CHAPTER 2

MARKEDNESS, ECONOMY, AND *STRUC

2.1 Introduction

In Optimality Theory (Prince and Smolensky 1993), to be marked means to violate a markedness constraint. Yet without formal restrictions on the content of markedness constraints, practically everything can be and sometimes is assumed to be marked. In this chapter, I propose an amendment to this view. I argue that markedness constraints are limited in what they can assign violation marks to—for every markedness constraint, there is at least one non-null structure that fully satisfies it. In this sense, markedness constraints are *lenient*.

This view is formally implemented as a theory of the constraint module CON. Markedness constraints are derived from harmonic scales that compare non-null structures with each other. No markedness constraint penalizes the most harmonic element on a scale, and no harmonic comparison is nihilistic. This means that no individual constraint is set up to prefer the absence of structure to every other alternative—there are no economy constraints in the grammar.

Although no individual constraint is an economy constraint, the interaction of constraints in a language-specific grammar can result in what appears to be minimization of structure—that is, *economy effects*. Yet there is nothing about economy effects that would suggest an overarching "principle of least effort" or general economy principle—the effects can always be reduced to the interaction of independently motivated constraints. These constraints can be shown to have other effects in the grammar—effects that do not result in economy of any kind. The reason for this is that every marked configuration can be avoided in a variety of ways—McCarthy (2002b) dubs this property of OT grammars *homogeneity of target, heterogeneity of process*. Deletion of structure is just one way to remove a marked configuration, but because there is always a less marked thing out there, *change* of structure should also be an option.

This view of economy effects is not universally accepted. Formal economy principles are often thought to be a necessary property of generative grammar because human language is recursive, which means that grammars must be able to produce structures of unbounded size. To limit this troubling but necessary ability, both syntacticians and phonologists have relied on economy principles, which range from the very general "Avoid Structure" (Rizzi 1997) to the fairly specific constraint against syllables *STRUC(σ) (Zoll 1993, 1996), its precursor the Syllable Minimization Principle (Selkirk 1981), and many others.[4]

One of the consequences of the present proposal is that economy constraints like *STRUC(σ) are excluded from CON as a matter of principle. This turns out to be a wecome result, because economy constraints are redundant in the theory where all economy effects result from constraint interaction. Not only are economy constraints redundant—

[4] For discussion of economy principles in syntax, see Chomsky and Lasnik 1977, Chomsky 1989, 1995, Grimshaw 2003, Poole 1998, Woolford 1995, and various papers in Barbosa et al. 1998. For discussion of economy principles in phonology, see Broselow 1995, Lindblom 1983, McCarthy 2002b, Noske 1984, Prince and Smolensky 1993/2002, Zoll 1993, 1996. For various applications of *STRUC constraints, see Causley 1997, Davis and Zawaydeh 1996, de Lacy 1999, Fukazawa 1999, Hewitt and Crowhurst 1996, Orgun 1996, Prince and Smolensky 1993/2002, Raimy and Idsardi 1997, Selkirk 2000, Truckenbrodt 1999, Walker 2003, Zoll 1993, 1996. Several of these works will be addressed in some detail in the coming pages.

they are also harmful. Their very presence in CON predicts that certain deletion processes should target structure that is unmarked (e.g., syllables regardless of metrical context), and this prediction is not supported by typological evidence.

This proposal for the reformation of CON puts another set of constraints in a questionable position: gradient alignment constraints (McCarthy and Prince 1993a, Prince and Smolensky 1993). Although gradient alignment constraints are not formally equivalent to economy constraints, their effects are very similar—both sets of constrains can keep track of the lengths of outputs. Some of the typological arguments against *STRUC constraints readily extend to alignment constraints. Interestingly, the present theory encounters some difficulty in relating alignment constraints to scales—they require either scales of infinite length or additional formal mechanisms. Thus this work adds to the arguments of McCarthy (to appear) that gradience cannot be a property of OT constraints.

The rest of the chapter is organized as follows. Section 2.2 presents the theory of the constraint set CON and discusses some of its implications for the formulation of constraints. In §2.3, I show how the interaction of independently motivated constraints produces a wide range of economy effects, and in §2.4 I provide a formal definition for *STRUC constraints and show how and why they should be excluded from the theory. Section 2.6 concludes.

2.2 The theory of CON: scales and Lenient Constraint Alignment

2.2.1 Introduction

Markedness is a matter of comparing non-null forms to each other rather than an abstract, platonic property: no form is marked except insofar as it compares to another non-null form.[5] Null structures vacuously satisfy all markedness constraints—they do not need to be specially favored by them. This section presents a theory of the constraint module CON that formally develops this idea. The theory has two components. First, all markedness constraints must be derived from harmonic scales and can never penalize the least marked member on a scale—they are lenient. Second, the scales themselves must meet certain requirements: they cannot imply that ∅ is more marked than a non-null form.

In the remainder of this section, I start by looking at harmonic scales and harmonic alignment of Prince and Smolensky 1993, which forms an important background to the proposal. Section 2.2.4 presents Lenient Constraint Alignment and §2.2.5 lays out the principles that harmonic scales must obey. Section 2.2.6 explores some of the issues in relating various kinds of markedness constraints to scales. Section 2.2.7 discusses the Null Output, which plays an important role in the proposal, and addresses its status in the present theory.

2.2.2 Harmonic scales

Optimality Theory does not necessarily offer guidelines for what markedness constraints can militate against, though a constraint's validity can be tested by examining the typological consequences of introducing it into CON. The theory of CON developed here looks at markedness constraints from another angle. Whether or not M is a valid constraint depends on the harmonic comparisons it implies; some comparisons are argued

[5] The proposal developed here is quite distinct from Comparative Markedness (McCarthy 2002c):

to be invalid. For every constraint, the markedness comparison must be encoded in a *harmonic scale*.

A harmonic scale orders linguistic entities along some dimension of markedness (Prince and Smolensky 1993). For example, nasal vowels are universally more marked than oral ones (McCarthy and Prince 1995). This is reflected in the following binary harmonic scale ("$\succ$" means "is more harmonic than"):

(1) *Vowel nasality scale*: oral vowel $\succ$ nasal vowel

Similarly, voiced obstruents are universally more marked than voiceless ones (Lombardi 1995, 2001), which can also be stated in terms of a scale:

(2) *Obstruent voicing scale*: voiceless obstruent $\succ$ voiced obstruent

Harmonic scales are not new or unique to this theory. Prince and Smolensky 1993 introduce harmonic scales that encode the relative well-formedness of syllable onsets (margins) and nuclei (peaks) depending on their sonority; the more sonorant a nucleus, the better. For onsets, the opposite is true:

(3) *Peak harmony scale*: pk/a $\succ$ pk/ i $\succ$ ... $\succ$ pk /t

(4) *Margin harmony scale*: m/t $\succ$ ... $\succ$ m/i $\succ$ m/a

These scales are derived from *prominence scales*. *Prominence scales* are not statements of markedness; rather, they are orderings of linguistic entities according to salience. For example, a syllable peak is a more prominent position than a syllable margin, and a sonorant segment is more prominent than an obstruent (">" stands for "is more prominent than"):

(5) *Peak/margin prominence scale*: peak > margin

(6) *Sonority scale*: a > i > ... > t

There is a preference for prominent positions to be occupied by prominent segments, and vice versa. The formal mechanism Prince and Smolensky devise for capturing this preference is called *Harmonic Alignment*:

(7) Suppose given a binary dimension D_1 with a scale X > Y on its elements {X, Y}, and another dimension D_2 with a scale a > b > … > z on its elements. The *harmonic alignment* of D_1 and D_2 is the pair of Harmony scales:

H_X: $X/a \succ X/b \succ \ldots \succ X/z$ [more harmonic … less harmonic]

H_Y: $Y/z \succ \ldots \succ Y/b \succ Y/a$ (Prince and Smolensky 1993:155)

Harmonic Alignment has been used extensively in OT to derive harmonic scales—it has been applied to sonority and stress (Kenstowicz 1996b), syntactic person and subject/object (Aissen 1999, Artstein 1998), and tone (de Lacy 2002b).

So, some harmonic scales are primitive (e.g., the vowel nasality scale and the obstruent voicing scale), while others are derived by Harmonic Alignment.[6] Primitive scales may be based on substantive principles: nasal vowels are perceptually weaker than oral ones, while voiced obstruents are marked for aerodynamic reasons. Apart from expressing linguistically sound tendencies, scales must meet certain formal requirements—these will be discussed in §2.2.5. I now turn to the procedure for mapping harmonic scales to constraints.

---

[6] De Lacy (2002a) lays out some principles for determining which scales are derived and which are primitive. In his theory, featural markedness scales (e.g., vowel nasality) never combine with structural elements for the purposes of constraint construction, while prominence scales (e.g., sonority) always do. This is basically what I assume here.

2.2.3 The Constraint Alignment of Prince and Smolensky 1993

Harmonic scales are not constraints: they cannot evaluate candidates and they cannot interact with other constraints in a ranking. For creating constraints from harmonic scales, Prince and Smolensky 1993 propose a different operation: Constraint Alignment (defined in (8)). Constraint Alignment assigns each element on a harmonic scale to a negatively stated markedness constraint. The result is a fixed hierarchy of constraints, whose order is the reverse of the relevant harmonic scale.

(8) The *constraint alignment* is the pair of constraint hierarchies:

a. C_X:*X/Z >> … *X/B >> *X/A [more marked >> … >> less marked]
b. C_Y:*Y/A >> *Y/B >>…>>*Y/Z (Prince and Smolensky 1993:155)

When this version of Constraint Alignment applies to the peak/margin hierarchies, it yields the following two constraint hierarchies:[7]

(9) *Peak constraints*: *NUC/t>>...>>*NUC/i >>*NUC/a

(10) *Margin constraints*: *ONS/a>>*ONS/i>>...>>*ONS/t

From the vowel nasality scale, a binary hierarchy is produced, where the constraint against unmarked oral vowels is universally ranked below the constraint against nasal vowels:

(11) *NASALV>>*ORALV (McCarthy and Prince 1995)

Fixed rankings are not a necessary aspect of this theory of markedness—the same markedness relationship can be expressed through constraints in a stringency relation (de Lacy 2002a, Prince 1997a). De Lacy proposes a version of Constraint Alignment that

---

[7] Prince and Smolensky call the constraints *P/x and *M/x instead of *NUC/x and *ONS/x. I will use *NUC/x and *ONS/x throughout to distinguish the syllable peak/margin constraints from the foot peak/margin constraints (Kenstowicz 1996b).

produces not fixed rankings but rather stringent constraint hierarchies, which impose the same harmonic orderings on the candidate set even when their ranking is permuted. For example, based on the obstruent voicing scale, there will be two constraints formulated in such a way that their ranking never results in voiceless obstruents being more marked than voiced ones, as shown in (12). The relative markedness of voiced and voiceless obstruents is invariant under re-ranking: regardless of the ranking of *VOICEDOBS and *OBS, the voiceless obstruent candidate incurs fewer constraint violations and is therefore universally less marked.

