just as easily as by a stress shift – i.e. by a process affecting the antecedent property of the constraint just as easily as by a process affecting the consequent property. In a language with default initial stress, an input like /pitʰa/ can surface with stress on the second syllable to satisfy the constraint ASPIRATE/σ́. The following tableau, repeated from Chapter 1, illustrates the unwanted 'aspiration-driven stress' language. Assuming that stress in this hypothetical language is initial by default, the mapping /pitʰa/ → *pitʰá* shows that the default can be overridden just in case aspiration is present on a non-initial syllable.

(11)

			ASPIRATE/σ́	DEP-h	STRESS INITIAL	MAX-h
/pitʰa/		pítʰa	*!			
	☞	pitʰá			*	
		pʰíta		*!		*
		píta	*!			*

In the more general case, the constraint 'If *P* has property *x*, then *Q* has property *y*', where *P* is any structure (segment, syllable, foot, adjacent consonants, etc.), can be satisfied by shifting *x* as well as by shifting *y*. The following tableau (12) illustrates this too-many-solutions problem in the general case. In the input, the properties *x* and *y* are located on different syllables (as indicated by the subscripts). A procedural constraint states that if something has property *x*, then it must also have property *y*. Different faithfulness constraints are violated by unfaithfully mapping *x* and *y*. Then the factorial typology contains at least three patterns: the fully faithful candidate which violates the markedness constraints, and two unfaithful candidates which violate the two faithfulness constraints. In order to satisfy the procedural constraint, both a solution that modifies *x* and a solution that modifies *y* are available. As argued in Chapter 1, these two solutions are too many.

(12)

			If *P* has property *x*, then *Q* has property *y*	FAITH-*x*	FAITH-*y*
/$\sigma_x\sigma_y$/	☞	$\sigma_x\sigma_y$	*		
	☞	$\sigma\sigma_{x,y}$		*	
	☞	$\sigma_{x,y}\sigma$			*

My solution is to replace the constraints ASPIRATE/$\acute{\sigma}$ and others responsible for similar too-many-solutions problems with procedural versions. The violation pattern of these constraints will ensure that, for tableau (12), no candidate with non-default (non-initial) stress will be able to win, and, likewise, in the general case in tableau (12), the candidate $\sigma\sigma_{x,y}$ will not be a potential winner. Forms such as *pitʰá* in tableau (12) and $\sigma\sigma_{x,y}$ in tableau (12) will become perpetual losers.

The strategy I use is to attack the problem in the most straightforward way. The phonological generalization behind the losing status of candidates like *pitʰá* and $\sigma\sigma_{x,y}$ is best stated in terms of input-output mappings. Procedural constraints directly refer to those mappings. The Implicational Constraint Principle introduced below serves as the mechanism for assigning violation marks to procedural constraints. Constraints like '✦If *P* has property *x*, then *Q* has property *y*' will receive violation marks not only when the implicational statement is false, but also when the constraint affects the phonological property mentioned in its antecedent. The notion of 'process' in the following statement refers to the definition in (10) above.

(13) **The Implicational Constraint Principle** (henceforth **ICP**)
A candidate *c* violates a procedural constraint '✦If *P* has property *x*, then *Q* has property *y*' iff:
a. In *c*, *P* has property *x*, and *Q* does not have property *y*, **OR**
b. The constraint forces a process affecting *x* in *c*.

Let me unpack this definition. It relies on the notion of 'process' introduced in Section 3.3.1, in the definitions (9) and (10). By these definitions, when a constraint affects *x* in a candidate, the actual state of *x* differs from the designated state of *x*. The designated state, in turn, is the location (and value, if it is a feature) of *x*'s t-correspondent in the optimal candidate in the grammar with the implicational constraint taken out of the ranking. In other words, what (13) says is that an implicational constraint is violated not only by those candidates where the implicational statement is false, but also by candidates in which the antecedent property mentioned in the constraint is not in its designated state.

As implied by the definition in (13), all procedural constraints subject to the ICP are assumed to be binary – each candidate incurs either 0 or 1 violations of such constraints, even if both of the conditions in (13) are met. This is needed to ensure that the typological consequences of the proposal hold. As I will show below in Section 3.3.4, this way of assigning violation marks ensures that the unwanted candidates can never win.[14] Before I take up that formal argument, for the remainder of this Section and in Section 3.3.3 I illustrate how (13) applies to concrete examples.

Let us take the tableau (12), repeated below as (14), to illustrate the proposal.

[14] I leave for future research the question whether, in case where one of the conditions of the ICP is violated in two different loci in the same candidate, it still incurs only one violation of the constraint. What matters here is that violation of both of the conditions of (13) does not lead to two asterisks.

(14)

		ASPIRATE/σ́	DEP-h	STRESS INITIAL	MAX-h
/pit^{h}a/		pít^{h}a	*!			
	☞	pit^{h}á			*	
		p^{h}íta		*!		*
		píta	*!			*

As discussed at length above, the problematic predictions of OT arise due to the action of the constraint ASPIRATE/σ́, which trumps the stress constraint(s) and forces a non-default stress to be assigned to the form. We can now replace this ordinary OT constraint with its procedural counterpart, ✦ASPIRATE/σ́, that would penalize not only any mismatch between stress and aspiration, but also candidates where stress shifts are used to repair that mismatch.[15] In other words, the new procedural constraint ✦ASPIRATE/σ́ requires something more than just for stress and aspiration to be located on the same syllable. 'Stress' is the antecedent property of this constraint, the property that is subject to clause (13)b of the ICP. The designated state of stress is determined with reference to the winning candidate in the evaluation where ✦ASPIRATE/σ́ is not present. The tableau for this evaluation is given below in (15); it shows that the winner has stress on the first syllable. The constraint taken out of the ranking is shaded.

(15)

			✦ASPIRATE/σ́	DEP-h	STRESS INITIAL	MAX-h
/pit^{h}a/	☞	pít^{h}a				
		pit^{h}á			*!	
		p^{h}íta		*!		*
		píta				*!

Thus, the designated locus of stress for the input /pit^{h}a/ and the constraint ✦ASPIRATE/σ́ is on the first syllable. According to the ICP, any candidate where a

[15] I postpone until Section 3.5 on constraint grounding the general discussion of how to determine which constraints are ordinary OT constraints and which are procedural.

stress shift has occurred violates this procedural constraint. We have the tools to find such candidates: a stress shift has occurred in any item whose stressed syllable does not t-correspond to the initial syllable in (15). All such candidates now receive extra violation marks by the ICP. To make the tableaux easier to read, I mark any such violations, incurred by clause (13)b of the ICP, with the sign ✦. There is no theoretical difference between ✦ used as a violation mark and the regular * violations. The following tableau shows the ✦ASPIRATE/σ́ constraint with the new violation pattern.

(16) DS of stress: initial syllable

			✦ASPIRATE/σ́	DEP-h	STRESS INITIAL	MAX-h
/pit^{h}a/		pít^{h}a	*!			
		pit^{h}á	✦!		*	
	☞	p^{h}íta		*		*
		píta	*!			*

The candidates *pít^{h}a* and *píta* incur ordinary violation marks of the ✦ASPIRATE constraint, because they contain stressed syllables with an aspirated onset. The candidate *pit^{h}á* incurs a ✦ violation of the constraint ✦ASPIRATE/σ́ because the location of stress is different from its designated location. The winner now is *p^{h}íta*, a candidate with default initial stress.[16]

Let me emphasize that a procedural constraint that is subject to the ICP is not a typical OT constraint: its violation profile for a given candidate set depends on the ranking of other constraints in the grammar. The reason for this is that such constraints are meant to rule out certain processes that are unattested typologically, and the notion of 'process' in OT, as discussed at length above in Section 3.3.1, only makes sense with reference to the ranking of the constraints in the language.

3.3.3 Tudanca Spanish

In Chapter 1, I went over laxness harmony in Tudanca Spanish, a case where a process is sensitive to the boundaries of a prosodic constituent. Harmony-prosody interactions present the familiar too-many-solutions problem for OT. Here I return to those Spanish data to illustrate the operation of the ICP in excluding unwanted interactions. First, to recapitulate the facts: final high vowels are lax, and laxness spreads leftward until it reaches the stressed syllable. Examples, taken from Flemming 1994, are shown below, with capitalization indicating laxness. The forms in (17)a have a final high vowel [U], which is lax. All vowels preceding it, up to the stressed syllable, are also lax; that laxness does not spread beyond the stressed syllable is shown by the form [se(kÁlU)]. The forms (17)b show the corresponding alternants where the final vowel is not high, and therefore not lax, and (7)c shows laxness spreading in forms where the stress falls on a syllable other than the penult.

(17)	a.	(pÍntU)	'male calf'	b.	(pínta)	'female calf'
		(čÍkU)	'boy'		(číka)	'girl'
		se(kÁlU)	'to dry him'		se(kálo)	'to dry it'
	c.	o(rÉgAnU)	'oregano'			
		(pÓrtIkU)	'portico'			
		ra(kÍtIkU)	'rachitic'			

Recall that the OT analysis of these facts made use of the following constraints.

(18)	a.	AGREE[tense]$_{\varphi}$	'All vowels within a foot have same value of [tense]'
	b.	*[+high, +tense]#	'No final tense vowels'
	c.	STRESS	Cover constraint for penultimate stress
	d.	IDENT[tense]	

[16] Once again, the typological argument that no stress shift candidate can EVER win will be made in section

The constraint driving laxness harmony within the stress foot, AGREE[tense]$_{\varphi}$ (18)a, can be satisfied either by violating the segmental faithfulness constraint – i.e. by applying harmony – or by violating a prosodic constraint by moving the prosodic domain boundary to accommodate the segments. This is the usual situation where at least two repairs are predicted to exist, given a constraint that mentions two categories (a prosodic and a segmental one). In the actual language, the constraint (18)a forces a segmental repair: harmony applies within the stress foot. This is illustrated in the following tableau, repeated from Chapter 1. The form ending in a high vowel surfaces with laxness harmony, /sekalu/ → [sek(ÁlU)], while the form ending in a non-high vowel surfaces faithfully, /sekalo/ → [sek(álo)].

(19) Tudanca Spanish

			STRESS	AGREE[tense]$_{\varphi}$	*[+high, +tense]#	IDENT[tense]
/sekalu/		se(kálu)			*!	
		seka(lÚ)	*!			*
		se(kálU)		*!		*
	☞	se(kÁlU)				**
/sekalo/	☞	se(kálo)				
		seka(lÓ)	*!			*
		se(kálO)		*!		*
		se(kÁlO)				*!*

Just as in other prosodically sensitive harmony processes, in Tudanca Spanish reranking the STRESS constraints to the bottom of the hierarchy produces a stress system where the stress feet are built in such a way that harmony does not get a chance to apply. In such a grammar, the harmony constraint AGREE[tense]$_{\varphi}$ would be satisfied not by an unfaithful mapping of segments, as in the actual Tudanca Spanish, but by violating prosodic markedness. This hypothetical situation is illustrated in the tableau below. The form ending in a non-high vowel still surfaces with the default stress, /sekalo/ → [se(kálo)]. However, the form /sekalu/ has a final stress, [seka(lÚ)], because the default stress would violate either the harmony constraint if harmony does not apply

3.3.4. For now, the examples simply serve to illustrate the action of the ICP.

(*[se(kálU)]), or the constraint against final high tense vowels if tensing does not apply (*[se(kálu)]), or the faithfulness constraint militating against changes in the [tense] feature (*[se(kÁlU)]). This leaves the candidate with final stress as the only viable option, as illustrated below.

(20) Pseudo-Tudanca Spanish

			AGREE[tense]$_\varphi$	*[+high, +tense]#	IDENT[tense]	STRESS
/sekalu/		se(kálu)		*!		
	☞	seka(lÚ)			*	*
		se(kálU)	*!		*	
		se(kÁlU)			*!*	
/sekalo/	☞	se(kálo)				
		seka(lÓ)			*!	*
		se(kálO)	*!		*	
		se(kÁlO)			*!*	

Let us now see how my proposal rules out the system in (20). Intuitively, while the "job" of the constraint AGREE[tense]$_\varphi$ is to produce harmony, in the unwanted mapping /sekalu/ → [seka(lÚ)] it has caused something other than harmony, viz. a stress shift. The solution is to make AGREE a procedural constraint whose only job is to enforce harmony. In order to rule out the unwanted interaction, the constraint AGREE[tense]$_\varphi$ must be prevented from causing stress shifts, i.e. must mention the location of metrical structure in its antecedent. I propose the following form of the constraint.

(21) ✦AGREE[tense]$_\varphi$
'If V_1 and V_2 are in the same foot, then they have the same value of the [tense]'

The antecedent property mentioned in the constraint (21) is 'being in the same foot': this is the property which the constraint ✦AGREE cannot force to change. By the ICP, the constraint ✦AGREE is violated not only in candidates where there is no agreement within the stress foot, but also by any candidate where foot boundaries are not in their designated state.

In order to determine the designated locus of feet, we take the constraint ✦AGREE out of the ranking. This is shown in the tableau below. The winners for both of the inputs have penultimate stress, and thus a main-stress foot comprising the final two syllables.

(22) Pseudo-Tudanca Spanish

			✦AGREE[tense]$_{\varphi}$	*[+high, +tense]#	IDENT[tense]	STRESS
/sekalu/		se(kálu)		*!		
		seka(lÚ)			*	*!
	☞	se(kálU)			*	
		se(kÁlU)			*!*	
/sekalo/	☞	se(kálo)				
		seka(lÓ)			*!	*
		se(kálO)			*!	
		se(kÁlO)			*!*	

The vowels *a* and *U* are not in the same foot in [seka(lÚ)], while their t-correspondents *a* and *U* in [se(kálU)] are. The designated state of these two vowels is 'being in the same foot'; any candidate where these two vowels are separated by a foot boundary thus incurs a ✦ violation by the ICP (as would any other candidate whose footing pattern differs from the designated state). Now, using this information, we can assign violation marks to the constraint ✦AGREE. The following tableau illustrates this for the input /sekalu/; the unwanted candidate is fatally penalized.

(23) Designated locus of stress: penultimate

			✦AGREE[tense]$_{\varphi}$	*[+hi, +tense]#	IDENT[tense]	STRESS
/sekalu/		se(kálu)		*!		
		seka(lÚ)	✦!		*	*
		se(kálU)	*!		*	
	☞	se(kÁlU)			**	

The new violation pattern does not affect the winner for the input with a final non-high vowel, /sekalo/, as the following tableau illustrates. Once again, the pathological candidate [seka(lÓ)] gets an extra ✦ violation mark because its stress foot is different from its designated state; this has no effect on the outcome.

(24) Designated locus of stress: penultimate

			✦AGREE[tense]$_{\varphi}$	*[+hi, +tense]#	IDENT[tense]	STRESS
/sekalo/	☞	se(kálo)				
		seka(lÓ)	✦!		*	*
		se(kálO)	*!		*	
		se(kÁlO)			*!*	

The discussion so far focused on the hypothetical language, similar to Tudanca Spanish except that the harmony constraint was able to force a stress shift under the standard OT analysis. Returning now to the actual language, let us make sure that the new violation pattern of the ✦AGREE constraint does not change the outcome. The designated locus of stress for the ✦AGREE constraint for both of the inputs, /sekalu/ and /sekalo/, is penultimate. No tableau should be necessary: because the STRESS constraints are high ranked, no candidate with a stress other than in the penultimate position will be optimal, regardless of whether or not ✦AGREE is taken out of the ranking.

Now the ✦AGREE[tense]$_{\varphi}$ assigns extra ✦ violation marks to any candidate with stress other than in the penultimate position, for both of the inputs. Since any such candidate is defeated by the higher-ranked STRESS, the new violation marks will not affect the outcome. The following tableau illustrates.

(25) Tudanca Spanish. DS of stress: penultimate

			STRESS	✦AGREE[tense]$_{\varphi}$	*[+high, +tense]#	IDENT[tense]
/sekalu/		se(kálu)			*!	
		seka(lÚ)	*!	✦		*
		se(kálU)		*!		*
	☞	se(kÁlU)				**
/sekalo/	☞	se(kálo)				
		seka(lÓ)	*!	✦		*
		se(kálO)		*!		*
		se(kÁlO)				*!*

To summarize this section: I returned to the Tudanca Spanish data from Chapter 1, and showed that the new violation marks assigned by the ICP prevent the unwanted stress shift candidates from winning. So far, the claim is not typological; I have merely shown that a particular ranking of the constraints does not produce a pathological winner. In the next section I move on to the broader consequence of my proposal, and show that in fact ANY candidate that incurs a ✦ violation mark is a perpetual loser. This will guarantee that the typological generaliztaion about prosody-segmental interactions is accounted for, and will provide a general mechanism for handling procedural generalizations in OT.

3.3.4 Candidates that incur ✦ violations are perpetual losers

In this section I offer a formal demonstration that candidates that incur ✦ violations by clause (13)b of the ICP cannot be optimal. This argument is crucial to the typological claim of my theory. Recall that the candidates that incur ✦ violations are those that involve undesirable input-output mappings that procedural constraints are designed to penalize. Thus, if no such candididate can win, the constraints make a typological prediction about processes, and we have a general way of accounting for generalizations about input-output mappings in OT.

My argument here will be made in three steps, as follows.

- **Step 1**: The 'designated state' is determined uniquely for each constraint.
- **Step 2**: For systems with only one procedural constraint, candidates with ✦ violations are perpetual losers.
- **Step 3**: Adding more procedural constraints to the system does not compromise the typological results.

I begin by taking up a question from Section 3.3.1 that I did not address there: is the designated state uniquely determined by the definition (9)? Let us assume there is a grammar *G* with procedural constraint ✦C, and the phonological category mentioned in the antecedent of the constraint is *P*. Does the grammar *G*', identical to *G* except that ✦C has been removed, uniquely determine the behavior of *P*? In order for my theory to get off the ground, the answer to this question must be affirmative. I argue here that it is.

To show that the grammar *G*' uniquely determines the behavior of the property *P*, it is sufficient to demonstrate that any two candidates that differ in the location or value of *P* have different violation profiles in *G*'. Having a different profile of violations means that the grammar distinguishes these two candidates, and that therefore they cannot both be equally harmonic.

Note first that ✦C, by definition, is a markedness constraint. Grammars *G* and *G*' do not differ in their sets of faithfulness constraints. As long as there are faithfulness constraints referring to *P*, the violation profiles of any candidates differing in *P* can be distinguished by faithfulness alone – and, a fortiori, can be distinguished by *G*'. This means that if *P* is a segmental feature or a piece of prosodic structure to which faithfulness can refer (such as stress), its designated state is uniquely determined by *G*'.

However, there is one important prosodic property which arguably is not subject to faithfulness constraints, namely syllable structure. In these cases, I suggest, the designated state can be determined by markedness alone. I will argue below and in Chapter 4 that the syllable structure markedness constraints ONSET, NOCODA, and *COMPLEX are not procedural. Thus, they are present in any grammar that identifies a designated state of some property. In order for syllable structure to be identified uniquely by such a grammar, it is sufficient that any two candidates that differ in syllable structure incur a different set of violations of these three constraints. This appears to be true, though I have no proof.[17]

[17] It is in fact possible to design pairs of candidates that differ in syllable structure but incur the same set of violations of the three constraints, e.g. [pa.trat.ra] vs. [pat.ra.tra], each incurring one violation of NOCODA and *COMPLEX. However, such candidates have a bounding set (Samek Lodovici & Prince 1999) containing the 'consistent' candidates [pa.tra.tra] and [pat.rat.ra]. This issue needs further exploration.

Now I am ready to proceed to Step 2 of the argument, viz. that, for grammars with only one procedural constraint, candidates incurring ✦ violations of that constraint are perpetual losers. One more definition is necessary here. In the discussion of the definition of 'process' above I referred to the 'designated state' of some phonological property for an input and a constraint as the state of that property in the optimal candidate in a grammar where the constraint in question has been taken out. Now we will need a convenient way to refer to that candidate in the alternative grammar; I will simply call this the DESIGNATED CANDIDATE (henceforth DC).

(26) **THE DESIGNATED CANDIDATE** (Definition)
For a grammar *G*, a constraint *C*, and an input /i/, a candidate is called the DESIGNATED CANDIDATE (DC) of constraint *C* for /i/ if it is the optimal candidate for that input in the grammar *G′* that is identical to *G* except that *C* is removed from the constraint set.

I leave open the question of whether the designated candidate is uniquely determined by *G*'. What is crucial is that, by the argument given above, each designated state is uniquely given by *G*'. This means that if *a* and *b* are two different designated candidates, they do not differ with respect to the property mentioned in the antecedent of the constraint *C*.

Suppose there is a procedural constraint '✦If *P* has property *x*, then *Q* has property *y*' and an input /i/. There is a certain designated state for *P* and some (possibly non-unique) DC for that input. The set of candidates *CAND* can be divided into two non-overlapping subsets. First, there are those candidates that have the property *P* in the designated state. All DCs belong to this set; there are others as well. Let us call this set *D*. Let *N* be the complement set of *D* in *CAND*. The set *N* includes all candidates that do not have *P* in the designated state. All and only members of *N* incur ✦ violations by the procedural constraint. The task is to show that no candidate belonging to *N* can be optimal.

Let us assume that constraints are functions that take a set of candidates and return a non-empty subset of that set (Tesar & Smolensky 2000, Samek-Lodovici & Prince 1999). Let *C* be the function corresponding to our implicational constraint, *H* be the function corresponding to all constraints higher ranked than *C*, and *L* be the function corresponding to all the constraints ranked lower than *C*. (*H* and *L* are compositions of the functions corresponding to individual constraints). If *C* is highest-ranked (lowest-ranked), let *H* (*L*) be the identity function.

By definition, the DCs are the winners in the system which has *H* and *L* but not *C*: the procedural constraint has been removed from the set, and the winner is determined by the remaining constraints. Those winners are given by the expression

(27) $\mathrm{DC} = L(H(\textit{Cand}))$.

First we apply the function *H* corresponding to the constraints ranked higher than the implicational constraint, and then the function *L* corresponding to all lower-ranked constraints.

The winner of the entire evaluation, with the implicational constraint put back into the ranking, is given by the expression

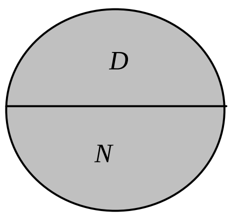
(28) $w = L(C(H(\textit{Cand})))$.

The application of *H* to the full candidate set yields the set $H(\textit{Cand})$; this is the set of candidates passed down to the lower ranked constraints. This set contains at least one member of *D*, namely the designated candidate *DC*, and possibly other members of *D*. It may also contain some members *N*.

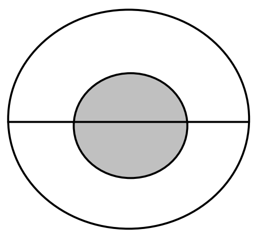
The function *C* applies to this set. Because implicational constraints are, by definition, binary, there are two possibilities:

(29) a. Some remaining candidates incur violations of *C*, some do not, OR
b. All remaining candidates incur an equal number of violations of *C*.

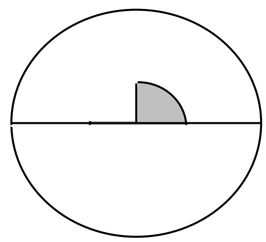
Let us consider the case (29)a first. By the *ICP*, all candidates belonging to *N* incur ✦ violations. By assumption (29)a at least some candidates do not incur any violations of the constraint; therefore, all such non-violating candidates belong to *D*. These are the only candidates that are passed down to the lower-ranked constraints. Therefore, after the application of *C*, all remaining candidates belong to *D* and none to *N*. No matter what the lower-ranked constraints do, no candidate belonging to *N* can win, because no such candidate survives the application of *C*. The following diagrams illustrate this argument for case (29)a. Shading indicates that the candidates are still in the running.



Full candidate set



After application of *H*



After application of *C*

Now consider the case (29)b. Because, by assumption, all candidates in the set passed down to *C* incur an equal number of violations of the implicational constraint, the function *C* is an identity function: it does not eliminate any candidates but simply passes down to the lower-ranked constraints the full set of candidates that are still in the running. Because *C* can be ignored, the winner of the overall evaluation is $L(C(H(C\textsc{and}))) = L(H(C\textsc{and}))$. This winner is known: it is the possibly non-unique DC (27). In other words, in the case (29)b the constraint *C* has no effect, so the winner is the same as when the constraint is taken out of the set; by assumption, this winner is the DC, which belongs to the set *D*. Therefore, in the case (29)b as well no member of *N* can be optimal. This completes the demonstration.

In the next section I move to the more complicated situation where more than one implicational constraint is present in the system. There, too, I will argue that no candidate that incurs a ✦ violation may be optimal.

3.3.5 More than one procedural constraint

In the toy examples so far, only one procedural constraint was involved. Clearly, this is an unrealistic simplification. I have to address the question of how two or more such constraints in a grammar can interact, and show that the typological results demonstrated for one implicational constraint are not compromised by adding more constraints to the system. This is the third step of my argument for the perpetual loserdom of candidates that incur ✦ violations.

Because the ICP makes the violation profile of each procedural constraint dependent on the rest of the grammar, it is not immediately clear how the ICP can be applied when there is more than one constraint. In fact, at first blush, the situation appears rather difficult. It is easy to see that, if there are two constraints, applying the evaluation procedure to the two constraints one after another can create unwanted results. Suppose a language has, like English, both an aspiration constraint that attracts [ʰ] to onsets of stressed syllables, and a constraint that flaps posttonic coronal consonants. Assume, as usual in our hypothetical examples, default initial stress covered by the STRESS constraints.