(12) Stringent constraints: {*VOICEDOBS, *OBS}

	*VOICEDOBS	*OBS
a. pa		*
b. ba	*	*

Whether these hierarchies are freely rankable or in a fixed ranking, they share a common feature: the hierarchies contain constraints against the least marked thing on the scale. *OBS or *VOICELESSOBS are essentially economy constraints—they have no other purpose but to penalize unmarked structure (I will return to constraints of this sort in §2.5). I propose to modify Constraint Alignment so that constraints against the unmarked are excluded from CON as a matter of principle.

2.2.4 Lenient Constraint Alignment

In the model of CON advocated here, all markedness constraints are derived from harmonic scales by an operation similar to Prince and Smolensky's Constraint

Alignment.[8] The difference is that every element on every scale has a corresponding markedness constraint against it except for the least marked one. The least marked element on every scale gets an "exemption." This *Lenient* Constraint Alignment is defined as follows:

(13) **Lenient Constraint Alignment**
The Constraint Alignment of a harmonic scale $a_n \succ a_{n+1} \succ \ldots\ a_{m-1} \succ a_m$ is the constraint hierarchy $*A_m >> *A_{m-1} \ldots >> *A_{n+1}$.

The most harmonic member of every scale, a_n, does not correspond to any constraint. The lowest-ranked constraint in the hierarchy militates against the next most harmonic member, a_{n+1}. This is the chief difference between (13) and Prince and Smolensky's version.

To see how LCA works, consider the obstruent voicing scale. The least marked element in the scale is *voiceless obstruent*. According to LCA, every element in the scale except the least marked one is assigned to a markedness constraint. There is only one such element in the scale, *voiced obstruent*, so only one constraint is derived: *VOICEDOBS. The unmarked element in the scale, *voiceless obstruent*, has no corresponding markedness constraint against it.

(14) *VOICEDOBS: *[+voice, -son] "voiced obstruents are prohibited."
Harmonic scale: voiceless obstruent $\succ$ voiced obstruent

When LCA applies to a longer scale, the result is the same: the constraint against the least marked element in the peak harmony scale, low vowels, is left off the resulting constraint

[8] In a footnote on p. 453, Ito and Mester 1997 suggest that constraints may be "formally understood as zero-level preference relations holding between linguistic structures." This is exactly what Lenient Constraint Alignment allows us to do.

hierarchy. For syllable onsets, the result is the same: the scale does not contain a constraint *NUC/t against voiceless obstruent onsets.

(15) *Syllable Peak Constraints:* *NUC/s>> *NUC/n.... >>*NUC/i
*NUC/a is not a constraint

(16) *Syllable Margin Constraints*: *ONS/a>>...*ONS/n>>*ONS/s
*ONS/t is not a constraint

This approach formalizes an intuition that other researchers have expressed: constraints should penalize only marked things. For example, Clements 1997 voices a concern about "anti-tendency" constraints like *NUC/a and *ONS/t:

(17) ...Voiceless stops are optimal syllable margins across languages; all known languages syllabify voiceless stops as margins in at least some circumstances, and the great majority do in all circumstances. We might say instead that this constraint expresses an *antitendency*—the contrary of a universally observed tendency—which is regularly and consistently violated in all known languages...[*NUC/a] encapsulates the statement that 'members of sonority class *a* [low vocoids] must not be parsed as a syllable Peak.' This statement ... expresses an antitendency, since low vocoids constitute the optimal representative of the class of syllable peaks across languages.[9] (Clements 1997:299-300)

In the same vein, Pater 1997 excludes the constraint against voiceless obstruent onsets from his onset sonority constraint hierarchy, and de Lacy 2002a argues (following Kiparsky 1994) that unmarked things are not protected by special faithfulness constraints, whereas marked things are.

---

[9] Clements actually goes on to add that constraints against consonantal margins and vocalic nuclei in general are "antitendency" constraints—e.g., languages don't usually balk at parsing most consonants as syllable margins, just as they do not shrink away from vocalic nuclei. There is some evidence of these constraints' activity. Pater 1997 discusses evidence for constraints against the more sonorant consonants as onsets in child speech, and there is also evidence from reduplication in adult languages such as Sanskrit (Steriade 1988). In chapter 4, I discuss various evidence for the constraints against low-sonority syllabic nuclei.

Kiparsky 1994 also discusses markedness constraints, although his approach is to doubly punish marked things rather than favor unmarked things—for example, he has constraints against labial and dorsal place and constraints against consonantal place in general. The latter constraint is not possible under Lenient Constraint Alignment, assuming that unmarked consonantal place is the least marked element on the place scale. Lenient Constraint Alignment ensures that unmarked things enjoy a special, markedness-free status in the grammar: they are literally unmarked because they do not violate the relevant markedness constraints.

Anchoring all constraints in scales brings up the issue of how the resulting constraints express hierarchical markedness relations—stringently or through a universally fixed ranking. This issue arises whenever a scale has three or more levels, i.e., when two or more constraints are derived from it. Since the arguments about stringency/fixed rankings are of little relevance to the topic of economy and would detract too much from the main concern of this chapter, I refer the reader to the extensive discussion in the works of Prince (1997b, 1997c, 1999) and de Lacy (1997, 2002a). What I will do here is provide a modified version of Constraint Alignment that is compatible with the stringent formulation of hierarchical constraints.

The stringency version of Lenient Constraint Alignment is based on de Lacy's schema for scale-referring markedness constraints, given in (18). De Lacy's definition maps every element in the scale to a markedness constraint. In the Lenient theory, the modification is to exclude the least marked element (see (19)).

(18) **Featural scale-referring markedness constraints (**de Lacy 2002a:30)

For every element *p* in every scale S, there is a markedness constraint *m*.
m assigns a violation for each segment that either
(i) contains *p*
or (ii) contains anything more marked than *p* in scale S.

(19) **Lenient Constraint Alignment (stringent version)**

For every element $a_i| i > n$ in scale S ($a_n \succ a_{n+1} \succ ... a_{m-1} \succ a_m$), there is a markedness constraint C_M.
C_M assigns a violation to every element that
(i) contains a_i
or (ii) contains anything more marked than a_i in scale S.

Given a scale X≻Y≻Z, (19) yields two constraints—one that penalizes only Z, one that penalizes Z or Y, and none that refer to X. Regardless of the ranking of *Z and *Z-OR-Y, candidate X emerges as the least marked, Y as more so, and Z as the most marked member of the set. No constraint penalizes X, Y, *and* Z:

(20) Stringent constraints generated by LCA

	*Z	*Z-OR-Y
a. X		
b. Y		*
c. Z	*	*

Just like the fixed ranking version of LCA (13), the stringent LCA maps every member of the scale to a constraint except for the least marked member.

Simply leaving the least marked member of every scale off of the resulting constraint hierarchy does not by itself rid CON of economy constraints—for that, the harmonic scales themselves must meet certain requirements. These requirements are discussed in the next section.

2.2.5 Requirements for harmonic scales

Formally, scales are defined as *partial orders*: they are irreflexive, transitive, and asymmetric.[10] A scale cannot state that something is more marked than itself, and it cannot reverse the markedness relation that it itself imposes. This means that scales of the following sort are illegitimate:

(21) Illegitimate scales

a. $\mathbf{x} \succ \mathbf{x}$ (not irreflexive)
b. $\mathbf{x} \succ \mathrm{y} \succ \mathrm{z} \succ \mathbf{x}$ (not irreflexive or transitive)
c. $\mathbf{x} \succ \mathrm{y} \succ \mathbf{x}$ (not asymmetric or irreflexive)

Second, scales cannot state that ∅ is less marked than another member of a scale.[11] (For now, I will use ∅ in an intuitive sense, to mean roughly "something unpronounced." A more precise definition will be given in §2.2.7.) Zero already satisfies all markedness constraints vacuously—including it in every (or any) markedness comparison introduces a perilous redundancy into the grammar. To formally exclude such redundancies, the following condition must hold of harmonic scales:

---

[10] Irreflexivity: $\forall x(\neg Rxx)$; transitivity: $\forall x \forall y \forall z((Rxy \,\&\, Ryz) \rightarrow Rxz)$; asymmetry: $\forall x \forall y(Rxy \rightarrow \neg Ryx)$(Partee et al. 1993). Asymmetry implies irreflexivity: if x is more marked than itself through transitivity, it is more marked than itself.

[11] I will restrict my attention to comparisons in the *unmarked* direction, though the question whether a comparison can imply that ∅ is more *marked* than something is an interesting one. Given my framework, a scale like $\varnothing \succ x$ can only give rise to a constraint *∅, which is a general "have structure" constraint. Constraints that demand the presence of specific structures are numerous, e.g., ONSET, FTBIN, PARSE-σ, or Grimshaw's (2003) OBHEAD and OBSPEC (see §2.3.4). Yet general constraints like *∅ may present a problem that is the opposite of Economy—Profusion. For my purposes, it is sufficient to require that ∅ be banned from the unmarked ends of a comparison, though it may be necessary to exclude ∅ from scales altogether. This does not exclude things like syntactic traces from scales—a trace can be defined as an empty projection that is contained in a projection together with some non-empty projections.

(22) NOZERO: No harmonic scale containing x implies that $\varnothing \succ x$.

Scales that disobey NOZERO include trivial binary comparisons ("∅ is better than a syllable"), zero-extended scales ("∅ is better than a voiceless obstruent, which is better than a voiced obstruent"), or the more bizarre zero-linked scales ("a trace is better than ∅, but ∅ is better than a non-empty projection").

(23) Illegitimate scales

a. $\varnothing \succ x$
b. $\varnothing \succ x \succ y$
c. $x \succ \varnothing \succ y$

NOZERO applies to both primitive and derived harmonic scales, though it applies to derived scales only vacuously: Harmonic Alignment is simply not set up to produce zero-extended scales. Recall from §2.2.2 that Harmonic Alignment applies to prominence scales, whose high end is occupied by a prominent segment such as a low vowel or a prominent position, e.g., the syllable peak. Zero cannot belong at the prominent end of a prominence scale, because *anything* is more prominent than ∅. As a result, ∅ can never be at the unmarked end of a harmony scale. As for primitive harmonic scales (such as the obstruent voicing scale) and the more formal scales (discussed in §2.2.6), these are prohibited from containing ∅ by (22).

The NOZERO principle might seem redundant if all scales can be stated in stringent terms. In a stringent scale, the unmarked is the superset of the marked. For example, in the stringent version of the vowel nasality scale, *vowel ≻ nasal vowel* ($V \succ V_{nas}$), the marked nasal vowels form a subset of all vowels (This way of looking at markedness is reminiscent of underspecification—see Archangeli 1984, 1988, McCarthy and Taub 1992, Pulleyblank 1988, Steriade 1995). Zero-extending the scale to $\varnothing \succ$ *vowel* $\succ$ *nasal vowel* violates the subset relationship, because $\varnothing$ is not a superset of *vowel*.

It is doubtful whether this approach can be extended to all scales, however. The problem is that once we move past the relatively simple featural markedness, stating scales in stringent terms becomes very difficult. For example, although nasal vowels are marked in general, they are not marked when adjacent to a nasal consonant. Conversely, oral vowels are in general unmarked, but they are marked when adjacent to a nasal consonant: $V_{nas}N$ is more harmonic than $V_{oral}N$. A non-stringent scale for this is straightforward: $V_{nas}N \succ V_{oral}N$. Stating this markedness relationship in stringent terms is a challenge—neither of the unmarked sequences is a superset of the marked. The same is true of many other markedness relationships—in the majority of cases, it is not possible to identify the marked structure by labeling it with a feature that the unmarked structure lacks. For this reason, the NOZERO principle is a necessary part of the theory.

Even though *scales* cannot state that $\varnothing$ is more harmonic than a non-null structure, a *ranking* can still select $\varnothing$ as the most harmonic candidate. This is a crucial aspect of the theory to which I will return in §2.2.7.2.

At this point, it is appropriate to consider a broader range of constraints and the harmonic scales on which they are based.

2.2.6 Relating markedness constraints to scales

The purpose of Lenient Constraint Alignment and the principles governing scales that were identified in §2.2.5 is to prevent constraints from penalizing all structure indiscriminately, as economy principles do. This theory of economy can only succeed if *all* constraints are derived from scales—otherwise there is no way to ban arbitrary antistructure constraints like *STRUC(σ) from CON.[12] This subsection identifies some issues in relating various kinds of markedness constraints to scales.

Scales are the real primitive in this theory—constraints are not. Ultimately, finding appropriate scales for previously proposed constraints is a problem for the analyst, not for the theory proposed here. For the purposes of this proposal, scales are required to express the relative ill-formedness of a particular form or structure and give a viable non-null alternative to it, but exactly how this is done is a separate matter. In this section, I discuss some possible formulations of scales for paradigmatic, syntagmatic,[13] and alignment constraints, though it should be kept in mind that there is no general "recipe" for scales.

Paradigmatic constraints are context-free constraints that ban segments with certain combinations of features—for example, *ŋ, *FRONTROUNDV, *VOICEDOBS, and *NASALV. Scales for such constraints are not hard to find: they reflect the relative markedness of some feature combination, e.g., "front rounded vowels are more marked than front unrounded and back rounded vowels."

---

[12] Alan Prince (p.c.) remarks that this is a necessary condition but not a sufficient one. Even if all constraints are lenient and derived from proper scales, it is also crucial that inputs be unrestricted. If inputs are restricted in any way, the theory will not achieve its results. For example, if the vowel inventory of a language is somehow artificially limited to {ɨ, y, ə}, the constraints against these vowels will act as economy constraints. For this reason alone, richness of the base must be a crucial assumption in the present theory. In chapter 4, I discuss cases where constraints against marked vowels interact with MAXV to produce economy effects, but these effects hold only over words that have such vowels—the rest of the language is unaffected precisely because inputs are unrestricted.

[13] The terms "syntagmatic" and "paradigmatic" in reference to constraint varieties are due to Pulleyblank 1997.

(24) *FRONTROUND “If a vowel is front, it is not round.”
Vowel rounding scale: {[+front, -round], [-front, +round]} $\succ$ [+front, +round]

Syntagmatic, or context-sensitive constraints, are based on more complex scales. The levels of these scales are occupied not by simple feature combinations but by sequences of segments and by structural configurations. For example, the scale for ONSET must state that consonant-initial syllables are superior to vowel-initial syllables:

(25) *Onset scale*: $[_{\sigma}$C... $\succ$ $[_{\sigma}$V...

Syntagmatic (context-sensitive) constraints don’t always refer to linear sequences of segments—many such constraints prohibit structural configurations. The scales for these constraints scales may be based on formal principles as opposed to the more phonetically oriented ones. For example, Cohn and McCarthy 1994/1998 derive the constraint[14] *(HL) from a scale based on the Grouping Harmony principle (Prince 1990). This scale shows a preference for a greater weight ratio between the second and the first syllable of a foot:

(26) GRPHARM, or *(HL)[15]
Grouping Harmony scale: (LH) $\succ$ (LL), (HH) $\succ$ (HL)

Again, just like the nasalization and onset scales, the Grouping Harmony scale orders structural configurations from most harmonic (LH) to least harmonic (HL). By Lenient

---

[14] H stands for “heavy syllable,” L stands for “light syllable,” and round brackets () are placed around feet throughout.

[15] Based on a ternary scale like Grouping Harmony, one would expect a constraint that bans HH and LL, as well. Cohn and McCarthy do not propose one. Of course, HH is ruled out by Prince’s (1990) WSP. LL violates SWP (see §2.3.2.3 and chapter 3).

Constraint Alignment, there is a constraint against (HL), but none against (LH).[16] The scale in (26) contains all the necessary information for formulating a constraint: it describes the most marked configuration, (HL), and offers some viable alternatives to it, i.e., (HH), (LL), and (LH).

In addition to paradigmatic and syntagmatic constraints of the sort already discussed, a third subtype of markedness constraints has been proposed: Alignment constraints (McCarthy and Prince 1993a, Prince and Smolensky 1993). These raise a formal issue of some importance to scales. Alignment constraints evaluate forms gradiently: for example, ALL-FT-L (a.k.a. ALIGN (Ft, L, Wd, L)) assigns a violation mark for every syllable that separates the left edge of a foot from the left edge of a prosodic word. This gives Alignment an economy flavor: the longer the word, the worse its violations will be. (The economy potential of Alignment is well-known; see §2.3 and especially §2.5.2.2).

Interestingly, there is no straightforward way to relate Alignment constraints to harmonic scales. The problem is that gradient constraints of this sort are able to make an infinitely large number of markedness distinctions, and therefore they require scales of infinite length. Yet scales of infinite length are an impossibility in Optimality Theory: CON is finite, so scales must be as well (see McCarthy (to appear) for some related discussion).

Thus, the least marked element on the scale for ALL-FT-L is *not* null—it is a foot that is perfectly aligned (in this, alignment constraints differ from *STRUC constraints; see

[16] Which is not to say that LH is a universally well-formed foot. LH may be banned in a trochaic system by a high-ranked WSP, but it will never be ill-formed in an iambic system.

§2.5.2.2). Yet the scale does not end by stating that a misaligned form is more marked than a perfectly aligned form—it goes on to state that a perfectly aligned form is more harmonic than one misaligned by one syllable, which is in tern less harmonic than a form misaligned by two syllables, which is less harmonic than a form misaligned by three syllables, and so on *ad infinitum*.

(27) Gradient ALL-FT-L: $[_{PrWd}\ (_{Ft}... \succ [\sigma\ (_{Ft}... \succ [\sigma\sigma\ (_{Ft}... \succ [\sigma\sigma\sigma\ (_{Ft}... \succ ...$

The infinite scale problem is a distinctly different matter than a constraint's ability to order candidates according to their magnitude of violation of a categorical constraint. For example, the ONSET scale states that a consonant-initial syllable is more harmonic than a vowel-initial syllable. In ordering *candidates*, ONSET will impose the ordering {*a* ≻ *a.a* ≻ *a.a.a* ≻ ...}, but as McCarthy (to appear) argues, the ability to keep track of multiple loci of violation is a necessary aspect of EVAL.[17] It is unnecessary and undesirable for *scales* to count loci of violation—it is sufficient that constraints do so.

It is possible to avoid the infinite scale problem by reformulating the scale in (27) in a more elegant form (see (28)). Note that this particular formulation distinctly resembles an economy principle, since size is a matter of comparison here:

(28) Gradient ALL-FT-L: $[_{PrWd}\sigma_n\ (_{Ft}... \succ [_{PrWd}\sigma_{n+1}\ (_{Ft}...$

The *n~n+1* aspect of this scale is a property that scales for categorical constraints lack, since those constraints are finite orderings. Nothing in the present theory rules out scales like (28), but there are other ways of excluding them from CON: gradient alignment

[17] As Prince and Smolensky (1993) repeatedly emphasize, EVAL does not really "count," rather, it compares the magnitude of violation of a constraint by different candidates.

constraints violate McCarthy's (to appear) definition of an OT constraint. Prohibiting gradience at scale level is not formally necessary to exclude it from the theory.[18]

The issue is actually more general: what about scales of the form $\sigma \succ \sigma\sigma \succ \sigma\sigma\sigma \succ \sigma\sigma\sigma\sigma \succ ...$ or $\sigma_n \succ \sigma_{n+1}$ (where $n \neq \varnothing$)? Scales of this form will give rise to constraints that do not necessarily prefer ∅ to any other candidate but are still intuitively economy constraints—they favor smaller structures over larger ones.[19] The problem here is that scales of this sort have no formal or substantive grounding. In addition to meeting the formal requirements on scales set forth in the present theory, scales need to express real linguistic tendencies; there is not evidence that the markedness of a form is proportional to the number of syllables in it. Another problem with "counting" scales is that languages—to put it simply—do not count. For all of these reasons, "counting" scales cannot be a part of the grammar.

To anticipate the upcoming discussion, it may now be apparent that economy constraints cannot be readily derived from any legitimate scales. The hallmark of a true economy constraint is its preference for ∅ above all other structures along a particular dimension of markedness; e.g., to *STRUC(σ), ∅ is better than a syllable, and to *VLESSOBS, ∅ is better than an obstruent. This point will be made precise in §2.4.4, where I will show that all *STRUC constraints share a common property in their relation to scales and are thereby prohibited from CON under the Leniency hypothesis.

---

[18] This problem with gradient constraints is not an issue in any of the case studies in this thesis—categorical constraints are used throughout. See §2.3.2.2 for an introduction to ENDRULE-L and ENDRULE-R, which take over some of the functions of ALLFT-L and ALLFT-R. The analyses in chapters 3 and 4 make extensive use of categorical constraints.

[19] Thanks to Andries Coetzee for bringing this to my attention.

In summary, this subsection examined some issues in relating different kinds of constraints to scales. For my purposes, scales simply state that some configuration is marked relative to at least one other. I applied this general approach to just a few context-sensitive and context-free constraints. In the chapters that follow, I provide scales for all the markedness constraints used in the analyses.

2.2.7 Null Outputs

2.2.7.1 <u>Defining the Null Output</u>

The notion of a Null Output,[20] or ∅, is of great importance to the proposal, since scales in CON are prohibited from implying its relative well-formedness. This section discusses the structural nature of ∅ and addresses its status in the theory.

Formally, the Null Output can be a number of things: a prosodic structure that is segmentally empty, an output in which every input segment has been deleted, or a segmentally empty output that bears no correspondence to the input at all. What I will do here is talk about how the present theory can be reconciled with the various proposals regarding the nature of the Null Output, though the theory need not be committed to any one of these proposals.