(30) a. ✦ASPIRATE/σ́: If a syllable is stressed, then its onset is aspirated;
b. ✦FLAP: If a syllable is stressed, it is followed by a flap.

Let us see the consequences of computing the designated state of stress separately for each constraint. First we will apply the procedure to the constraint ✦ASPIRATE, and then to ✦FLAP. Suppose the language has the ranking IDENT[son] ⟩⟩ ✦ASPIRATE ⟩⟩ DEP-*h* ⟩⟩ ✦FLAP ⟩⟩ STRESS. Consider the input /pitaɾa/. To compute the designated state of stress for the constraint ✦ASPIRATE, we take it out of the system and find the stress in the optimal candidate. Because the designated state of stress for ✦FLAP has not been found yet, we have to treat it like a normal OT constraint. Given the assumed ranking, the optimal candidate is necessarily *pitáɾa*, with stress attracted by flapping. The computation of the designated state of stress for ✦ASPIRATE is illustrated below.

(31)

		IDENT[son]	✦ASPIRATE	DEP-*h*	✦FLAP	STRESS
/pitaɾa/	pítaɾa				*!	
	☞ pitáɾa					*
	píɾaɾa	*!				
	pʰítaɾa			*!	*	
	pitʰáɾa			*!		*
	pʰíɾaɾa	*!		*		

The optimal candidate is *pitáɾa*, with a non-default stress that has been attracted to a flap. This is exactly the type of pattern that the theory is designed to rule out! The problematic stress in *pitáɾa* now becomes the designated state of stress for ✦ASPIRATE, meaning that it would ✦-penalize all candidates where stress is on any other syllable. This is illustrated by the tableau below.

(32) DS of stress for ✦ASPIRATE: second syllable (*ta*)

		IDENT[son]	✦ASPIRATE	DEP-*h*	✦FLAP	STRESS
/pitaɾa/	pítaɾa		✦!		*	
	pitáɾa		*!			*
	píɾaɾa	*!	✦			
	pʰítaɾa		✦!	*	*	
	☠ pitʰáɾa			*		*
	pʰíɾaɾa	*!	✦	*		

This result is fatal for the typological predictions: the designated state of stress, as well as the stress in the winner, is now in a place where it should not be. The candidates with initial stress – the ones that should be optimal – end up with fatal ✦-violations of the ✦ASPIRATE constraint.

Moving on to the designated candidate for the constraint ✦FLAP, taking this constraint out of the ranking does not change the winner as illustrated by the tableau below (33).

(33)

		IDENT[son]	✦ASPIRATE	DEP-*h*	✦FLAP	STRESS
/pitaɾa/	pítaɾa		✦!		*	
	pitáɾa		*!			*
	píɾaɾa	*!	✦			
	pʰítaɾa		✦!	*	*	
	☠ pitʰáɾa			*		*
	pʰíɾaɾa	*!	✦	*		

The designated location of stress for ✦FLAP is also on the second syllable *ta*. This means that any candidate with a stress on a different syllable will be ✦-penalized by ✦FLAP as well. In this particular case, as shown by the tableau below, the extra violations make no difference: the output is still the pathological candidate *pitʰáɾa*.

(34) DS of stress for ✦ASPIRATE: second syllable (*ta*)
DS of stress for ✦FLAP: second syllable (*ta*)

		IDENT[son]	✦ASPIRATE	DEP-*h*	✦FLAP	STRESS
/pitaɾa/	pítaɾa		✦!		✦	
	pitáɾa		*!			*
	píɾaɾa	*!	✦		✦	
	pʰítaɾa		✦!	*!	✦	
	☠ pitʰáɾa			*!		*
	pʰíɾaɾa	*!	✦	*	✦	

This example shows that as soon as more than one constraint is present, applying the procedure to separate constraints overturns the result that antecedent properties cannot be modified.

This very serious problem arises because both of the constraints ✦ASPIRATE and ✦FLAP mention stress in their antecedent. Thus, if we try to determine the designated state of stress of one of these constraints, the other one can 'interfere' and produce an unwanted effect; this is what gave the bad result in the tableaux above.

Crucially, such interference only occurs if the two procedural constraints share their antecedent property. In the example under discussion, it would be sufficient to compute the designated state of stress ONCE for both the constraints ✦ASPIRATE and ✦FLAP, by taking them both out of the ranking simultaneously and finding the location of stress in the optimal candidate of the resulting grammar.

More generally, the set of procedural constraints can be partitioned into non-overlapping ANTECEDENT CLASSES based on what property is mentioned in the first part of the constraint: there is a STRESS class of constraints that have the form 'If a syllable is stressed, then *Q*', a TONE class, a WEIGHT class, a NASALITY class, etc. All constraints mentioning property *P* in their antecedent belong to the antecedent class *P*. The designated state of each property *P* can then be computed only once for each antecedent class, as given by the following definition, modified from (9).

(35) THE DESIGNATED STATE (Definition; modified)
Let there be a grammar *G*, an **antecedent class of constraints** *C*, an input /i/and its output [o] in *G*, and a phonological object *p* in [o]. Then the DESIGNATED STATE of *p* for /i/ and *C* in *G* is *p*'s correspondent in the optimal candidate in the grammar *G*' that is identical to *G* except that **all of the constraints** in *C* have been removed.

The intuition behind this modified definition of designated state is that we first determine the default behavior of some property (e.g. stress), and then apply all processes which are sensitive to that property.

Now the unwanted predictions do not arise. ✦ASPIRATE and ✦FLAP both have 'stressed syllable' as their antecedent property, and thus belong to the same antecedent class. The designated state of stress is thus computed by taking both constraints out of the ranking at the same time, as shown in the tableau below. The STRESS constraints ensure that the designated state of stress is on the initial syllable.

(36)

			IDENT[son]	✦ASPIRATE	DEP-*h*	✦FLAP	STRESS
/pitaɾa/	☞	pítaɾa					
		pitáɾa					*!
		píɾaɾa	*!				
		p^{h}ítaɾa			*!		
		pit^{h}áɾa			*!		*
		p^{h}íɾaɾa	*!		*		

Now, any candidate with stress on a syllable other than the initial one would receive a fatal ✦ violation, and thus the typological predictions are preserved. The final evaluation is illustrated below.

(37) DS of stress: initial syllable

			IDENT[son]	✦ASPIRATE	DEP-*h*	✦FLAP	STRESS
/pitaɾa/		pítaɾa		*!		*	
		pitáɾa		✦!		✦	*
		píɾaɾa	*!	*			
	☞	p^{h}ítaɾa			*	*	
		pit^{h}áɾa		✦!	*	✦	*
		p^{h}íɾaɾa	*!		*		

The remaining question is how to deal with more than one antecedent class in the same system. Suppose there are *n* antecedent classes of constraints; call them C_1, C_2, …, C_n. For each input, each of these antecedent classes is associated with a designated state of its antecedent property. The question is, how to determine that property? For each class, the designated state must be determined by taking all of the constraints in that class out of the system, and letting the remaining grammar decide the behavior of the relevant property. However, once this operation is performed for some class C_i, the violation pattern of all of the constraints in C_i changes. Therefore, the order in which the designated states for the different antecedent classes is determined can potentially affect the outcome. Furthermore, and more importantly, it is not immediately clear that the typological predictions of the theory remain with the addition of more than one antecedent class to the grammar.

I argue here that the typological predictions stand: no matter what order the designated state is computed for the several antecedent classes of constraints, no candidate where the antecedent property is modified can win. However, an unsettling result is that the order in which the operation is performed does in fact affect the outcome. Thus, while the system is typologically well-behaved, it is indeterminate: the same grammar can produce several outputs. It will then be necessary to fix an order for computing the designated states arbitrarily, by stipulation. This area of the theory remains for future research: it remains to be seen what empirical predictions are made by this arbitrary choice of the order of computation of designated states.

Let me now go through the first part of the argument, viz. that the typological predictions are not compromised by the presence of more than one antecedent class of constraints. Consider some antecedent class, C_i, whose antecedent property is P_i. The constraints express certain procedural generalizations about P_i, by putting this property in their antecedent and thus penalizing certain input-output mappings of it, by the ICP. By assumption, ALL procedural generalizations concerning P_i are expressed by the constraints in C_i; that is how the constraint system is designed. To take a specific example, if P_i is stress, then all generalizations about what stress cannot be attracted to are expressed by the constraints in its antecedent class, C_i. It contains constraints like '✦If stress, then aspiration', '✦If stress, then flapping', etc. – constraints that make it impossible to assign stress based on certain segmental properties. No constraint outside of C_i has the power to limit the set of properties to which stress can be sensitive, either because those constraints are not procedural, or because they belong to a different antecedent class.

Once again, by assumption, no grammar with all of the constraints in C_i taken out can produce a typologically undesirable stress pattern (e.g. aspiration-driven stress), because all constraints that can cause such effects are members of C_i. It immediately follows that the designated state of the property P_i is typologically well-behaved, and thus the ICP cannot cause an unwanted candidate to win. This completes the argument that the typological predictions of the theory are not compromised by the presence of more than one antecedent class of constraints.

However, there is an indeterminacy to the theory. If there is more than one antecedent class of constraints, each of which is associated with a designated state of various properties for each input, in what order should the designated state be computed? A simple example shows that the choice of the order can affect the outcome. Suppose there are two procedural constraints in the grammar, shown below. One is the familiar aspiration constraint, the other is the sonority-driven stress constraint that attracts stress to syllables with the nucleus *a*.[18]

(38) a. ✦ASPIRATE: If a syllable is stressed, then the onset is aspirated.
b. ✦STRESS-*a*: If the nucleus of a syllable is *a*, then the syllable is stressed.

Consider the input /pitaku/ in a grammar with default initial stress. Then, depending on the order in which the designated state-computing procedure is applied to this input, the winner can be either *pit^{h}áku* or *p^{h}ítaku* (I will spare the reader the tableaux that show this). Neither of these presents a typologically undesirable pattern (stress attraction to aspiration), but the grammar is indeterminate.

It is therefore necessary to fix the order in which the designated states are computed for several antecedent classes. I am not aware of any independent motivation for such an order. In the remainder of this dissertation, I will apply the procedure in the order in which the constraints are ranked, starting from the highest-ranked and proceeding downward. Once again, making this choice into an empirical one remains for future research.

---

[18] The exact form of this second constraint does not matter for now; I will come back to sonority-driven stress below. What is crucial here is that 'being *a*' is in the antecedent, while the location of stress is in the consequent.

3.4 Bidirectional interactions

The discussion so far was devoted to those cases where stress interacts asymmetrically with some segmental property. I have shown how my proposal to introduce a new type of constraint into the theory allows generalizations about the stress assignment process to be handled in OT. I suggested that by linking the asymmetry of the implicational statement of a constraint to the direction of interaction between the two properties mentioned in it, undesirable processes like aspiration-driven stress can be ruled out.

As explained in detail in Chapter 1, the typology of interaction between stress and segments contains not only asymmetrical interactions such as those analyzed in this chapter so far, but also the three broad classes of cases where stress interacts bidirectionally with some other property. The three properties that are able to influence stress are quantity, tone, and vowel sonority. In this section I explore the role of implicational constraints in accounting for these symmetrical interactions.

There are two general approaches one could take to these cases. One would be to argue that whenever stress interacts freely with some property, the constraints responsible for it are not implicational, but standard OT constraints expressing output-based generalizations that do not invoke the special interpretation mechanism proposed here. This approach is, of course, assumed in standard OT, where constraints like stress-to-weight (SWP) and weight-to-stress, stress-to-tone and tone-to-stress (de Lacy 2003) are interpreted as standard output-oriented markedness constraints.

The second approach to the bidirectional interaction of stress would be to employ, for each of the features, two implicational constraints with opposite directionality. If the procedural constraints 'If stressed, then heavy' (✦SWP), and 'If heavy, then stressed' (✦WSP) exist in the system, then both directions of interaction would be accounted for: each of the two constraints would be responsible for one of the directions of interaction. The former constraint would produce stress-driven weight, and the latter constraint would account for weight-driven stress. Here I argue that the second option, with two opposing implicational constraints is the correct view.

The empirical difference between the two approaches lies in the types of processes BESIDES stress-segmental interactions that the constraints are predicted to force. If the theory has only two constraints, ✦WSP and ✦SWP, then the only two kinds of repairs for stress-weight misalignment that are predicted to exist are stress attraction to heavy syllables, and lengthening/gemination in stressed syllables to make them heavy. The two constraints cannot cause any other process, under the theory proposed here. Because the effect of the ICP is to prevent constraints from modifying the property mentioned in its antecedent, ✦SWP cannot cause any repair that involves a stress shift, while ✦WSP cannot cause any repair that modifies the original heaviness of a syllable.