Under Prince and Smolensky's Containment model of input-output mappings, material can never literally removed from the output, but it can be prosodically underparsed. Thus a candidate in which every segment is deleted is formally the same as an unprosodified segmental string. Under Containment, there is only one type of Null

[20] The Null Output is often discussed in the context of absolute ill-formedness. For discussion and applications of the Null Output, see Bakovic and Keer 2001, Benua 1997, Cohn and McCarthy 1994/1998, Kager 2000, Legendre et al. 1998, McCarthy to appear, Orgun and Sprouse 1999, Raffelsiefen 1996. See also the review in McCarthy 2002b:230.

Output—a partially or fully unprosodified candidate, which is “uniquely unsuited to life in the outside world” (Prince and Smolensky 1993:51). To be “partially unprosodified” means to lack an entire layer of prosodic structure. Thus, an output that has at least some of each of morae, syllables, feet, and prosodic word structure is fully prosodified in their sense, even if it has some extraprosodic material. This Null Output does not have any faithfulness violations, but it has egregious violations of constraints of the PARSE family (PARSESEG, PARSE-σ, and so on), which require elements to belong to proper levels of the Prosodic Hierarchy. Every segment of such an output is literally extrametrical.

Under Correspondence Theory (McCarthy and Prince 1995), more than one kind of output can be null because there is more than one way for a candidate to be unfaithful. There are two kinds of Null Output: ⊙, whose correspondence relation to the input is undefined (McCarthy to appear), and *e*, where every input segment has been deleted (Benua 1997). These two kinds of Null Outputs differ in their faithfulness violations: ⊙ violates Prince and Smolensky’s M-PARSE (which militates against non-realization of morphemes), *e* violates IO-MAX (which militates against the deletion of individual segments):

(29) A Null Output is any candidate that

a. violates M-PARSE (McCarthy to appear),
b. contains no correspondence relations that satisfy IO-MAX (Benua 1997),
c. lacks one or more PH levels (Prince and Smolensky 1993).

Despite formal differences, all of these Null Outputs share a common trait: they lack phonetic realization. The theory may not be so rich as to permit all of these versions of the Null Output, but no scale can imply that a structure without a phonetic realization (regardless of its formal nature) is more harmonic than a non-null structure.

2.2.7.2 <u>The status of Null Outputs in the theory</u>

Although the Null Output cannot be more harmonic than a non-null structure on a harmonic scale, the Null Output *can* be less marked than another candidate with respect to a markedness constraint. This is crucial to the theory of economy effects developed here: no individual constraint prefers a Null Output to every other candidate, but a ranking can. This is because markedness constraints do not include any instructions on how to fix the markedness problem, as in: "replace a nasal vowel with an oral one." The grammar is free to select any alternative to a nasal vowel—a nasal consonant, an oral vowel, ∅, or any other form that is selected by other markedness and faithfulness constraints in the ranking.

This is schematically shown in (30). Given these constraints, any one of the candidates {x, y, ∅} is a possible winner in some language. If all the constraints in (30) dominate MAX, candidate (c) will be selected as the winner.

(30) The set of possible winners

/x/	*X	MAX	IDENT
a. x	*		
b. y			*
c. ∅		*	

This is actually a point of difference between the theory presented here and Targeted Constraint Theory (Wilson 2000, 2001, see also McCarthy 2002a). In Targeted Constraint Theory, constraints are also based on comparisons between forms, but there is a significant difference. Targeted constraints are not capable of comparing two candidates unless they are explicitly set up to compare them. For example, a constraint "Y≻X" will impose the harmonic ordering {y ≻x} on the candidates in (30), but they cannot assess the harmony of x relative to ∅ or of y relative to ∅. Moreover, Targeted Constraint Theory does not necessarily rule out constraints of the form "∅ ≻ X." In the Lenient theory, every constraint is capable of evaluating every candidate: even though ∅ is not on the scale that *X in (30) is based on, *X is still able to compare ∅ to x or to y. The reader is referred to Wilson 2000, 2001 and to McCarthy 2002a for further discussion.

To sum up, although individual markedness constraints are not set up to favor ∅ above all other candidates, the grammar can do so under a particular ranking This is a crucial ingredient for economy effects—we want deletion to be an option in at least some cases.

2.2.8 Section summary

In this section, I outlined a proposal for the structure of the constraint set CON. According to this proposal, all markedness constraints must be based on scalar comparisons between marked structures and non-null unmarked structures. This approach offers a new way to look at markedness: to say that x is marked is to say that there is a non-null y that is less marked than x. One of the mechanisms of the theory is a lenient reformulation of Prince and Smolensky's Constraint Alignment, whereby the least marked element on every markedness scale is not mapped to a constraint but other levels are.

This modification of CON has a significant consequence: no constraints can penalize structure for the sake of penalizing structure. Any dispreference for structure, also known as economy, must follow from the interaction of constraints in language-specific grammars. The next section explores this in more detail by demonstrating how several economy effects are derived in the theory.

2.3 <u>Economy effects through constraint interaction</u>

2.3.1 Introduction

While economy principles and constraints do not exist, economy effects do. Broadly speaking, there are two kinds of structural economy effects. The first might be called *limited structure building*—the number of structural nodes in a given input is minimized. For example, instead of giving each of two syntactic phrases its own phonological phrase, the two syntactic phrases are lumped into a single phonological phrase whenever possible (see Selkirk 1995a, Truckenbrodt 1999 and others). The second is a more aggressive effect that results in actual deletion of input elements, such as truncation, syncope, and other processes that visibly make the output smaller.

I argue that the dispreference for structure can always be reduced to the interaction of other factors—there is never an overarching economy principle at work. As long as deletion is an available option in the grammar, some markedness constraints will be satisfied by deletion at least some of the time. Crucially, though, deletion is never the *only* option for satisfying a particular markedness constraint—it may be so in a given grammar, but there will be other grammars that achieve the same markedness goal in another way.

Recent work in OT has been rather successful in explaining many economy effects in terms of independently motivated constraints. In the remainder of this section, I will review some of the existing work on the subject and discuss a few new possibilities for analyzing economy effects.

2.3.2 Limited structure building

2.3.2.1 <u>One big structure is better than two smaller ones</u>

First, let's look at the preference for *fewer* structures. Consider the aforementioned preference for "lumping" several syntactic phrases into a single phonological phrase. Truckenbrodt 1999 proposes that this lumping is the effect of a constraint WRAP-XP, which requires each XP to be contained inside a phonological phrase. This constraint conflicts with ALIGN(XP, PhP). When several smaller XPs are contained in a larger XP, WRAP-XP penalizes all outputs that place smaller XPs into their own phonological phrases without "wrapping" the larger XP into one, but alignment constraints ban XP edges that do not coincide with phonological phrase edges:

(31) WRAP and ALIGN, after Truckenbrodt (1999)

	WRAP-XP	ALIGN (XP, PhP)
$(_{PhP}[_{XP1}\ [_{XP2}\ \ ]\ [_{XP3}\ \ ]])$	✓	*(XP3)
$[_{XP1}(_{PhP}[_{XP2}\ \ ])(_{PhP}[_{XP3}\ \ ])]$	*(XP1)	✓
$(_{PhP}[_{XP}\ \ ])$	✓	✓

Intuitively, neither of the constraints in (31) is an economy constraint: they do not *count* phonological phrases, since only the correspondences between edges matter. These are also not economy constraints from the formal point of view, since they can be related to scales that compare two non-null structures: a well-phrased one and a poorly phrased one. Yet if WRAP-XP dominates ALIGN, the effect will be a preference for fewer but larger phonological phrases—i.e., a structural economy effect in the sense of Chomsky 1991, 1995 and Rizzi 1997 but without economy principles or constraints.[21]

2.3.2.2 The "one foot per word" effect: one structure is better than many

Another class of limited structure building effects involves situations where only one constituent is built even though more than one is possible, but the size of the constituent is constant. An example of such an effect is non-iterative foot parsing.

First, a little background. In the theory of foot parsing of McCarthy and Prince 1993a, b, whether a language has iterative footing or non-iterative footing depends on the relative ranking of gradient alignment constraints and PARSE-σ. PARSE-σ demands that every syllable belong to a foot, while ALL-FT-L and ALL-FT-R require that every foot in a word stand at an edge, assigning violation marks for every syllable that stands between the edge of a foot and the edge of a prosodic word. Economy of footing, or the "one foot per word" effect, is obtained when either ALL-FT-L or ALL-FT-R dominates PARSE-σ; the relative ranking of the alignment constraints determines whether the single foot is at the left or the right edge.

(32) The "one foot per word" effect in gradient alignment theory

	ALL-FT-L	ALL-FT-R	PARSE-σ
a. (σσ)σσ		**	**
b. σσ(σσ)	**		**
c. (σσ)(σσ)	**	**	

[21] Paradoxically, Truckenbrodt still employs a *STRUC constraint in his system, *P-PHRASE, though it is never crucially active—it never makes any distinctions that other constraints do not make.

Kager 2001 argues that this constraint set overgenerates, imposing a symmetry on the typology of iambic systems that is not matched by the observed data (see also McCarthy (to appear) for other arguments against gradience in OT). An alternative to gradient alignment for deriving the "one foot per word" effect are the categorical ENDRULE constraints (McCarthy to appear), which are OT adaptations of Prince's (1983) proposal. The definitions of these constraints and their harmonic scales are given below.

(33) ENDRULE-L: "The head foot is not preceded by another foot within the prosodic word" (McCarthy to appear).
Harmonic scale: [$_{PrWd}$ x (HdFt)...] ≻ [$_{PrWd}$ ....(Ft)... (HdFt)...] x not a foot

(34) ENDRULE-R: "The head foot is not followed by another foot within the prosodic word" (McCarthy to appear).
Harmonic scale: [...(HdFt) x $_{PrWd}$] ≻ [...(HdFt) (Ft) $_{PrWd}$]

ENDRULE constraints interact with PARSE-σ as shown in (35). A word with just one foot and no unfooted syllables satisfies both of the ENDRULE constraints and PARSE-σ: the main stress foot is not preceded or followed by another foot in the word. A word with a single foot that contains some unfooted syllables still satisfies both of the ENDRULE constraints, but it incurs some violations of PARSE-σ—the longer the word, the more violations. Exhaustively footed words with more than one foot will violate either ENDRULE-L or ENDRULE-R, depending on the position of the main stress foot.

(35) ENDRULE constraints and the "one foot per word" effect

	ENDRULE-L	ENDRULE-R	PARSE-σ
(σσ)			
(σσ)σ			*
(σσ)σσ			**
σσ(σσ)			**
(σσ)σσσ			***
(σσ)σσσσ			****
(σ́σ)(σσ)		*	
(σσ)(σ́σ)	*		
(σ́σ)(σσ)(σσ)		*	
(σσ)(σσ)(σ́σ)	*		

Collectively, these constraints distinguish between words with one foot and words with more than one foot, but feet are not counted beyond that. The only counting is done by PARSE-σ, which assigns violation marks for every additional instance of an unfooted syllable. This constraint set turns out to make all the necessary distinctions: in chapter 3 we will see languages where the number of *unfooted* syllables is minimized, but the number of *footed* syllables is never minimized except as a function of the foot's well-formedness (more on this in the next subsection.) No constraint forbids feet per se.