On the other hand, if stress-weight interactions are handled by the standard OT pair of constraint like SWP and WSP, no such predictions about possible repairs are made. My argument here is that the more restrictive predictions of the implicational version of ✦SWP and ✦WSP are empirically correct.

An important set of cases where SWP and ✦SWP make different predictions concerns vowel syncope. I will touch briefly on these cases here; they will be the topic of a more detailed investigation in Chapter 4. The standard OT constraint SWP can cause the deletion of a stressed vowel just in case the stressed syllable in the outcome is heavy, while the implicational version ✦SWP cannot do so. Let me illustrate this point with a hypothetical example. Suppose the input has three light syllables with stress on the second syllable, /$CV_1C\acute{V}_2CV_3$/. A high-ranking SWP constraint can potentially force the deletion of a the stressed vowel $\acute{V}_2$ just in case all other potential repairs creating a heavy stressed syllable in the output are ruled out by other constraints. The tableau below illustrates such a case: the fully faithful candidate $(CV_1C\acute{V}_2)CV_3$ is violates SWP; lengthening of the stressed vowel $(CV_1C\acute{V}_2{:})CV_3$ violates DEP-μ; and the deletion of the final syllable to give $(CV_1C\acute{V}_2C)$ is precluded by the high-ranking NONFIN constraint. This leaves the deletion of the stressed vowel $\acute{V}_2$ with resyllabification of its onset as the coda of the preceding syllable as the only viable option that would satisfy SWP. As a result, the constraint SWP causes the deletion of a stressed vowel and a concomitant stress shift.

(39) /CV₁CV́₂CV₃/

		SWP	DEP-μ	NONFIN	PARSE-σ	MAX-V
	$(CV_1C\acute{V}_2)CV_3$	*!			*	
	$(CV_1C\acute{V}_2{:})CV_3$		*!		*	
	$(CV_1C\acute{V}_2C)$			*!		*
☞	$(C\acute{V}_1C)CV_3$				*	*

On the other hand, the implicational version ✦SWP cannot produce such an outcome, because this constraint has stress mentioned in its antecedent, and therefore any candidate with a stress shift is a perpetual loser. This is illustrated in the tableau below.

(40) /CV₁CV́₂CV₃/
Designated location of stress: V_2

		✦SWP	DEP-μ	NONFIN	PARSE-σ	MAX-V
	$(CV_1C\acute{V}_2)CV_3$	*!			*	
	$(CV_1C\acute{V}_2{:})CV_3$		*!		*	
☞	$(CV_1C\acute{V}_2C)$			*!		*
	$(C\acute{V}_1C)CV_3$	✦!			*	*

Although several analyses of syncope processes that make use of the power of SWP to produce stressed vowel deletion have been proposed, I will argue in Chapter 4 that such analyses are incorrect, and better accounts are available for the relevant cases. As I will show, this inability of SWP to force stressed vowel deletion is part of a more general typological generalization that stressed vowels cannot be deleted through metrically-driven syncope. If that claim is correct, then it follows that the more restrictive constraint ✦SWP is to be preferred over the less restrictive version SWP, because the typology of repairs predicted by ✦SWP more tightly fits the observed typology.

3.5 Grounding the constraints

The mechanism of constraint interpretation introduced in this chapter allows constraints to refer directly to processes, and to penalize those candidates that involve undesirable processes. The proposal is intended to replace SOME of the current OT markedness constraints with constraints of the new type, but, crucially, output-oriented constraints are still present in the theory.

The rationale for allowing these two types of constraints in the system is that phonological generalizations are of two types. Some are surface-based, and require the standard OT constraints to handle them. Other generalizations are procedural, and require the new ✦-constraints to directly express them. This richness of the formal machinery in the proposed theory relative to standard OT is necessary because of the richer set of phonological generalizations than has been standardly assumed.

The next question then is, is it possible to predict generalizations are output-based and which generalizations are procedural? From the theoretical standpoint, is it possible to tell which constraints should be subject to the standard OT mechanism of constraint evaluation, and which constraints should be subject to the ICP? In other words, we have the problem of GROUNDING the constraints in some extratheoretical reality.

In this section I discuss some psycholinguistic evidence in favor of the observed proposed directionality of the stress constraints. However, I will argue that the best way of determining the nature of a given constraint is the typology of the interactions between the properties that it mentions.

3.5.1 Typological grounding

Rather than grounding the directionality of the asymmetrical procedural constraints in some extralinguistic factor, I rely on the more standard technique used by OT practitioners: grounding constraints in the observed typologies, and selecting, from the set of possibilities provided by the formalism, those constraints that account best for the crosslinguistic distribution of a particular feature. To use a familiar example, there is nothing a priori which leads to the two syllable structure constraints being ONSET and NOCODA: as far as the formal machinery is concerned, these constraints might have been the opposite, CODA and NOONSET. However, the typology of syllable structure points to the correct formulation of the constraints (Prince and Smolensky 1993).

The typological predictions on the interaction of phonological properties depend on the correct allocation of phonological categories among the two parts of the constraint, the causer (the antecedent) and the causee (the consequent). The form of each constraint, as stated above, must be decided on a case by case basis: whatever the direction of interaction the typology indicates must be reflected in the statement of the constraint. This lack of external, independent grounding is perhaps the weakest aspect of the current proposal.

As mentioned elsewhere in this chapter, there are yet more degrees of freedom than the choice of directionality for the procedural constraints. Whether or not a given constraint is procedural in the first place, and thus controlled by the violation assignment mechanism introduced here, is also subject to the analyst's choice. Once again, this choice must be made with reference to the observed typology.

The main objection of this dissertation to canonical OT is that its hypothesis that all phonological generalizations are in the output is too radical. I do not intend to replace this radical hypothesis with its opposite, however. Nothing I have written here should be interpreted as offering a counterargument to the observation, established since Kisseberth 1970, that phonologies often conspire to produce a certain output pattern. This fundamental claim stands. And yet, the systematic gaps in the typologies of interaction of some properties show that SOMETIMES phonologies respond to pressures other than purely surface-driven. These two sources of phonological generalizations, surface pressures and input-output mappings, lead, in my theory, to two types of constraints. Procedural constraints reflect the latter kind of generalization, and standard

OT constraints reflect the former. It is the analyst's job to assign each markedness constraint to one class or the other.

The argument in this chapter is that there are two kinds of constraints – the standard OT constraints that refer to output structures and prohibit marked configurations, and those that refer to processes and prohibit certain input-output mappings. The former type of constraints encode surface-based generalizations, while the latter type refer to procedural generalizations. For a given constraint, the observed typology of interactions serves as a clue to assigning the constraint to one type of the other. If the relevant phonological categories interact in a way that is most insightfully described in terms of a procedural generalization, then the constraint must be stated implicationally and is subject to the ICP. If, on the other hand, the generalization is best stated in terms of outputs, then it must be handled by a standard OT constraint. I will use this reasoning in Chapter 4 to argue that syllable structure constraints like ONSET and NOCODA, as well as sonority-sequencing and syllable contact constraints, must remain of the standard output-oriented type (section 4.1.4).

3.5.2 Psycholinguistic evidence

In this section I survey the psycholinguistic evidence that is relevant to the interaction of prosody and segments. If the directionality of the statement of constraints is to be grounded in such extragrammatical facts, the mode of explanation here would be analogous to Hayes' (1995) proposal to link the asymmetrical foot typology (the Iambic-Trochaic Law) to the effects of rhythm psychology: "I posit an EXTRASYSTEMATIC motivation, in a law of rhythm, for internal formal principles of the linguistic system" (Hayes 1995: 81). Cf. also the programmatic statement in Anderson 1981: "[W]hile linguistic constraints proper need not mirror the restrictions of other cognitive structures, nonetheless these other structures, insofar as they are involved in the development of linguistic knowledge, can be expected to have their own consequences for actual grammar… [T]he character of the linguistic system depends on the interaction of particular substantive considerations NOT specific to Language with an irreducible component which is" (1981: 536).

While the evidence I discuss below suggests that the asymmetrical statements of constraints are not arbitrary and have some basis in the human perception of stress and rhythm, I will remain cautious in directly grounding the claims about the formal structure of constraints in evidence about processing. There are two reasons to remain cautious in this respect: first, the evidence on processing and perception of stress is not complete, most importantly lacking in broad cross-linguistic data that would allow one to see similarities and differences between typologically distinct stress systems. Secondly, and more importantly, the theory I am proposing here is a theory of knowledge of language, which is in principle independent of processing and performance, and thus data about how humans process stress and segments cannot be directly used to argue for one or another formalism. With these caveats in mind, I turn to discussing the relevant experimental studies.

On the whole, the evidence indicates a psycholinguistic difference between those features with which stress is typologically observed to interact bidirectionally and those which stress can only influence in a unilateral fashion. Two sets of arguments from the psycholinguistic literature suggest that the prosodic component has a special status in phonology vis-à-vis the segmental component. First, patterns of speech errors and tip-of-the-tongue phenomena differ in the prosodic and segmental domains in a way that points to the logically prior status of stress with respect to segments. Second, the special status of quantity, intensity, and tone as correlates of stress is confirmed by the literature on the psychology of rhythm, which shows these same features to play a role in perception and production of non-linguistic prosody.

In this section I will review the available evidence that bears on the role of prosody in grammar. However, while the arguments from this external domain are suggestive, and at least do not contradict a theory of prosody-segmental interactions that gives special status to prosodic structure with respect to segmental features, I will not argue for a direct grounding of the asymmetrical prosody-segmental constraints in psycholinguistic data. There are two reasons to be cautious in this respect. First, the claims of this section must be taken with a grain of salt because most of the studies presently available have been done on English, so we know little about the crosslinguistic differences in processing and acquisition of prosodic structure, such as the difference between pitch accent and dynamic stress languages, let alone the psycholinguistic status of covert prosodic structure. The second reason why such external evidence does entail a particular view of constraint interaction in a direct way is that the theory I am pursuing is a theory of the knowledge of language, not of processing.

Let me begin by discussing the data on speech errors suggesting that prosody is processed prior to the segmental details. One source of information are tip-of-the-tongue (TOT) phenomena, which involve a temporary inability to produce a phonological output for a word whose semantics has already been accessed (Levelt 1989: 320). It TOT states, speakers have the intuitive feeling that they know the word they are trying to produce, but cannot remember how it sounds exactly. Because priming with phonologically similar but semantically unrelated words can cause a TOT state, it is generally accepted that such states are caused by an unsuccessful search for a phonological form of a given lemma, not for the lemma itself. Speakers in TOT states are typically aware of the prosodic structure of the form whose segments they cannot access, viz. the number of syllables and the stress pattern (60-80%). The initial segment of the target word is also available to TOT state speakers in 60-70% of the time. What is missing, however, are the details of the word's segmentation: the distribution of consonantal and vocalic features in the form. This clearly suggests that prosodic information is accessed earlier than segmental structure in word generation. TOT states occur when a speaker fails to proceed from the prosodic to the segmental stage.

Speech error data also suggest a separate behavior of prosody as opposed to segments. Particularly relevant is the study of malapropisms – word selection errors that are not due to substitutions, anticipations, and similar phonological problems, and are not semantic errors. In a survey of English malapropisms, Fay and Cutler (1977) showed that the mistakenly selected word in 98% of the cases has the identical stress pattern to the target word. This rate of agreement is higher than the 82% agreement in stress patterns found in errors classified as semantic (Fay and Cutler 1977: 508). The malapropisms typically have the same syllable count and the same grammatical category as the target word. This suggests that, at the point at which the error occurs, morphological and prosodic information, but not segmental structure, has been accessed. The locus of error in malapropisms is the same as in TOT states: it occurs between the prosodic and the segmental stages in processing. The difference between the two types of errors is that in the latter case the speaker fails to access any form, while in the former case he proceeds in the wrong direction. This confirms what TOT phenomena also suggest, that prosody is accessed relatively early in speech production.

Equally informative are error types and TOT states that are not reported to occur. While it is common for a speaker in a TOT state to have accessed the prosodic structure of a word while still searching for segmental information, the reverse type of TOT states has not been reported, i.e. having accessed and being able to produce the segmental content but searching for its correct prosodic parse, or its right stress pattern. Likewise, while malapropisms involve segmental mistakes with the correct prosodic structure, the reverse type of error appears to be unreported. Such a prosodic malapropism would contain the correct segmental structure with a prosodic aberration that cannot be attributed to substitution or anticipation errors.

To be sure, stress errors do occur in speech (Cutler 1980). However, they are almost entirely errors of a single type: an erroneous stress pattern is 'borrowed' from a morphologically related word. For example, Cutler reports the error *certífication*, due to the stress pattern of the related word *certíficate*. This type of error is clearly different from segmental malapropisms, where the incorrect segmentism is borrowed not from morphologically related words, but from PROSODICALLY similar unrelated words. Cutler's survey shows that in stress errors the prosodic pattern is not borrowed from segmentally similar but morphologically unrelated words. Once again, the error data indicate that prosody is accessed before segments.

These facts have lead researchers like Levelt to posit that prosodic information is accessed early, perhaps concurrently with morphological structure, and prior to most segmental information.

While the error studies have concentrated on English, there is some crosslinguistic data available on the comprehension side. The summary of previous research in Cutler and Van Donselaar (2002) indicates that stress is an important factor in word recognition in Dutch but not in English. Members of stress-based minimal pairs in Dutch, like *vornáam* 'first name' and *vórnaam* 'respectable' do not prime each other, while the members of analogous English pairs like *forbéar* and *fórbear* do. The relatively greater importance of stress in word recognition in Dutch vs. English was linked by these researchers to the relatively greater amount of vowel reduction in English. Because the functional load of stress per se is not great in a language with significant vowel reduction – in English, there are not many pairs like *forbéar* and *fórbear*, where the stress difference does not correlate with some vowel quality difference – stress is simply not a useful tool in word recognition in English. Instead, Cutler and Van Donselaar argue, English speakers, unlike Dutch speakers, act more efficiently by focusing on segmental cues in word recognition.

Note, however, that this result does not necessarily show that PROSODIC STRUCTURE itself is not relevant to word recognition. Another way to interpret the findings is simply that intensity is not a clue to prosodic structure used by English speakers; instead, they rely on the more robust segmental cues such as vowel reduction as a source of information about prosody, but use prosody to distinguish words. It would follow that pairs where the clues are abundant – those with vowel reduction – would be more easily distinguishable, while the prosody of pairs like *fórbear* and *forbéar* would be less accessible.

Another relevant line of research involves the relationship between the type of stress system found in a language, the presence of lexical exceptions in stress placement, and the ability of speakers to 'hear' stress (Peperkamp and Dupoux 2002, Peperkamp 2004). These researchers observe that speakers of some languages, such as French, are "stressdeaf" – are not able to identify explicitly where stress goes in a form, and are poor at distinguishing stress contrasts when presented with stimuli from a language that has contrastive stress. On the other hand, speakers of Spanish are relatively more aware of the location of stress and are better at distinguishing members of accentual minimal pairs. Peperkamp and Dupoux link this difference in behavior between the two languages to how predictable the location of stress is from utterance boundaries. A child learning a language like French can infer the location of stress solely with reference to utterance boundaries – no information on word or morpheme boundaries is necessary to arrive at the correct stress rule. Stress then can be acquired 'prelexically' – that is, before actual words and independently of them – and will not be stored in the mental lexicon for each lexical entry. Conversely, in a language like Spanish, where morphological and lexical information figures in the stress rule, that rule cannot be acquired solely from the position of stress relative to word boundaries, and thus cannot be available to the child prelexically. As a consequence of this relatively late acquisition of the stress rule, Dupoux and Peperkamp hypothesize, in Spanish-like languages accentual information is redundantly recorded in lexical entries. This accounts for the relatively greater awareness of stress by Spanish speakers than by French speakers. Peperkamp (2004) further links this difference to the existence of lexical stress exceptions in Spanish-type languages vs. French-type languages.

This work suggests that there may be a crosslinguistic difference in processing of stress. Because English, with its numerous stress exceptions, morphologically based accentual generalizations, and so forth, clearly belongs to the Spanish type rather than the French type, work done exclusively on English speech error data might present a skewed picture of the processing of prosodic structure. Fay and Cutler's (1977: 511) suggestion that stress patterns in English serve as an important organizational principle of the lexicon that aids both production and comprehension may not hold for languages of a prosodic type differing from English. For this reason the suggestive psycholinguistic data discussed in this section must be taken with a grain of salt.

So far I have discussed the psycholinguistic differences between prosodic and segmental structure in a given language with an assumption that it is established a priori which features count as segmental and which features count as prosodic. The next question, then, is: given that there are these two categories of phonological properties, why do intensity, quantity, and tone fall on the prosodic side of the divide, while the segmental features fall on the other side? To address this question I turn to the literature on the psychology of rhythm.

The function of rhythmic stress is to organize the speech string into manageable units for processing and storage, and that is why the nature of stress systems can be made clear by looking at the psychology of rhythm (cf. Hayes 1995). I suggest here that the special status of intensity, duration, and tone in stress systems, and ultimately the availability of bidirectional constraint sets linking these properties with abstract prominence, is due to the role of these features in the psychology of rhythm.

Duration, intensity, and pitch stand out among the linguistically relevant characteristics of a speech signal in that these properties are all and only that can be meaningfully applied to non-linguistic signals as well.[19] One can talk about, say, a musical note having duration and pitch and loudness, but not nasality or spread glottis or aspiration. What makes prosodic properties special in language is that these properties are coopted for linguistically relevant use from a non-linguistic domain. In language, prosodic features are also the only ones that express what can be called "paralinguistic meaning": the expression of emotional states such as surprise, disgust, excitement, and so forth, is done via exaggerating the prosodic features of the speech signal, or overlaying it with new prosodic content. One would not expect to find a language where, say, extreme surprise is expressed by nasality or voicing rather than by intensity or shifts in pitch register. And finally, the prosodic features are the only ones that can be meaningfully applied to non-linguistic vocalizations that are part of the speech signal, i.e. expressions like "arghh, hmm, phew", whistles, signs, and so forth.

[19] This insightful observation is due to Larry Hyman, p.c.

All such items must necessarily have some duration, some loudness, and, at least if voice is involved, some pitch – but not necessarily any segmental features.

In sum, prosodic properties are all and only properties that exist independently of language, but can be used by phonological systems for linguistically relevant ends.

This extralinguistic status of prosody is manifested also in its role in the psychology of rhythm. There is extensive evidence showing that humans have an innate tendency to organize sequences of like stimuli into small, regularly recurring, periodic constituents. Prosodic properties have a special role with regards to the perception and production of rhythm.

The basic human propensity for rhythmic organization is demonstrated by experiments showing that a succession of identical sounds is perceived as a rhythmic pattern, at a rate of between 0.5 and 5 Hz (Bell 1977, Fraisse and Oléron 1954, Allen 1975). The length of the rhythmic units is determined by the rate of the stimulus presentation: the higher the rate, the more units per group.

The two prosodic properties of intensity and duration have been shown to correlate with rhythmic strength in a way that has been subject to much discussion in the metrical stress literature. The iambic-trochaic law (Hayes 1995) expresses this asymmetry in perception: loud sound first, long sound last. A succession of sounds with alternating intensity is perceived as a trochaic sequence (Fraisse and Oléron 1954), though the effect diminishes at slow rates, below about 3.5 Hz (Bell 1977). If the successive sounds are distinguished by length rather than loudness, they are perceived as an iambic sequence, provided the difference in length is by a factor greater than about 1.5. Pitch seems to have a weak tendency to cause left prominence (Bell 1977).

A further relevant result of Bell 1977 is the relative language independence of the rhythmic effect. He found that Polish speakers are not influenced by the regular penultimate default stress of their language and do not perceive the penultimate sound in a series of identical sounds as more prominent.

This experimental work shows that the prosodic characteristics of at least intensity and duration have what we may call 'inherent salience': they are associated with rhythmically prominent positions. While we may remain agnostic as to the innateness of the human psychology of rhythm, it is clear from the experimental work cited above that, first, this ability and tendency to group stimuli into periodic units is extralinguistic, and second, that duration and intensity are inherently associated with rhythmic prominence. Tone may also have such a relationship with rhythm, although the details are less clear.

Once again, prosodic features in language are those that also have extralinguistic relevance – in this case, relevance to the psychology of rhythm – and have been coopted by languages to signal linguistically relevant rhythm, i.e. stress. This relationship between the extralinguistic and phonological manifestations of prosody becomes especially clear if we consider a typological asymmetry so obvious that hardly anyone ever bothers to note it. As I have mentioned many times above, stressed syllables tend to be both longer and louder than unstressed syllables. However, nothing in principle prevents the reverse use of prosodic cues: why are there no languages where stressed syllables are not louder, but LESS LOUD than unstressed syllables? Or languages where stressed syllables are not longer but SHORTER than unstressed ones?[20] Naturally, such a hypothetical language would appear absurd to most phonologists, but it is not as unreasonable as it seems. Stressed syllables in a language where they are cued by shortness rather than length would be just as easily distinguishable from unstressed stressed syllables as in a 'normal' language, so the absence of such languages from the typology has nothing to do with perceptibility. Instead, such counter-natural way of cuing prosodic prominence would go against the inherent association between rhythmic strength and prosodic features, and would arguably be unlearnable by a human being.

To sum up the discussion so far, the evidence from the psychology of rhythm suggests that prosodic features have a special status in language in two ways: first, they are the features that have non-linguistic meaning, and the only properties that can be

[20] There is no circularity here. Some readers may ask if in such a language we would simply call the shorter and the less intense syllables unstressed. The situation I claim here is unattested would be distinguished by having only ONE shorter or less intense syllable per word, on analogy to systems that have only one louder or longer syllable per word.

predicated of non-linguistic stimuli; and second, these prosodic features (with some complications for tone) have an inherent association with the human sense of rhythm. Evidently, these two facts are related: it is no surprise because rhythm psychology is an extralinguistic ability, and humans have a sense of rhythm in stimuli devoid of any linguistic content, that the features used to signal that rhythm are also not specifically linguistic.

Crucially, none of what I said so far with regards to the three prosodic features applies to any of the other potential effects of stress – i.e. any of the segmental effects having to do with consonantal features such as voicing, sonorancy, etc. These features are language-specific, not relevant to extralinguistic signals such as music, and have no inherent association of any kind with rhythm. This fact may suggest that the cause-effect relationship between prosodic features and linguistic prosodic constituent structure is bidirectional, but the effect of stress on consonants is not reciprocated by the like effect of consonants on prosody. As I stressed above, however, despite the suggestive nature of the psycholinguistic evidence, formal phonology must remain cautious about directly using that evidence in support of a particular formalism.

3.6 Sonority-driven stress and prominence

3.6.0 Introduction

Stress systems are often sensitive to vowel quality, more specifically, to the position of a vowel on the sonority scale (Kenstowicz 1994, de Lacy 2003). The standardly accepted sonority scale below corresponds to vowel height, with the low vowels being more sonorous that non-low vowels, and central vowels being the least sonorous of all. Sonority-driven stress systems typically have a preference for assigning stress to syllables more sonorous vowels.

(41) ə, ɨ ⟨⟨ i, u ⟨⟨ e, o ⟨⟨ a

De Lacy's (2003) theory of sonority-driven stress analyzes the interactions between stress placement and vowel quality with sets of markedness constraints in a stringent relationship, as shown below.

(42) a. *$\mathrm{Hd}_{\mathrm{Ft}}$/{ə}
*$\mathrm{Hd}_{\mathrm{Ft}}$/{ə, i/u}
*$\mathrm{Hd}_{\mathrm{Ft}}$/{ə, i/u, e/o}
*$\mathrm{Hd}_{\mathrm{Ft}}$/{ə, i/u, e/o, a}

b. *Non-$\mathrm{Hd}_{\mathrm{Ft}}${a}
*Non-$\mathrm{Hd}_{\mathrm{Ft}}${a, e/o}
*Non-$\mathrm{Hd}_{\mathrm{Ft}}${a, e/o, i/u}
*Non-$\mathrm{Hd}_{\mathrm{Ft}}${a, e/o, i/u, ə}

The constraints in (42)a penalize vowels in prominent positions starting at the low-sonority end of the hierarchy, while the constraints in (42)b penalize vowels in non-prominent positions starting at the high end of the hierarchy. In case these constraints are allowed to outrank stress placement constraints, the predicted effect is to force stress onto more sonorous vowels and repel it from less sonorous vowels. Conversely, if the constraints in (42) outrank vowel quality faithfulness constraints, the predicted effect is for vowels in prominent positions to move up on the sonority scale and for vowels in non-prominent positions to move down the scale.

As I mentioned in Chapter 2, the existence of sonority-driven stress is not easily compatible with de Lacy's the fixed ranking approach, because the constraints in (42) are the only constraints that mention segmental information that are allowed to outrank stress placement constraints. Sonority-driven stress can be analyzed using the proposal on procedural constraints in this chapter as long as the direction of the implicational statement of the constraint mentions stress in the consequent, not the antecedent. If the constraint says 'If a syllable is stressed, then its nucleus is an [a]', the stress cannot shift to satisfy it. If, on the other hand, the constraint is 'If a syllable's nucleus is [a], then that syllable is stressed', then stress can be attracted to low vowels.

While the theory proposed above can deal with the general case, there are some types of sonority-driven stress, uncovered by Vaysman, that are not compatible with any of the current theories. In section 3.5 below I will propose a further refinement of the analysis of stress-quality interactions that allows such cases to be accounted for, arguing for formally separating prominence from metrical structure.