2.3.2.3 A smaller structure is better than a bigger one

This particular class of effects is in a way the opposite of the kind discussed in §2.3.2.1, which reveals a certain lack of real unity to economy effects—a problem for overly economy principles like Rizzi's (1997) "Avoid Structure." At a certain level of analysis, the preference for smaller structures over larger ones is really just a variation on the "one is better than many" effect. This is true of the preference for monosyllabic, heavy trochees (H) over trochees that consist of two light syllables (LL), which is instrumental in the case study of Tonkawa in chapter 3.

In the metrical theories of Prince 1990 and Hayes 1995, H and LL trochees are treated equivalently: they are both binary at the moraic level and they are both even (in terms of weight). For Prince 1990, this is the cumulative effect of FTBIN and GRPHARM, since both feet are equally unmarked with respect to these constraints. Yet it is not the case that *no* constraint distinguishes between H and LL trochees—the STRESS-TO-WEIGHT PRINCIPLE does. H satisfies the requirement for foot heads to be heavy, yet it is not the only foot to do so—as shown in (36), HL feet do as well. Only H satisfies both SWP and GRPHARM:

(36) Syllable economy in trochees through constraint interaction

/pata/	SWP	GRPHARM
a. (pát.ta) HL		*
b. (páa.ta) HL		*
c. (pát) H		
d. (pát)ta H		
e. (pá.ta) LL	*	

Neither of the constraints in (36) prefers smaller structures to larger ones or counts syllables, yet collectively they converge on H as the best foot. The fact that it is monosyllabic is not a virtue by itself—rather, its weight distribution is its best attribute. Note also that among iambs, there is no preference for H over LH—unevenness is praised in iambs, and both H and LH satisfy the requirement for foot heads to be heavy. In §2.5.2.1, I will argue that their harmonic equality is supported by typological evidence.

To summarize, limited structure building effects result from the interaction of regular markedness constraints—no economy principles are necessary to derive them. In the next section, I turn to the more aggressive economy effects—ones that actually involve deletion of input material.

2.3.3 Deletion of input structure

Deletion is one of the most striking economy effects—it visibly makes the output shorter. Early on, Zipf (1949) observed that frequently used words and names undergo truncation (e.g., *popular* → *pop*), which he attributed to a general Principle of Least Effort that, he argued, governs many aspects of human behavior.[22] Since then, several linguists have shown that deletion (including truncation) is governed by the same constraints that are instrumental in non-economy processes. In this subsection, I show how a number of size maximum restrictions can be derived by appealing to regular markedness constraints for which there is independent motivation outside of economy processes.

2.3.3.1 Foot-sized maxima derived

A major player in truncation is the metrical foot. Ito 1990 demonstrates that truncated forms of English loanwords in Japanese must be large enough to fit a disyllabic trochaic foot template (e.g., *herikoputaa*→ *he.ri* 'helicopter,' not **he*). The same is true of hypocoristics and other forms of truncation, where the foot restricts minimal size (Bethin 2002, Crowhurst 1992, McCarthy and Prince 1986, 1990, Weeda 1992, Woodbury 1985). If economy is really all that matters, then why not go with the shortest pronounceable word, e.g., one that is just a single light syllable? Clearly, crucial here is not size per se but prosodic well-formedness.

Particularly telling are cases where the foot is not only the size minimum but also the size maximum. Consider truncation in the speech of child learners of English (Pater

[22] For a nice overview of Zipf's Law (a.k.a. the Zipf-Mandelbrot-Pareto Law) and critique of Zipf's work, see Rapoport 1982.

and Paradis 1996, Pater 1997). Adult words of three syllables or longer are clipped to two syllables, but some disyllabic words (e.g., *giraffe*) are also truncated:

(37) Truncation in child speech (Pater 1997)

a. wǽ:dit 'rabbit'
b. té:do 'potato'
c. wæ:f 'giraffe'
d. ga:bɛdʒ 'garbage'

Pater 1997 observes that truncated words in child speech are not conforming to a disyllabic template—rather, the output of truncation is invariably a trochaic left-aligned foot. This explains why disyllabic words like *giráffe* undergo truncation—the adult form contains an unfooted syllable at the left edge, which is marked. Pater's analysis is an extension of McCarthy and Prince's (1994a) analysis of Diyari foot-sized reduplicants (discussed shortly). Pater argues that the foot-sized size maximum emerges from the interaction of ALL-FT-L, PARSE-σ, and MAX (see (38)). Disyllabic words that already have trochaic stress, e.g., 'rabbit,' do not undergo truncation. Disyllabic words that are stressed on the last syllable must be shortened so they are exhaustively parsed:

(38) Truncation without economy constraints in child speech

		ALL-FT-L	PARSE-σ	MAX
'rabbit'	a. ☞(wǽ:dit)			
	b. (wæb)			**!
'giraffe'	c. ☞(wæf)			**
	d. gi(wæf)	**!	*	
'hippopotamus'	e. ☞(pómus)			***!***
	f. (hìppo)(pómus)	**!		**
	g. (hìppo)(póta)mus	**	*	

This is really a variation on the "one foot per word" effect discussed in §2.3.2.2, except here it is coupled with a "no unfooted syllables" restriction. (Note also that the same effect can be obtained if ALL-FT-L is replaced with the non-gradient ENDRULE-L constraint in this tableau.)

Truncation in child speech is not shortening for the sake of making words shorter—clearly, it matters whether the adult word violates certain constraints. Shortening 'rabbit' to something like *wæb* would produce a more economical output that is also a trochaic, binary foot—witness 'giraffe' → *wæf*. The reason shortening does not apply here is that no metrical markedness constraint calls for it. An economy trigger/markedness blocker explanation (e.g., FTBIN>>*STRUC(σ)>>MAX) would incorrectly predict that all words should be clipped down to a CVC or CVV binary trochaic foot.

Pater argues that, although metrical markedness constraints are ranked below MAX and can be violated in adult English, they still have visible effects. The interaction of ALL-FT-L and PARSE-σ produces the so-called initial dactyl effect (McCarthy and Prince 1993a): when a trisyllabic sequence precedes the main stress, secondary stress usually appears on the initial syllable, e.g. *(Tàta)ma(góu)chi* not *Ta(tàma)(góu)chi*. ALL-FT-L enforces the requirement for the first syllable to be footed in adult English and in child English alike.[23]

Just like words in child speech, reduplicative morphemes in many adult languages are limited to a foot-sized unit (McCarthy and Prince 1986, 1993b). A famous example of this is reduplicant disyllabicity in Diyari (McCarthy and Prince 1994a). Although non-

[23] In a theory with only categorical constraints, the effect has to be attributed to a different constraint. McCarthy (to appear) suggests PARSE-σ1, which requires the first syllable of the word to be footed—a kind of positional markedness constraint.

reduplicated forms can be longer than two syllables, the reduplicant is limited to the size of a trochaic foot:

(39) Diyari reduplicant disyllabicity (McCarthy and Prince 1994a)

a.	/RED-wila/	wila-wila	'woman'
b.	/RED-ŋankanti/	ŋanka-ŋankanti	'catifish'
c.	/RED-tʲilparku/	tʲilpa-tʲilparku	'bird species'

McCarthy and Prince argue that the reduplicant is not just squeezed into a disyllabic template—rather, it has all the properties of the prosodic word in the language, including separate stress and no word-final codas (Austin 1981). The difference between the marked base and the unmarked reduplicant is that the reduplicant must be an exhaustively footed monopod, whereas the base does not have to be either. Again, well-formedness is important here, not shortness.

An interesting variation on the size maximum restriction holds of prosodic words in Maori, which de Lacy 2002b also analyzes in terms of metrical well-formedness constraints. The twist is that Maori words can contain unfooted syllables, but they cannot be *footable*—trisyllabic words are acceptable but quadrisyllabic words are not.

Truncation in Maori is often used to clip words down to the maximally trisyllabic size, but sometimes truncation does not reduce the size enough—there are still footable syllables in the word. De Lacy argues that in these cases, epenthesis applies, so that part of the word can form a separate prosodic word. Thus, in the first word in (40), *hikáia*, the suffix *–ia* is mapped faithfully because the word fits into the single-foot limit, but in *kopóua*, the suffix loses its first vowel. In longer words, though, deleting the single vowel does not produce the necessary improvement; in words with three moras and longer, the suffix heads its own prosodic word (square brackets indicate prosodic word boundaries, periods indicate syllable boundaries):

(40) Maori maximal words: truncation and augmentation (de Lacy 2002b)

a. /hika-ia/	→ [hi.(kái).a]	‘plant passive’	
b. /kopou-ia/	→ [ko.(póu).a]	‘appoint passive’	
c. /tapuhi-ia/	→ [(tá.pu).hi] [(tí.a)]	‘sort out’	not *(tápu)hia

De Lacy’s gradient constraint analysis can be easily recast in terms of ENDRULE constraints, since ENDRULE constraints subsume the functions of his constraint *FT- “no non-head feet.” This analysis is sketched out in (41).

(41) Maori maximal words

		ENDRULE(L/R)	*LAPSE$_{FT}$	DEP-C	MAX
/karaŋata/	a. ☞ (kára)ŋa				*
	b. (kára)ŋata		*!		
	c. (kára)(ŋáta)	*!			
/kopou-ia/	d. ☞ko(póu)a				*
	e. ko(póu)ia		*!		
	f. ko(póu)(ía)	*!			
	g. [ko(póu)][(tía)]			*!	
/tapuhi-ia/	j. [(tápu)hi] [(tía)]			*	
	k.[(tápu)hia]		*!		*
	l. [(tápu)(hía)]	*!			*

ENDRULE dominates MAX together with *LAPSE$_{FT}$ “adjacent unstressed moras must be separated by a foot boundary” (which de Lacy adopts from Green and Kenstowicz 1995, Prince 1983, Selkirk 1984b). Words that are just the right size (e.g., *hi(kái)a*) will not truncate, since they can be served with just one foot without any lapses. A hypothetical input like /karaŋata/ will have to be truncated because the only alternatives are lapses and iterative feet, as will *ko(póu)a*. Inputs like /tapuhi-ia/ are simply too long for deletion to make any difference—witness the failure of *(tápu)hia* to satisfy *LAPSE$_{FT}$. The only solution is to parse this word as two prosodic words, which requires the epenthesis of *t* in Maori. (The reader is referred to de Lacy's paper for a complete analysis of this complex pattern.)

Being shorter is not a goal in itself here—the well-formedness conditions that hold of the Maori prosodic word are just as possible to satisfy by insertion as by deletion. Deletion just happens to be preferred because MAX is ranked below DEP.

In general, the "one foot per word" effect results from the interaction of metrical constraints with MAX. These constraints are not economy constraints—none of them prefer smaller structures to larger ones. The preference emerges from their interaction in language-specific rankings.