First, however, I will discuss an alternative view of sonority-driven stress that relates it to quantity-driven stress.

3.6.1 Quantity or sonority?

Stress attraction by high-sonority vowels appears related to stress attraction by heavy syllables and long vowels. High sonority vowels are longer and, in some sense, more prominent than low-sonority vowels. The question is then, are sonority-driven stress and quantity-driven stress two manifestations of the same phenomenon, and can they be collapsed theoretically? In other words, can the sonority hierarchy be treated as a manifestation the weight hierarchy? Here I review arguments for and against such a move. It appears that the evidence at present is inconclusive.

Kenstowicz (1994) and de Lacy (2003) treat sonority-driven stress as a theoretically separate phenomenon from quantity-driven stress. The weight hierarchy CVV 〉〉 CVC 〉〉 CV is separate from the sonority hierarchy a 〉〉 e,o 〉〉 i,u 〉〉 ə, and the two can in principle act independently in the same language. An approach that treats weight- and sonority-driven stress as the same phenomenon would collapse the two hierarchies into one general quantity hierarchy. One way to do this is to subdivide each of the elements of the weight hierarchy based on sonority of the nucleus vowel, as shown below.

(43) Cā 〉〉 Cē,ō 〉〉 Cī,ū 〉〉
CaC 〉〉 CeC, CoC 〉〉 CiC, CuC 〉〉 CəC 〉〉
Ca 〉〉 Ce, Co 〉〉 Ci, Cu 〉〉 Cə

There are two arguments in favor of collapsing the weight and sonority hierarchies, and three arguments against doing so. I will go through these five arguments in this section, but withhold judgment on what the correct analysis is.

First, the majority of sonority-driven stress languages have either an 'unstressable *ə*' system, where stress is repelled from schwas, or a 'stressed *a*' system, where stress is attracted to low vowels, and no vowel length distinctions. All such systems can be analyzed as quantity-driven, with quantity predictable from vowel quality. At least in some cases such an analysis seems preferable. In Kara (Schlie & Schlie 1993, de Lacy 2003), for example, stress is assigned according to the following algorithm.

(44) a. Rightmost syllable containing [a]
b. Else rightmost closed syllable
c. Else stem-initial

The vowel system of Kara is shown below. Apart from [a] and [ə], all vowels exist in tense-lax pairs, whose distribution is at least in part predictable.

(45)	i ɪ				ʊ u	
	e ɛ		ə		ɔ o	
			a			

Omitting the application of laxness harmony and some other low-level rules, the distribution of tense-lax vowels is shown below in (46). The distribution of the mid vowel pairs is entirely predictable, while for high vowels, there appears to be a contrast in closed syllables. Schlie and Schlie, however, analyze tense high vowels in closed syllables as coming from underlying [iə]; if this analysis is correct, then the distribution of tenseness in high vowels is also entirely predictable from syllable structure. In contrast, [a] and [ə] are distinct in all syllables.

(46) a. [e, o]: in open syllables
[ɛ, ɔ]: in closed syllables
b. [i, u]: open syllables
[i ~ ɪ, ʊ ~ u]: contrast in closed syllables (S&S analyze [i] in closed σs as /iə/)
c. [a ~ ə]: contrast in all syllables

These facts suggest that the [a~ə] distinction is different from the tense-lax distinction in the other vowels. In addition to the distributional evidence given above, there are two other reasons not to treat the [a~ə] pair as tense-lax. First, when there is a partial tense-lax contrast in a language, it is usually the HIGHER, not the lower vowels that are contrastive (as in English): Kara would then be typologically unusual if [a~ə] are a tense-lax pair. Secondly, only the high and mid vowel pairs but not the [a~ə] pair participate in laxness harmony, as summarized below.

(47) **Kara laxness harmony**
In open σs: [ɛ] before [ɪ,ɛ]
[ɔ] before [ɔ] (no data on [ʊ])
[a, ə] do not undergo this process

I take these facts to suggest that [a~ə] is not a tense-lax pair, but should be analyzed in terms of quantity: [a] is a long (bimoraic) vowel, and [ə] is its short counterpart. If this is the case, the stress rule then takes the following form.

(48) a. Stress the rightmost long vowel
b. Else the rightmost closed syllable
c. Else initial

This new formulation of the stress rule makes no reference to vowel sonority, but only to the weight hierarchy CVV ⟩⟩ CVC ⟩⟩ CV. There happens to be an indirect piece of evidence that this analysis is correct. The vowel system of Nalik, a closely related language (Volker 1998), shows a vowel length distinction only in the low vowels: there is a [a~ā] contrast, but no other quantity contrast. Unfortunately, Volker provides no description of stress in Nalik.

(49) Nalik vowel system (Volker 1998)

i		u
e		o
	a, ā	

In sum, at least some systems that superficially involve sonority-driven stress can, and perhaps should, be reanalyzed in terms of quantity. The proposed reanalysis of Kara is applicable to any system that has either a 'stressed *a*' or 'unstressable *ə*' generalization, and no other quantity distinctions.

The second argument in favor of collapsing the weight and vowel sonority hierarchies comes from the typology of interaction between weight and sonority. Arto Anttila (p.c.) observes that whenever a stress system has both quantity and sonority sensitivity, SONORITY NEVER TRUMPS QUANTITY. In other words, preferential assignment of stress to syllables with higher-sonority nuclei is only observed in case there is no other quantity distinction. While a syllable like [Ca] may attract stress over a syllable like [Ci], one never observes syllables like [Ca] attracting stress over BOTH [Ci] and [Cī]. This fact suggests that a hierarchy like the one in (43) might reflect the typological facts in a more restrictive way than two separate hierarchies for weight and vowel sonority.

To sum up, there are two arguments in favor of treating sonority-sensitivity as a kind of weight-sensitivity: the preponderance of 'unstressable *ə*' and 'stressed *a*' systems, and the typology of interaction between weight and sonority.

Now let me turn to the three arguments against collapsing weight and sonority (I rely on the discussion in Kenstowicz 1994). First, a very complex hierarchy like the one in (43) raises the question as to how weight is represented. Weight is thought of standardly as a structurally encoded property. The mora count, a unit of timing, is the device by which heavy syllables are distinguished from light syllables. Collapsing the weight and sonority hierarchies leads to more weight distinctions in a single language than can be reasonably handled by a moraic theory. Some languages which show sonority sensitivity in their stress systems do have stress-relevant weight distinctions. Asheninca (Payne 1981) and Finnish (Anttila ms.) are two such languages. In the most extreme cases, where both weight and sonority play a role in stress assignment, a distinction of as many as five levels of weight may be required. A theory that collapses the weight and sonority hierarchies would have to allow a range of syllables from mono- to pentamoraic.

(50)	CVV, CVC	μμμμμ
	Ca	μμμμ
	Ce,o	μμμ
	Ci,u	μμ
	Cə	μ

The second argument against collapsing weight is sonority is that, as pointed out by Kenstowicz 1994, vowel quality does not necessarily correlate with phonetic length even in those languages that have sonority-driven stress: it is not generally true that length decreases significantly with vowel height.

The third and final argument that the sonority hierarchy in its role in stress assignment is not simply a refinement of the stress hierarchy is that sonority fails to play a role in other weight-sensitive phenomena, most notably word minimality. I know of no languages where the distinction between an acceptable and a subminimal word is made according to sonority: e.g. a language where Ca would be an acceptable word, but Ce and Ci would not.

Based on this discussion, it appears that at present it remains inconclusive whether the vowel sonority hierarchy can be collapsed with the weight hierarchy for the purposes of stress assignment.

3.6.2 Separating prominence from metrical structure

Let me now turn to my proposed refinement of the analysis of sonority-driven stress.

There are three primary correlates of metrical prominence: pitch, duration, and intensity. On the standard view, pitch and duration behave differently than intensity in that they can be used contrastively independently of metrical structure. Both pitch contrasts and contrasts in vowel length can be misaligned with metrical prominence:

some languages tolerate high tones on unstressed syllables, and some languages tolerate unstressed heavy syllables. At the same time, there are universal preferences in languages that attract stress to tones and heavy syllables, or, conversely, that raise the pitch of stressed vowels or lengthen them. While for duration and pitch participate in a bidirectional interaction with metrical structure, the relationship between intensity and stress is commonly thought to be unidirectional: in dynamic stress systems, intensity realizes metrical prominence, but intensity itself, as is generally assumed, cannot be used contrastively independently of metrical structure.

I suggest here that intensity can be brought into line with the other two correlates of stress and treated as a contrastive feature. Constraints on the relationship between stress and intensity would then function analogously to stress-to-weight, weight-to stress, stress-to-tone, and tone-to-stress constraints. Such an analysis empirically depends on finding cases where there is a mismatch between intensity and metrical structure. I will suggest below that cases of this type do indeed exist.

Furthermore, I propose to treat stress-sonority interactions as mediated by intensity. This move has the advantage of limiting the bidirectional interactions of stress to the three main correlates of intensity, duration, and pitch, and for permitting an analysis of systems where metrical structure mismatches with sonority-driven prominence.

3.6.2.1 The dual system of Mari

In a recent paper, Olga Vaysman uncovered a sonority-sensitive stress system that may shed light on the nature of quality-sensitive stress in general (Vaysman 2005). Mari (Finno-Ugric, Russia) distinguishes two levels of sonority: [ə] and the rest of the vowel system. Main stress is assigned in a default-to-opposite fashion: the rightmost full vowel is stressed, but if every vowel in the word is a [ə], then stress falls on the initial syllable. The standard analysis (e.g. Halle and Vergnaud 1987) of such stress systems involves right-headed unbounded feet and an initial default. In a rule-based theory, every full vowel and every initial syllable project a line 1 grid mark, and a right-headed unbounded foot is constructed on line 1.

However, there is direct evidence that Mari also has foot structure which does not coincide with the sonority-sensitive stress. Underlying schwas can surface either as short vowels, or as full vowels whose exact nature depends on the harmonic context. Whether a vowel surfaces as [ə] or as a full vowel depends on syllable count: if the schwa is in an odd-numbered syllable, it will surface as a full vowel, otherwise it remains [ə].

The schwa vocalization process renders the stress assignment rule opaque: in words with no underlying full vowels, stress falls on the initial syllable even if some of the underlying schwas surface as full, because it is assigned at the stage in the derivation when all vowels in the word are still schwas.

Cases of metrical opacity, when different strata in the phonology have different stress rule, have been well-known in the literature. For example, in Huariapano, there is a segmental process of *h*-insertion sensitive to left-to-right syllabic trochees, while the surface stress pattern in most words involves right-to-left syllabic trochees (Parker 1998). In Jarawara (Dixon 2004), surface stress is assigned with right-to-left trochees, while there is a host of segmental processes all sensitive to left-to-right feet. What these cases of metrical opacity have in common is that it is the last round of stress assignment that survives on the surface qua stress; evidence for the earlier rounds comes from segmental processes conditioned by metrical structure that does not survive in the output. The Mari situation is different, however. If Mari were to be treated as a case of opacity, with one stress rule applying at one stratum and the other stress rule at a later stratum, the sonority-sensitive stress must apply first, because it is counterfed by schwa vocalization. The rhythmic stress rule that is responsible for the feet that condition schwa vocalization must then apply at a later stage in the derivation. And yet, it is the output of the earlier rather than the later stress rule that survives on the surface.

Vaysman's way of addressing the problem of the two incompatible stress systems in Mari is to suggest that only one of the two systems, the binary left-to-right feet responsible for schwa vocalization, has to do with metrical constituency. The sonority-sensitive system, contrary to the standard analysis, does not involve unbounded feet.

The Mari case analyzed by Vaysman demonstrates that not all stress systems involve metrical constituent structure, and it is the sonority-sensitive system that operates independently of foot placement. The question then becomes, can the generalization that foot-placement constraints have no access to the segmental details such as vowel quality be saved if all sonority-driven stress systems are analyzed in a Mari-like way that does away with metrical constituency? In this section I pursue an analysis that makes such a separation between constituent-based and prominence-based stress systems.

The separation between prominence and rhythm is not a new idea in the theory of stress. Functionally speaking, stress serves two separate functions: on the one hand, rhythmic organization of the segmental material, and on the other, providing cues to word boundaries. Stress placement constraints fall into two natural classes depending on which of the two functions of stress they cater to: on the one hand there are rhythmic constraints governing the placement of binary feet, such as FTBIN, *LAPSE, *CLASH, etc., and on the other hand, there are constraints such as CULMINATIVITY and edge alignment constraints that serve the second function of stress. The separation between these two sets of preferences and functions of stress has always been at least implicit in stress theories. For example, the analysis of prominence-driven systems in Walker (1997) assumes that not all stress systems involve metrical constituent structure: unbounded systems, on her analysis, are purely prominence-driven, with factors such as syllable weight and sonority, or lexical factors, determining the placement of stress, and with no iterative or unbounded feet. Languages that fall into this category are those with default-to-same and default-to-opposite stress patterns driven by quantitative factors (weight, sonority, length, such as in Mari and Mongolian), or by lexical factors (such as Indo-European accent). Walker's analysis relies on the following sets of constraints.