2.3.3.2 The syllable-sized limit on reduplicants: OO-correspondence

The prosodic explanation of foot template effects is now uncontroversial, but maximal size can be limited to a unit that is even smaller than the foot. Thus, reduplicants in many languages seem to copy as little as possible of the base (e.g., a syllable or even just one segment), which several researchers have attributed to economy constraints (Feng 2003, Riggle 2003, Spaelti 1997, Walker 1998, 2000, 2003). Interestingly, this size restriction is not widely attested outside of reduplication,[24] which makes it doubtful that general economy constraints are the answer. I propose that the size restrictor in these

---

[24] Walker 2003 argues that in Yuhup, *all* morphemes are limited in size to a single syllable: there is a requirement that morphemes and syllables correspond one to one. However, in the data Walker cites from Lopes and Parker 1999, every syllable also happens to be either CVV or CVC, which suggests that the real generalization concerns *feet*, not syllables. The fact that the foot is monosyllabic falls out under a trochaic analysis, assuming that SWP is high ranked. Walker notes that stressed syllables lengthen in Yuhup, which suggests that this analysis is on the right track.

cases is not economy but rather Output-Output faithfulness (Benua 1997, Burzio 1994, Kenstowicz 1996a).[25]

The reduplicated form stands in transderivational correspondence with the non-reduplicated form, which serves as the base in the OO-corespondence relationship. OO-DEP (Benua 1997) requires that every segment in the reduplicated form have a correspondent in the base, which effectively puts a limit on how much can be copied—a violation is incurred for every segment of the reduplicant. The reason anything is realized at all is MORPHREAL, which requires every morpheme to have a phonological exponent. In most cases, then, reduplicants (underlined in (42)) will copy just enough to give the reduplicant some realization, but not more:

(42) OO-correspondence and minimal copying in reduplication

base: pa.ta input:/RED-pata/	MORPHREAL	OO-DEP
a. ☞ <u>pa</u>-pa.ta		p
b. pa.ta	*!	
c. <u>pa.ta</u>-pa.ta		pata!

Spaelti (1997) discusses several such cases. For example, in the Rebi dialect of West Tarangan, a single consonant is copied wherever possible, while a syllable is added only where necessary. The reduplicant always immediately precedes the stressed syllable. As the patterns below show, the reduplicant copies a single consonant if it can serve as a coda to the pretonic syllable, as in *bimtɛ́mana* and *tar̲púran*. Single segment reduplication

---

[25] Alber 2001 suggests that another pressure can act as a size restrictor for reduplicants: the requirement that every segment of the output be in the root-initial syllable (cf. Beckman 1998 on MAX-POSITION constraints). The full implications of this remain to be seen.

is blocked if the preceding syllable is closed and a single consonant cannot be appended to it, in which case the entire firsts CVC of the base is copied, as in *paylawláwana* (not **payw.láwana*). Single segment reduplication is also blocked by the constraint against geminates, so *nánay* reduplicates as *nanánay* not **nan.nánay*:

(43) Rebi West Tarangan reduplication (Spaelti 1997)

a.	/RED-bitɛma-na/	bimtɛ́mana	'small 3s.'	cf. bitɛ́mana
b.	/RED-tapuran/	tarpúran	'middle'	cf. tapúran
c.	/RED-paylawa-na/	paylawláwana	'friendly 3s.'	cf. payláwana
d.	/RED-nanay/	nanánay	'hot'	*nan.nánay

The reduplicated and the non-reduplicated forms look quite similar in the default pattern—cf. *tarpúran* and *tapúran*. This similarity is achieved by copying as little as possible, i.e., just a single segment, while still realizing the reduplicative morpheme:

(44) Minimal copying in Rebi West Tarangan: just one segment

base: ta.pú.ran input: /RED-tapuran/	MORPHREAL	OO-DEP
a. ☞ tar.pú.ran		r
b. ta.pú.ran	*!	
c. ta.pú.ran-ta.pú.ran		tapuran!
d. ta.pur.pú.ran		pur!

In words that begin in a CVC syllable, infixation of a single consonant is ruled out by *COMPLEX, which overrides the effects of OO-DEP:

(45) Minimal copying in Rebi West Tarangan: just one syllable

base: pay.lá.wa.na input: /RED-paylawa-na/	*COMPLEX	MORPHREAL	OO-DEP
a. ☞ paylawláwana			law
b. paywláwana	*!		w

Walker 2000 discusses a similar pattern for Mbe, where she argues a nasal coda is the only exponent of the reduplicated part of a complex morpheme—in Mbe, if the reduplicant would have to copy an entire syllable, copying is blocked altogether. Here, the relevant constraints on codas actually dominate MORPHREAL. The common thread to these and other similar patterns is that reduplicants seem to be under a restriction against increasing the size of the word, but only with respect to another word *in the same derivational paradigm*. This is not syllable economy—it's paradigm uniformity.

This account of minimal copying predicts that such size restrictions will hold only of affixes (including reduplicative affixes) but not of stems. OO-DEP cannot have a size limiting effect on stems, since these add nothing new to the base. Only affixes do, so only they are limited in size to units smaller than a foot. This analysis also eliminates the need for the oft-criticized templatic constraint AFFIX≤σ (McCarthy and Prince 1994b). See McCarthy and Prince 1999 for some discussion.

An alternative explanation for minimal copying is in terms of *STRUC(σ) or syllable alignment (Feng 2003, Riggle 2003, Spaelti 1997, Walker 2000, 2003). These constraints apply not only to affixes but also to stems, so in principle it is possible for them to limit the size of every morpheme to a single segment or a single light syllable—both effects are unattested. Since the OO-DEP analysis is sufficient and makes just the right predictions, I suggest that the economy constraint analysis of minimal copying be abandoned, especially since economy constraints are not needed for any other reason.

2.3.3.3 Haplology and the OCP

A group of deletion processes that might be called economy effects involve adjacent identical segments (OCP effects) or sequences (haplology). These are not economy effects in the most obvious sense of the word, but they do result in shorter outputs, and they have been analyzed in terms of economy constraints.

In (46), two kinds of deletion are shown: in the first case, dubbed anti-antigemination by Odden 1988, vowels delete between identical consonants, which appear as a geminate on the surface. Deletion does not apply between different segments.[26] In the second case, Basque, deletion targets one of two adjacent obstruents that are both continuant (or both non-continuant). In some cases, the entire consonant does not delete but instead deaffricates (Fukazawa 1999 argues that this is still deletion of *features*). Deletion does not apply otherwise, as in the last two examples.

(46) OCP deletion

a. Syncope between identical segments and gemination in Mussau (Blust 2001)

/pa̲pasa/	ppasa	‘outrigger poles’	
/gagaga/	gagga	‘tidal wave’	
	biliki	‘skin’	*bilki, *bliki
	karasa	‘whet, grind a blade’	*karsa, *krasa

b. Consonant deletion and de-affrication in Basque (Hualde 1991)

/bat̲ paratu/	baparatu	‘put one’	
/irabas-ts̲en/	irabasten	‘earn, win’	
/hit̲s-t̲egi/	histegi	‘dictionary’	
/itʃ-tsen/	iʃten	‘open’	
	ibiltsen	‘walk’	*ibilten
	esne	‘milk’	*ene

Morphological haplology can be defined as the non-realization of a morpheme when it is attached to a stem that contains an adjacent identical sequence of phonemes, as with the French suffix *–iste* [ist]. When the suffix attaches to a base that ends in a sequence that is partially or fully homophonous with *–iste*, part or all of the suffix is not realized:

[26] Here, the notion of adjacency has to be stretched to include consonants separated by a vowel—see McCarthy 1986 and Rose 2000b.

(47) French Haplology (de Lacy 1999, (a) and (b) from Corbin and Plénat 1992)

a.	/deiksis-ist/	deiksist	‘deixis + *ist*’	*deiks<u>isis</u>t
b.	/ametist-ist/	ametist	‘amethyst +*ist*’	*amet<u>istist</u>
c.	/ego-ist/	egoist	‘egoist’	

These processes should be discussed in the context of a rather general constraint against identity, the OCP (Fukazawa 1999, Goldsmith 1990, Keer 1999, Leben 1973, McCarthy 1986, Myers 1997, Odden 1988, Rose 2000b, Suzuki 1998, Yip 1988, 1998; see also chapter 4). A rather striking thing about the OCP is just how many ways there are to satisfy it: dissimilation, allomorphy, lexical gaps, consonant deletion, syncope, and suppletion are all observed effects. It appears, then, that there is nothing at all special about deletion being part of this set—the interaction of the OCP with MAX straightforwardly predicts it.

Despite this range of effects, some have argued that structure-reducing operations of the sort illustrated above are in some way special and indicate that all structure is marked. Thus, de Lacy (1999) argues that morphological haplology is economy-driven coalescence. He observes that haplology does not always target morphemes with marked features, as in the case of Arabic /ta + ta + kassaru/ → *takassaru* ‘it (fem.sg.) breaks,’ **tatakassaru* (Wright 1971). Assuming that there is a markedness constraint against everything, even the apparently unmarked *ta*, haplology can be analyzed using Economy constraints of the *STRUC family and without resorting to constraints against adjacent identical sequences. De Lacy presents several arguments against an OCP analysis of haplology, but the OCP analysis has a strong virtue that *STRUC lacks: only the OCP can be satisfied by dissimilation, allomorphy, and other processes that do not involve deletion or coalescence.

OCP-driven deletion of single segments has similarly been analyzed in terms of economy principles. Because the OCP can target a sequence of any identical features and not just marked ones, Fukazawa 1999 analyzes it as the Local Conjunction of Economy constraints.[27] As the following quotation shows, this analysis also relies on the assumption that the best structure is no structure:

(48) All the features are marked in a sense; therefore, the constraints which prohibit them exist in the grammar... Thus for example, although the [cor] feature is relatively unmarked compared to the [dor] or [lab] feature, it is still marked, and the constraint against the [cor] feature does exist, namely, *[cor]. The OCP effects on this relatively unmarked feature [cor] can be accounted for based on the self-conjoined markedness constraint, namely, *[cor][cor]. In this respect, there are no OCP effects which the self-conjunction approach cannot explain. (Fukazawa 1999:19)

I assume that what is marked here is repetition and identity of features, not their mere occurrence (cf. Yip 1998). Any features can be targeted for deletion because any features can be repeated.[28]

Economy principles can be used in this fashion to explain vowel harmony, tone spreading, assimilation, Verner's Law, and any other process that replaces a series of feature nodes with one shared feature. To my knowledge, not all of these avenues have been pursued, and for a good reason: these are processes that can just as well be explained as regular markedness effects. All economy effects can and should be analyzed in terms of markedness constraints.

---

[27] Local Conjunction combines the power of two constraints to create a third constraint that is active in a specific domain (Smolensky 1995). For example, the conjunction of ONSET and NOCODA in the domain of a syllable, $[\text{ONS\&NOCODA}]_\sigma$, is a constraint that is violated by a syllable that simultaneously has a coda and lacks an onset, but not by a syllable that violates only one of the two conjoined constraints.

[28] Alderete 1997 argues that there is an implicational universal here—if unmarked features are targeted, then marked ones must be as well; see his analysis for more details

2.3.4 Economy of structure in Grimshaw's theory: a comparison

The approach to economy pursued here is inspired by Grimshaw 2003, who also argues that structural economy results from the interaction of independently motivated constraints rather than special economy principles. However, there is an important difference in the way the constraints in the two theories treat ∅.

Grimshaw 2003 shows for syntactic phrase structure that economy effects follow from Alignment and constraints that require syntactic positions to be filled—constraints needed for independent reasons. Although *individually* these constraints may prefer larger structures to small ones, *collectively* they prefer smaller structures. The more projections a form contains, the more violations of alignment it incurs in Grimshaw's system (alignment is reckoned gradiently, with one violation mark assigned for every projection that separates an element from the nearest phrase edge):

(49) Grimshaw's phrase structure economy

	HEAD-LEFT	SPEC-LEFT	COMP-LEFT
a. [Head]			
b. [Spec H Comp]	*		**
c. [[Spec H Comp] H Comp]	**		****
d. [[[Spec H Comp] H Comp] H Comp]	***		******

Candidate (a) in (49) is as small as possible for a non-null structure and is perfectly aligned because it contains only one element. Any more internal complexity results in additional violation marks (b)-(d). This is an economy result—more structure means more markedness, yet no special economy constraints are used.

A preference for smaller structures need not entail a preference for empty structures. An interesting result of Grimshaw's system is shown in (50): a null projection (a) is harmonically bounded by candidates like (b) and (c), which are just as well aligned and satisfy at least one constraint that requires positions to be non-empty (OB-HEAD stands for "obligatory head," OB-SPEC stands for "obligatory specifier"):

(50) Empty structure disfavored

	OB-HEAD	OB-SPEC	HEAD-LEFT, SPEC-LEFT
a. [__]	*	*	✓
b. [H]		*	✓
c. [Spec]	*		✓

No special constraints that prefer smaller structures are required in this system because "economy of phrase structure is a theorem of the theory of phrase structure" (Grimshaw 2003:81).

Note, however, that the constraint set in (50) can actually favor wholesale deletion of input material, because the deletion candidate ∅ satisfies all of the constraints better than any other candidate. The null candidate is structurally distinct from the empty structure [___]—it contains no projections, so it cannot violate OB-SPEC or OB-HEAD.

(51) Grimshaw's constraints can favor wholesale deletion

	OB-HEAD	OB-SPEC	HEAD-LEFT, SPEC-LEFT
a. [__]	*	*	
b. [H]		*	
c. [Spec]	*		
d. ∅			

Grimshaw assumes that deletion is not allowed in syntax and that underlying forms do not contain function words (see also Grimshaw and Samek-Lodovici 1995, Grimshaw 1997, Grimshaw and Samek-Lodovici 1998). Whether or not deletion (rather than underparsing) is actually allowed in syntax is not a settled issue. It might be argued that GEN is not allowed to alter the semantic content of the input (Ackema and Neeleman 1998, though see Bakovic and Keer 2001, Legendre et al. 1998 for alternative views).

However, deletion is necessary if inputs are unrestricted (Prince and Smolensky 1993): if an input contains too many pleonastics, for example, as in **Mary did buy the book* (with unstressed *did*), they must be deleted and inflection must be inserted so that a grammatical output is obtained. If deletion is not an option in syntax, then ∅ is not a problem for this theory of economy effects.

2.3.5 Section summary

In this section, I argued that a variety of economy effects follow from the interaction of constraints rather than from special economy principles. While individually these constraints do not prefer smaller structures to larger ones, collectively they may favor economical structure building and actual deletion of input material. Deletion is always just one of several solutions, however—none of the markedness constraints in the Lenient theory of CON are set up to favor ∅ above all other candidates.

One economy effect not yet discussed has long evaded a markedness explanation: vocalic syncope. Consider the following quote about a syncope process in Odawa:

(52) Why a rule should enter the language which simultaneously opacates a stress rule, destroys a surface alternating stress pattern and causes wholesale allomorphy, seems a question worth pondering. (Kaye 1974:149)

From the point of view of syllable structure, syncope is indeed puzzling, since it creates syllables with codas or complex onsets out of CV sequences. This has caused many researchers to appeal to economy principles (e.g., *V or *STRUC(σ)) and economy rules (e.g., V→∅) (Hammond 1984, Hartkemeyer 2000, Kiparsky to appear, Kisseberth 1970a, b, McCarthy 1986, Semiloff-Zelasko 1973, Taylor 1994, Tranel 1999). According to such analyses, syncope is a general, default operation—vowels are deleted whenever they are "unnecessary," just as "unnecessary" structure is deleted. This can be described as *Do*

Something Except When Banned (Prince and Smolensky 1993/2002). Under this view, the burden on the analyst is to explain only why deletion is blocked in certain contexts, but not why it is triggered in the first place.

Syncope is the empirical focus of chapters 3 and 4, where I argue that it results from the interaction of regular markedness constraints with MAXV. Because of the wealth and diversity of data, syncope is an ideal ground for the study of economy; yet I argue that there is no economy principle behind syncope—in fact, economy constraints are shown to be insufficient, unnecessary, or harmful.

The next section of this chapter focuses on *STRUC constraints, showing that they have harmful effects whether high ranked or not. Luckily, they cannot belong to CON if constraints are formulated leniently.

2.4 Ruling out *STRUC constraints

2.4.1 Introduction

In section §2.3 I argued that various economy effects follow from constraint interaction, without special economy principles. This is not an assumption shared in earlier OT work. To limit structure-building operations, Prince and Smolensky propose a special family of Economy constraints, *STRUC:

(53) Constraints of the *STRUC family ensure that structure is constructed minimally: a notion useful in syntax as well as phonology, where undesirable options (move-α; non-branching nonterminal nodes) typically involve extra structure... Pointless nonbranching recursion is ruled out by *STRUC, and bar-level can be projected entirely from functional information (argument, adjunct, specifier). In Economy of derivation arguments, there is frequently a confound between shortness of derivation and structural complexity, since each step of the derivation typically contributes something to the structure.

(Prince and Smolensky 1993:25, fn.13)

In the time since *STRUC constraints were originally proposed (Zoll 1993, 1996), they have been used in two senses that are not entirely distinct from each other: first, as a ban against nonterminal levels in some structural hierarchy (e.g., syllables), and second, as a ban on every element in the representation (e.g., features).

Intuitively, what all *STRUC constraints have in common is that they militate against all things, including those that are basic and unmarked. For example, *STRUC(σ) indiscriminately penalizes *all* syllables, whereas its more particular counterpart $*\sigma_{\mu\mu\mu}$ bans only *superheavy* syllables (see Chapter 3). Similarly, *C "no consonants" bans all consonants regardless of position, whereas NOCODA or *COMPLEX take syllable position into account. It is tempting to use this indiscriminateness as the unifying property of all *STRUC constraints. Nevertheless, non-*STRUC markedness constraints can be less complex or just as complex in definition as *STRUC constraints. No definitional property can usefully distinguish *NUC/t, a markedness constraint that expresses a strong cross-linguistic generalization, from *NUC/a, a *STRUC constraint whose only effect is economy (see §2.2.3 and §2.2.4). *STRUC constraints must therefore be identified by their external properties—the kinds of candidates that they penalize and their formal origins.

The theory of CON developed in §2.2 offers a way to define *STRUC constraints: they are the constraints that penalize the least marked non-null element on the relevant scale. In the remainder of this subsection, I will show how both kinds of *STRUC constraints are ruled out from CON under the proposed theory and why removing them from CON is necessary. But first let us review the two types of *STRUC constraints that have been proposed in OT (§§2.4.2, 2.4.3).

2.4.2 Prosodic Hierarchy-referring constraints

Prince and Smolensky's and Zoll's original *STRUC constraints ban the hierarchical structure that GEN imposes on the input: syllable structure, foot structure, or, in Prince and Smolensky's discussion, syntactic phrase structure. These constraints express the claim that *all structure is marked* and are a direct OT counterpart of Chomsky's (1991, 1995) Economy of Representation or Rizzi's (1997) "Avoid structure" principle.

In phonology, *STRUC constraints of this sort refer to the structure built by GEN that isn't necessarily present in the input:[29] *STRUC(μ) (Nishitani 2002), *STRUC(σ) (Kiparsky to appear, Zoll 1996), *STRUC(FOOT), *STRUC(PRWD), *STRUC(PHON-PHRASE) (Truckenbrodt 1999)—basically, they ban levels of the Prosodic Hierarchy. In the discussion that follows, these constraints will be called *PH-referring* *STRUC constraints.

Apart from the notional similarity between them, these constraints share an external property: only a Null Output can fully satisfy them. Thus, *STRUC(σ) can only be fully satisfied by a candidate that lacks the syllabic layer of prosodic structure or by one that contains no phonological material at all. The same is true for Truckenbrodt's (1999) *STRUC(PRWD) (see (54)). *STRUC(PRWD) assigns two violation marks to a candidate with two prosodic words *[pata][taa]* and one violation mark to the single prosodic word candidate *[patataa]*. Still, any null parse (c-e) will fare better than both *[patataa]* and *[pata][taa]*:

[29] In fact, many of the researchers cited here assume that prosodic structure is absent in the input and inserted only in GEN.

(54) PH-referring *STRUC and Null Outputs

/pata-taa/	*STRUC(PRWD)	MPARSE	MAX	PARSESEG
a. [patataa]	*			
b. [pata][taa]	**			
c. ∅$_{no\ correspondence}$	✓	*		
d. <patataa>$_{unprosodified}$	✓			*******
e. ∅$_{deleted}$	✓		*******	

This is a property common to all PH-referring *STRUC constraints: they assign zero violation marks *only* to Null Outputs. A PH-referring *STRUC constraint expresses a harmonic ordering of the sort shown in (55): zero is better than a mora, syllable, foot, and so on:

(55) Orderings imposed by PH-referring *STRUC constraints

Harmonic ordering	*STRUC constraint
$\varnothing \succ \mu$	*μ (Crosswhite 1999b, Nishitani 2002)
$\varnothing \succ \sigma$	*σ (Kiparsky to appear, Zoll 1993)
$\varnothing \succ$ Foot	*FOOT
$\varnothing \succ$ PrWd	*PRWD (Truckenbrodt 1999)

2.4.3 Nihilistic *STRUC constraints

The second category of *STRUC constraints shares little if any notional unity: they ban consonants, vowels (Hartkemeyer 2000, Kiparsky 1994), stress (Kiparsky 2003), coronal place (Fukazawa 1999), low and high vowels (Beckman 1998, Lombardi 2003), voiceless obstruents, and so on. Despite their diversity, these constraints have the character of economy principles: through their interaction with other constraints, these *STRUC constraints can very effectively duplicate the effects of classic economy principles.