(51) a. ALIGN constraints: aligning the stressed syllable with a word edge;
b. Licensing constraints: licensing marked stressed syllables (i.e. $\acute{\sigma}_{\mu}$) at an edge;
c. PKPROM: constraint expressing quantity sensitivity
d. NONFINALITY

None of these constraints mention metrical constituent structure: the preferences have nothing to do with rhythmic organization. Instead, the constraints in (51) express two conflicting preferences: to have stress uniformly aligned with an edge, and to preferentially stress prominent syllables.

While this set of constraints works for the systems that Walker analyzed – purely prominence-based stress systems without any evidence for footing – nothing precludes the coexistence of prominence-based stress with prosodic constituency below the word level, as Vaysman's Mari case clearly demonstrates. The interesting cases are those that involve an interaction between footing and prominence.

3.6.2.2 Stress or prominence?

Along the lines of stress-to-tone and tone-to-stress attraction investigated by de Lacy (2003) and the standard OT assumptions about the interaction of stress and weight, I propose the following constraints to regulate the relationship between metrical structure and prominence. Formally, I treat prominence as a feature of vowels, much like nasality or height: a vowel that is [+prom] is realized with increased intensity. The usual set of faithfulness constraints regulates the mapping of this feature between the input and the output, and stress-to-prominence and prominence-to-stress constraints, here formulated in a parallel fashion to ✦SWP and ✦WSP constraints, regulate the relationship between this feature and metrical structure.

(52) Faithfulness constraints
MAX-PROM 'Input [+prom] is realized as output [+prom]'
DEP-PROM 'Output [+prom] realizes input [+prom]'
NOFLOP-PROM '[+prom] is associated with the same segment in the input and the output'

(53) Attraction of prominence to metrical structure

Stress-to-prominence:	S→P	'if stressed, then prominent'
Prominence-to-stress:	P→S	'if prominent, then stressed'

In addition, I posit markedness constraints that regulate the appearance of the [+prom] feature in words independently of stress. These are the constraints responsible for sonority-driven prominence.

(54) a. Prominence-to-sonority
PROM{a}, PROM{a, o/e}, PROM{a, o/e, i/u}, PROM{a, o/e, i/u, ə}
*PROM{ə}, *PROM{ə, i/u}, *PROM{ə, i/u, o/e}, *PROM{ə, i/u, o/e, a}
'if a syllable contains *a*, it is prominent' …
'if a syllable contains *ə*, it is not prominent' …

b. Other constraints

CULM-PROM	'No more than one prominent syllable per phonological word'
*PROMCLASH	'No prominence clash'

On this analysis, the generalization that purely metrical, stress-related constraints do not make reference to the featural details of vowels. Stress can only be attracted to one of the three correlates of metrical structure: quantity, pitch, and loudness. All cases of stress-to-segments attraction are handled in a uniform way: constraints other than stress constraints may cause the attraction of quantity, pitch, or loudness to syllables with particular segmental features, and stress may in turn be attracted by constraints such as those in (53) to the resulting prominence, high pitch, etc. The general picture is given in (55) below.

(55) Metrical structure: FTBIN, ALL-FT-R…

S→P, P→S ToneToStr, StrToTone WSP, SWP

Intensity	Pitch	Quantity
MAX-PROM	MAX-H	MAX-μ
PROM{a}…	OCP, *Contour…	*μμμ, *HH…

At the core of my proposal is the idea that intensity can be treated on a par with the other two primary correlates of stress, pitch and duration. The set of constraints used to deal with intensity-stress interactions is entirely parallel to the constraints used for tone and weight: there is a pair of constraints aligning metrical heads with intensity, in two directions, and there is a set of constraints on the relationship between the feature [prom] and the content of a representation.

3.6.2.3 Factorial typology in standard OT

In this section I outline the predicted typology of stress given the enriched constraint set that separates stress from prominence. I go through the factorial typology generated by OTSoft using the following nine constraints (56). To simplify things, I consider under the heading RHYTHMIC the set of constraints that create a L→R syllabic trochee pattern, and of the sonority-prominence constraints I only include PROM{a}.

(56) a. RHYTHMIC (syllabic trochees L→R)
b. PROM{a}
c. S→P, P→S
d. MAX-PROM, DEP-PROM, NOFLOP-PROM
e. HAVE-STRESS, HAVE-PROM

I consider three inputs: an input with syllables without high-sonority vowels and without lexical stress, an input with a high-sonority vowel, and an input with a lexical stress. These inputs and the outputs included for each of the inputs are given below. Acute accents indicate prominence; the notation *sá* in the output indicates that a syllable has a stressed high-sonority vowel.

(57)	a.	/σ σ σ σ σ/	[(s s)(s s)s]
			[(ś s)(s s)s]
			[s s s s s]
			[ś s s s s]
	b.	/σ σ_a σ σ σ/	[(s s)(s s)s]
			[(ś s)(s s)s]
		2nd syl. has *a*	[s s s s s]
			[ś s s s s]
			[(s sá)(s s)s]
			[s (sá s)(s s)]
			[(ś sá)(s s)s]
			[s sá s s s]
			[ś sá s s s]
	c.	/σ σ́ σ σ σ/	[(s s)(s s)s]
			[(ś s)(s s)s]
		2nd syl. has	[s s s s s]
		lexical prom.	[ś s s s s]
			[(s ś)(s s)s]
			[s (ś s)(s s)]
			[(ś ś)(s s)s]
			[(s ś)(s ś)s]
			[ś ś s s s]

The factorial typology contains 45 output types, classified below based on the stress system type they generate. In each case, the three lines show the outputs of the three inputs from (57): the first line is for an input without any high-sonority vowels or lexical stress, the second line for the input with a high-sonority vowel in the second syllable, and the third line for the input with lexical stress on the second syllable. Thus, each column of three lines represents an output pattern.

The first pattern is the simplest: rhythmic stress with no sonority sensitivity. Here RHYTHMIC must outrank MAX-PROM, PROM{a}, and P→S. The ranking of S→P and DEP-PROM determines whether the binary rhythmic stress is realized overtly or covertly. There are many languages like this: any language with syllabic trochees built on the left edge without any sensitivity to factors such as vowel quality and without any exceptions.

I. Rhythmic stress, no sonority sensitivity, no faithfulness

RHYTHMIC 〉〉 MAX-PROM, PROM{a}, P→S

```
(ś s)(s s)s         (s s)(s s)s
(ś s)(s s)s      (s s)(s s)s
(ś s)(s s)s         (s s)(s s)s
```

The second pattern involves lexically unpredictable stress without sonority sensitivity, where MAX-PROM and P→S must outrank RHYTHMIC and DEP-PROM must outrank the sonority constraint PROM{a}. The system generates three possible outcomes. First, there is the system with a rhythmic default at the left edge but some lexical exceptions with stress on the second syllable. This involves the high-ranked S→P, forcing all stressed syllables to be realized as prominent. If this constraint is low-ranked, metrical structure will be covert just in those cases where there is no lexical stress (this pattern properly belongs with category VI below). In such a language, there could be a potential contrast between overt and covert metrical structure. The second sub-pattern in Class II involves metrical structure appearing only if there is unpredictable lexical stress, which also belongs in Class VI.

Languages instantiating the first subpattern include Polish and Macedonian, where a rhythmic default can have exceptions. The second and third subpatterns is harder to find for dynamic stress, but its analogue in pitch accent languages is common, where it is normal to have a contrast between accented and unaccented words.

II. Lexically unpredictable stress

MAX-PROM, P→S 〉〉 RHYTHMIC; DEP-PROM 〉〉 PROM{a}

```
(ś s)(s s)s         (s s)(s s)s
(ś s)(s s)s         (s s)(s s)s
s(ś s)(s s)         s(ś s)(s s)
```

The third class of outputs includes sonority-driven stress. Once again, there are two subpatterns here: one where metrical structure is overt in all words, and one where it is covert in words without high-sonority vowels. Examples of the former are any language with sonority-driven stress and footing dependent on it (e.g. Gujarati in de Lacy's description, Asheninca Campa). I am not aware of the examples of the second subpattern.

III. Sonority-driven stress

PROM{a}, P→S ⟩⟩ RHYTHMIC ⟩⟩ MAX-PROM, DEP-PROM

```
(ś s)(s s)s        (s s)(s s)s
s(sá s)(s s)       s(sá s)(s s)
(ś s)(s s)s        (s s)(s s)s
```

The system also predicts a combination of the patterns II and III, where both sonority sensitivity and lexical stress coexist.

II & III combined

PROM{a}, P→S, MAX-PROM ⟩⟩ RHYTHMIC, DEP-PROM

```
(s s)(s s)s        (ś s)(s s)s
s(sá s)(s s)       s(sá s)(s s)
s(ś s)(s s)        s(ś s)(s s)
```

Now we get to cases of mismatch between metrical structure and prominence. Class IV patterns have a default rhythmic structure throughout, but in case a word has a lexically marked prominent syllable, that syllable may not fall in the strong position of the foot.

IV. Mismatch between lexical prominence and metrical structure

RHYTHMIC, MAX-PROM ⟩⟩ P→S, PROM{a}

```
(ś s)(s s)s        (s s)(s s)s
(ś s)(s s)s        (s s)(s s)s
(s ś)(s s)s        (s ś)(s s)s
```

Next comes the type of stress system that was the original motivation for the reanalysis of stress proposed here: Mari-like mismatch between sonority-driven prominence and metrical structure. Once again, two variations on the system are predicted, depending on the ranking of S→P: cases where all feet are overt and cases where no prominence is realized just in case there is no high-sonority vowel in the word.

V. Mismatch between sonority-driven prominence and metrical structure

RHYTHMIC, PROM{a} ⟩⟩ MAX-PROM, P→S

```
(ś s)(s s)s        (s s)(s s)s
(s sá)(s s)s     (s sá)(s s)s
(ś s)(s s)s      (s s)(s s)s
```

The system also predicts a combined case of mismatch between rhythmic structure, sonority-driven stress, and lexical stress.

IV & V combined

RHYTHMIC, MAX-PROM, PROM{a} ⟩⟩ P→S

```
(ś s)(s s)s      (s s)(s s)s
(s sá)(s s)s       (s sá)(s s)s
(s ś)(s s)s        (s ś)(s s)s
```

Next come some potentially problematic cases. There is a large and heterogeneous class of outputs with some sort of contrast between lexically accented and lexically unaccented words. Not many of these outputs are represented in the typology; there appear to be few languages with dynamic stress where there is a contrast between accented and unaccented words, although such a pattern is common in the pitch accent analogues of the stress systems.

VI. Contrast between accented and unaccented words (possibly dependent on sonority)

```
s s s s s      (s s)(s s)s  s s s s s    s s s s s      s s s s s
s s s s s      (s s)(s s)s  s sá s s s   s(sá s)(s s)   s(sá s)(s s)
(ś s)(s s)s    (ś s)(s s)s  s(ś s)(s s)  (ś s)(s s)s    s s s s s

s s s s s      s s s s s    s s s s s    s s s s s      (s s)(s s)s
s(sá s)(s s)   s sá s s s   s sá s s s   s s s s s      (s sá)(s s)s
s(ś s)(s s)    (ś s)(s s)s  (s ś)(s s)s  (s ś)(s s)s    (ś s)(s s)s

(s s)(s s)s    (ś s)(s s)s  (ś s)(s s)s  (ś s)(s s)s    s s s s s
s(sá s)(s s)   s sá s s s   s sá s s s   s sá s s s     s s s s s
(ś s)(s s)s    (ś s)(s s)s  (s ś)(s s)s  s(ś s)(s s)    s(ś s)(s s)
```

Another systematic set of patterns that are not attested involves "too much stress", i.e. cases where intensity realizes two independent things in a language.

VII. Too much stress

```
(ś s)(s s)s  (ś s)(s s)s    (ś s)(s s)s   (ś s)(s s)s   (ś s)(s s)s
(ś s)(s s)s  (ś sá)(s s)s   (ś sá)(s s)s  (ś sá)(s s)s  (ś sá)(s s)s
(ś ś)(s s)s  (s s)(s s)s    (ś s)(s s)s   s s s s s     (ś ś)(s s)s

(ś s)(s s)s  s s s s s      s s s s s
s sá s s s   s s s s s      s sá s s s
(ś ś)(s s)s  (ś ś)(s s)s    (ś ś)(s s)s
```

Below I give the rankings that generate the patterns in VII.

(58) a. S→P, RHYTHMIC, PROM{X} ⟩⟩ MAX-PROM
both sonority-driven prominence and rhythmic stress
b. S→P, MAX-PROM, PROM{X} ⟩⟩ RHYTHMIC
both sonority-driven prominence and unpredictable prominence
c. S→P, RHYTHMIC, MAX-PROM ⟩⟩ PROM{X}
both rhythmic stress and unpredictable prominence
d. all four constraint types high-ranked
all three stress types in one system: sonority, unpredictable prominence, rhythm

A system such as (58)a would have both sonority-driven prominence and, independently of it, a rhythmic stress pattern realized dynamically. All of the other systems in (58) would also have two independent stress patterns realized with the same cue.

This problem is not particular to my analysis, however: the constraints on tone-to-stress alignment also predict parallel 'too-much-tone' systems where a rhythmic pattern realized tonally coexists with lexically unpredictable tone.

(59) RHYTHMIC, STRESS-TO-TONE, MAX-H ⟩⟩ TONE-TO-STRESS
Unpredictable tones combined with a rhythmic tonal pattern

As (59) shows, the problem is not limited to my account: any theory that has both tone-to-stress and stress constraints and tone faithfulness constraints predicts that having both high-ranked would create a language with two separate functions of tone: it would realize both metrical constituent structure and appear on some vowels unpredictably, protected by the faithfulness constraints.

Although the constraint system systematically does not rule out such 'too-much-stress' and 'too-much-tone' systems, it is plausible to assume that they do not occur for other reasons, external to the OT grammar. Once again, this problem is not unique to my proposal, but is faced by any theory of stress where faithfulness plays a role, and is not unique to prosodic phonology either: too much faithfulness in segmental phonology can also lead to predicted but implausible languages. I follow the orthodox view that the non-existence of such systems does not pose a serious challenge for the theory.

Another small set of patterns predicted by my system involve no metrical structure at all.

VIII. No rhythmic structure

```
s s s s s        s s s s s
s s s s s        s sá s s s
s s s s s        s s s s s
```

And finally, there are a few oddball cases where the system predicts a rather bizarre outcome: stress is *not* realized just in case it is marked lexically. This arises because high-ranked NOFLOP, RHYTHMIC, and S→P make it impossible to realize stress in non-default positions: realizing it on the second syllable would violate S→P, shifting foot structure one syllable to the right would violate RHYTHMIC, and realizing it on the first syllable would violate NOFLOP. It remains to be seen how such cases can be ruled out.

IX. Stress not realized just in case it's marked lexically

```
(ś s)(s s)s     (ś s)(s s)s     (ś s)(s s)s       (ś s)(s s)s
(ś s)(s s)s     (ś s)(s s)s     (s sá)(s s)s      s sá s s s
s s s s s       (s s)(s s)s     (s s)(s s)        s s s s s

(ś s)(s s)s     (ś s)(s s)s
s(sá s)(s s)    s(sá s)(s s)
(s s)(s s)s     s s s s s
```

Finally, what the ill-behaved cases in VI-IX have in common is "extreme arrhythmia" (Paul Kiparsky, p.c.). They have either adjacent stresses (ś ś), or no stress at all (s s s s s). Once again, such extremely dysfunctional systems, predicted by any theory of stress that has faithfulness to prosodic structure, can be ruled out on general grounds of culminativity.

Although the typology of stress outlined here is richer than the typology predicted by the standard theory that does not treat intensity as potentially contrastive, in some areas my proposal makes more restrictive predictions. A hypothesis implicit in the preceding discussion has been that lexically unpredictable stress arises due to the action of faithfulness constraints that refer not to stress (i.e. metrical structure) directly but to one of its correlates, e.g. intensity. I propose here that this is the only way for lexical stress to arise: there are no faithfulness constraints that refer directly to stress, expressing the idea that the function of prosodic constituent structure is not to create contrasts for the language but to organize the material into manageable chunks.

In other words, the only way that unpredictable stress can arise is through the action of faithfulness constraints that operate on intensity, pitch, and quantity. MAX-PROM can lead to unpredictable stress such as in Russian, as outlined above; MAX-H creates unpredictable stresses in pitch accent languages such as Ancient Greek, and MAX-μ creates weight-sensitive stress.

Doing away with stress faithfulness constraints that refer to metrical structure, e.g. NO-FLOP-STRESS, predicts that *covert prosodic structure is always predictable*. If a language has only abstract feet – structure created by the action of constraints such as FTBIN or AL-FT-L, but not aligned to either pitch or intensity, such a stress system ought to have no exceptions. There should be no languages like Capanahua where the covert feet are left-aligned in some arbitrary set of words and right-aligned in others. Likewise, there should be no covert contrast in the placement of inaudible foot boundaries: because the are no faithfulness constraints referring to foot boundaries, there should be no contrast between, say, (σ σ) σ σ and (σ σ σ) σ.

3.7 Some applications

My proposal in this section is not specific to prosody-segmental interactions. Rather, I suggest that the too-many-solutions problem that is most clearly seen in the domain of stress-segmental interactions is in fact present elsewhere in OT and result from its general claim that all generalizations are in the outputs. I have argued that this position is too radical; prosody-segmental interactions are subject to a different type of generalization. In Chapter 4 below I will discuss in detail another domain where procedural generalizations are crucial – the typology of vowel syncope and epenthesis. In this section I briefly survey some other cases in phonology where my proposal can prove fruitful, leaving a detailed investigation for future research.

In (60) I give a sample of segmental constraints stated procedurally and the typological consequences of the ICP for each constraint. In each case, the effect that is ruled out involves modifying the property mentioned in the antecedent of the constraint.

(60) a. '✦If a velar is followed by a front vowel, then it is palatalized'
☞No vowel backing after velars
b. '✦If a consonant is intervocalic, then it is voiced'
☞No vowel deletion to make voiceless consonants not intervocalic
☞No consonant epenthesis to make voiceless consonants not intervocalic
c. '✦Short front vowel followed by *r* are realized as [e]' (cf. Latin)
☞*r* cannot be deleted, turned into *l*, etc. just after short front vowels
d. '✦If a nasal is followed by a consonant, then it is homorganic with that consonant'
☞No epenthesis to relieve non-homorganic NC clusters
☞No deletion of C in non-homorganic NC clusters

Now consider harmony, which I will discuss in more detail. A notorious problem in any OT analysis of harmony is that the constraint penalizing disharmonic configurations does not say anything about how such configuration can be repaired. Consider a simple example of nasal harmony modeled after Pater 2003. Suppose that a language has progressive nasal harmony, and that glides are transparent to harmony while voiceless fricatives block it. Sample inputs and outputs and a tableau are given below.

(61) a. /ãwa/ → ãw̃ã
b. /ãsa/ → ãsa

(62)

			MAX[nas]	*s̃	AGREE	DEP[nas]
/ãwa/		ãwa			*!	
	☞	ãw̃ã				**
		awa	*!			
/ãsa/	☞	ãsa			*	
		ãs̃ã		*!		**
		asa	*!			

The problem with such an analysis is that the factorial typology is too rich. Given the input /ãsa/, the grammar in (62) chooses to sacrifice harmony for the sake of avoiding the marked segment *s̃*, and for the sake of avoiding a violation of MAX[nas]. However, nothing precludes a ranking where both harmony and **s̃* dominate the faithfulness constraint, resulting in a language that applies harmony where it can and deletes the harmonizing feature in blocking contexts. Such systems are unattested.

(63)

			*s̃	AGREE	MAX[nas]	DEP[nas]
/ãwa/		ãwa		*!		
	☞	ãw̃ã				**
		awa			*!	
/ãsa/		ãsa		*!		
		ãs̃ã	*!			**
	☞	asa			*	

The reason this problem arises is the same reason that was behind the incorrect predictions with respect to prosody-segmental interactions: there is a generalization about input-output mappings at work. This means that AGREE is a procedural constraint subject to the ICP. One possible statement of the constraint is given below, where the word 'domain' stands for the relevant harmony domain of the language (in our hypothetical example, the word).

(64) ✦AGREE
If *x* is a segment within a domain that contains a [+nas] segment, then *x* is [+nas].

The candidate *asa* in tableau (63) is exactly the type of candidate that the ICP is intended to rule out. In order to satisfy the implicational statement of the harmony constraint (64), this candidate modifies the property that is mentioned in the antecedent of the constraint rather than in the consequent. The ICP rules this out, as shown in the following tableau (65). Both the candidates *awa* and *asa* incur ✦ violations of the AGREE constraint by the ICP, because they have segments that are not within a domain that has a [+nas] segment while their t-correspondents in the designated candidates *ãwa* and *ãsa* are within such a domain. Incorporating the effects of ICP into the ranking will make it impossible for the candidates to win if the harmonizing feature is deleted just in those cases where harmony is impossible.

(65)

				*s̃	✦AGREE	MAX[nas]	DEP[nas]
/ãwa/		✦	ãwa		*!		
	☞		ãw̃ã				**
			awa		✦!	*!	
/ãsa/	☞	✦	ãsa		*		
			ãs̃ã	*!			**
			asa		✦	*!	

The full factorial typology for the four constraints contains only three patterns: (1) harmony applying to both /ãwa/ and /ãsa/, (2) harmony applying to neither, and (3) harmony applying to /ãwa/ but not to /ãsa/. In this particular case, the designated candidates of the constraint AGREE is always *ãwa* and *ãsa*, no matter what the ranking of the remaining three constraints, so the factorial typology can be computed simply by reranking the constraints as given in (65).[21]

(66) a. DEP ⟩⟩ AGREE (12 grammars)

Harmony applies in no forms

b. $\left\{ \begin{array}{c} \text{AGREE} \rangle\rangle \text{DEP} \\ {*}\tilde{s} \rangle\rangle \text{AGREE} \end{array} \right\}$ (6 grammars)

Harmony applies in /ãwa/ but not /ãsa/

c. $\left\{ \begin{array}{c} \text{AGREE} \rangle\rangle \text{DEP} \\ \text{AGREE} \rangle\rangle {*}\tilde{s} \end{array} \right\}$ (6 grammars)

Harmony applies in all forms

Now consider a more complicated situation. In Guaraní and some related languages, the domain of nasal harmony is defined by the main stress foot: nasal harmony proceeds right-to-left up to the stressed syllable. One problematic prediction of standard OT is that it should be possible for stress to shift onto final nasal vowels just to prevent nasal harmony from applying. Such an outcome is given in the tableau below.

(67)

			*s̃	AGREE$_{\varphi}$	MAX[nas]	DEP[nas]	STRESSINITIAL
/tawã/		áwã		*!			
		ã́w̃ã				*!*	
	☞	awã́					*
		áwa			*!		
/tasã/		ásã		*!			
		ã́s̃ã	*!			**	
	☞	asã́					*
		ása			*!		

However, given the ICP, the unwanted outcome is no longer possible. The designated candidates of the constraint AGREE always have initial stress, because there is no constraint that can override default stress assignment. The candidates *awã́* and *asã́*

[21] Adding other constraints to the picture will, of course complicate matters somewhat. For example, high-ranking *NASALVOWEL will cause the designated candidates for AGREE to be *awa* and *asa*, resulting

violate the constraint ✦AGREE$_\varphi$ by : the first vowel *a* in *awã́* does not have the property of being in a stress foot that contains a [+nas] segment, while its correspondent in the DC does have this property. Likewise, the first vowels of *áwa* and *ása* also do not have the property in question, due to the absence of [+nas] feature in the domain, while their correspondents in the respective DCs do have the property. These candidates therefore incur ✦violations of ✦AGREE$_\varphi$.

(68)

				*s̃	✦AGREE$_\varphi$	MAX[nas]	DEP[nas]	STRESSINITIAL
/tawã/		✦	áwã		*!			
	☞		ã́w̃ã				**	
			awã́		✦!			*
			áwa		✦!	*		
/tasã/	☞	✦	ásã		*			
			ã́s̃ã	*!			**	
			asã́		✦			*!
			ása		✦	*!		

Once again, the factorial typology can be computed in a simple way, since the choice of the designated candidates of the constraint ✦AGREE$_\varphi$ does not depend on the ranking of the other constraints. The typology contains only three output patterns: (1) harmony applies in both /awã/ and /asã/; (2) harmony applies in neither; (3) harmony applies in /awã/ but not in /asã/ (as is the case in (68). The rankings are given in (69).

(69) a. DEP ⟩⟩ AGREE (60 grammars)
Harmony applies in no forms

b. *s̃ ⟩⟩ AGREE ⟩⟩ DEP (30 grammars)
Harmony applies in /awã/ but not in /asã/

c. $\left\{\begin{array}{c}\text{AGREE} \rangle\rangle \text{ s̃} \\ \text{AGREE} \rangle\rangle \text{ DEP}\end{array}\right\}$ (30 grammars)
Harmony applies in all forms

---

in outputs where [nas] is removed from vowels across the board.