Not all economy processes reduce the number of moras, syllables, and feet. De Lacy 1999 discusses haplology in Russian (see also §2.3.3.3), where the suffix /sk/ 'inhabitant of' haplologizes with a homophonous adjectival suffix, e.g., /tom-sk-sk-ij/ → *tomskij*, **tomskskij* 'of Tomsk (city name).' If this is indeed a case of haplology,[30] it reduces not the number of syllables but the number of segments and features. De Lacy analyzes this haplology process using *STRUC, which he defines as a constraint that assigns a violation for every node in the output form. Every feature of the output incurs a violation of *STRUC, regardless of how unmarked it is. Constraints of this sort are very similar in spirit to PH-referring *STRUC constraints, since they embody the claim that *everything is marked*.

When *STRUC is generalized in this manner beyond PH-referring constraints, it includes the set of all regular markedness constraints plus a number of constraints against everything, including unmarked things: vowels, voiceless obstruents, sonorant syllable peaks, and so on. Consider Hartkemeyer's *V, which assigns violation marks to all vowels. Whether a vowel is oral (relatively unmarked) or nasal (relatively marked), it will incur one violation of *V. The only candidates in (56) without violations are the Null Output (c) and the non-vowel candidate (d):

[30] *Tomsk* itself is not monomorphemic but back-formed from *Tomskij ostrov* 'Tom' island'. The adjective *tomskij* is formed from the name of the river *Tom'* using the adjectival suffix –*sk*. A more accurate description of what happens in **tomskskij* may be that the adjectival suffix haplologizes with itself (Robert Rothstein, p. c.).

(56) Nihilistic *STRUC constraints

	*NASALV	*V
a. u		*
b. ũ	*	*
c. ∅		✓
d. w		✓

This is a point of difference between the classic, PH-referring *STRUC constraints and nihilistic *STRUC constraints: the Null Output is not the only candidate that receives zero marks from the latter type of *STRUC. The next section presents a way to unify both types of *STRUC constraints by looking at them in terms of harmonic scales, which allows us to eliminate them from the theory altogether.

2.4.4 *STRUC constraints are impossible to derive from proper scales

While *STRUC constraints differ in the sort of harmonic orderings they impose on candidates, they agree in the harmonic orderings they impose on the members of a scale. According to a PH-referring *STRUC constraint, ∅ is more harmonic than a given level of the Prosodic Hierarchy. According to a nihilistic *STRUC constraint, ∅ is more harmonic than the least marked member of a harmonic scale. We can therefore pin down the property common to all *STRUC constraints:

(57) A *STRUC constraint bans the least marked non-null element on some scale.

All nihilistic *STRUC constraints can be related to scales in a fairly straightforward way: *ONS/t is derived from the onset sonority scale, *NUC/a is derived from the nucleus sonority scale, *ORALV (or *V) can be derived from the vowel nasality scale, and so on:

(58) Onset sonority harmonic scale : $\underset{\substack{\uparrow \\ \text{*ONS/t}}}{\underline{\text{Ons/t}}} \succ$ ... ons/i $\succ$ ons/a

(59) Vowel nasality harmonic scale: <u>Oral vowel</u> ≻ nasal vowel
↑
*ORALV (or *V)

These are the constraints that are not produced by Lenient Constraint Alignment (see §2.2.4), since it maps every member of a scale to a constraint *except for* the least marked member. A way to sneak around Lenient Constraint Alignment is to zero-extend scales, tacking ∅ as the least marked member of every scale. If all scales begin with ∅, then obstruent onsets, oral vowels, and other unmarked things are no longer the least marked things on their scales, and Lenient Constraint Alignment will create constraints against them but not against ∅. Scales of this sort, however, are prohibited by the NOZERO principle: scales cannot make vacuous harmony comparisons; harmony relationships must hold between two non-null structures. Thus, a scale like (60) cannot be used to sneak in a constraint against oral vowels (or all vowels) into CON:

(60) *Vowel nasality harmonic scale*: ∅ ≻ <u>Oral vowel</u> ≻ nasal vowel

The NOZERO principle is also the stumbling block for PH-referring *STRUC constraints. They must also be based on scales, but they have not been traditionally conceived in terms of scales because these constraints are really not comparative. According to *STRUC(σ), the syllable is not marked relative to some other structure (e.g., the mora), it is marked absolutely—only nothing is better than a syllable. Because of this, though, *STRUC(σ) cannot be based on a scale like (61), since it violates the NOZERO principle:

(61) Syllable scale 1: ∅ ≻ <u>σ</u>
↑
*STRUC(σ)

Removing ∅ from the scale leaves the unary scale (62). Nothing in the theory rules out unary scales, but Lenient Constraint Alignment cannot create a constraint based on them because it skips the least marked member of the scale. The least marked member of the scale in (62) is also its only member, so it is not eligible for constrainthood—the schema is set up so that it can only apply to a minimally binary scale. The presence of unary scales in the grammar has no affect on the constraint set.[31]

(62) Syllable scale 2: σ

Neither variety of *STRUC constraints can belong to CON if markedness constraints are formulated in such a way that they cannot penalize a structure unless there is some other structure that is less marked. Markedness constraints express the markedness of one form *relative* to another. All legitimate markedness constraints are based on scalar comparisons of this sort—comparisons that *STRUC constraints are incapable of making, because of their nihilistic nature.

2.4.5 Section summary

*STRUC constraints are OT's counterpart to the traditional idea that there are general economy principles constraining linguistic structure. Notionally, *STRUC constraints come in two varieties. The first is structural economy constraints; in

---

[31] Several interlocutors have suggested that the harmonic scale for constraints like *STRUC(σ) is the Prosodic Hierarchy. The chief problem with this strategy is that there is no evidence that shows prosodic words to be more marked than feet or feet to be more marked than syllables. It would be extremely difficult to come by such evidence, since it would have to be of the sort that shows, for example, that two prosodic words are less marked than a single foot. This is impossible, because higher-level prosodic constituents imply the presence of lower-level prosodic constituents. The Prosodic Hierarchy is not really a harmonic scale but a theory of the hierarchical organization of phonological representations, so no constraints can be derived straight from it without intermediate formal principles (e.g., EXHAUSTIVITY of Selkirk 1995).

phonology, these are the constraints against various levels in the Prosodic Hierarchy. The second kind of *STRUC is a more diverse set of constraints that embody the claim that "everything is marked": voiceless obstruents, oral vowels, sonorant nuclei, and so on. Together with regular markedness constraints, the latter type of *STRUC constraints duplicates the effects of structural economy. The theory of constraints developed here offers a way to unite the two sets: a *STRUC constraint bans the least marked non-null structure on its harmonic scale. Since scales cannot make vacuous markedness comparisons with null structures or penalize the unmarked, *STRUC constraints are excluded from the theory as a matter of principle.

2.5 Harmful effects of *STRUC constraints

The argument against economy principles is two-pronged. On the one hand, economy constraints are unnecessary because economy effects follow from independently motivated constraints (§2.3; see also chapters 3 and 4). On the other hand, economy constraints have harmful effects as freely rankable constraints. This section examines some of these effects.

In OT, a grammar is a language-particular ranking of universal constraints, and any ranking of constraints must produce an actual or at least a plausible grammar. *STRUC constraints are unlike other markedness constraints in that they are not freely rankable. *STRUC constraints upset the factorial typology in two ways: when high-ranked, they produce defective languages, and when low-ranked, they can have odd effects that stem from their nihilistic dislike of structure.

2.5.1 Why *STRUC must always be low-ranked

When *STRUC constraints are called upon to perform their economy duties, they always come second to other, higher-ranked demands. This is generally true of all economy principles: they limit but never ban. For example, as Grimshaw 2003 notes, Rizzi's "Avoid structure" principle (Rizzi 1997:314) is always "overridden" by other structure-building principles, since structure is never successfully avoided. The same is true of economy principles in phonology: *STRUC is dominated by at least some constraints in every analysis that employs it. For example, Hartkemeyer 2000 observes that *V must always be dominated, because the ranking of *V above all Faithfulness constraints describes an impossible language that lacks all vowels. Likewise, in Zoll's original analysis of Yawelmani ghost segments, *STRUC(σ) is allowed only to check epenthesis and to require the deletion of subsegmental features but never of whole segments (Zoll 1993, 1996).

It is not difficult to see why *STRUC constraints must be artificially restricted to the bottom of every language-particular ranking. If constraints like *OBS or *V can be undominated, the result is languages without obstruents or vowels, both unattested. Similarly, the existence of constraints like *ONS/t predicts languages that have no onsets, since they penalize the least marked onset of them all (Pater 1997).

This banishment of *STRUC from the top of every hierarchy is surprising under traditional OT assumptions that constraints are freely rankable (with the possible exception of constraints based on multi-valued prominence/markedness scales). Since *STRUC constraints are not based on such scales, their obligatory low ranking is hard to justify.[32] It is not clear which constraints universally dominate *STRUC. Faithfulness constraints cannot universally dominate *STRUC, since *STRUC must at least dominate MAX in at least some languages for deletion economy effects. As for markedness, the constraints that must dominate *STRUC differ from language to language. For example, in Lillooet, syncope cannot create onset clusters with rising sonority but can result in final stress, while in Lebanese Arabic it is the other way around. The constraints that block syncope must be ranked in the opposite way in the two languages: in Lillooet, it's SONSEQ>>*STRUC>>NONFINALITY, while in Lebanese Arabic, it's NONFINALITY>>*STRUC>>SONSEQ (for detailed analyses of these cases without *STRUC constraints, see Chapter 4). Thus we cannot even be sure which constraints universally dominate *STRUC—we only know that *some* must.

2.5.2 Odd effects under re-ranking

Even when dominated by other constraints, *STRUC constraints can have odd effects. By penalizing all structure without reference to markedness, PH-referring *STRUC constraints can produce implausible patterns that hinge only on reducing the number of structural nodes in the output. The pre-eminent *STRUC constraint, *STRUC(σ), predicts one such unattested pattern.

---

[32] The obligatory low ranking challenge cannot be addressed in the same way as the question that is often brought up against OT by skeptics: "if constraints are freely rankable, why are there no languages in which all markedness dominates all faithfulness?" (McCarthy 2002b:243-244). The problem here is a different one: "why isn't there a language in which just one *STRUC constraint is undominated?" *None* of the *STRUC constraints proposed in the literature is ever found at the top of a language's hierarchy.

2.5.2.1 Syllable economy and syncope

To understand the oddity of this pattern, we need a little background on attested metrical syncope patterns (these will be discussed in more detail in chapter 3). In the metrical theories of Hayes 1995 and Prince 1990, H and LH feet are equally well-formed as iambs: both are binary and satisfy the weight requirements on iambic feet by having heavy heads. Although these feet are equally well-formed metrically, they are not equally economical: (H) has one fewer syllable than (LH). Economy processes that show a preference for (H) over (LH) are not atttested, yet they are possible if *STRUC(σ) is admitted into CON.

First, let us briefly review what economy effects are attested in iambic languages. In many iambic languages, syncope applies to /LL.../ to yield (H)... and to /LLL.../ to yield (LH). Deletion of a vowel here frees up a consonant to serve as a weight-bearing coda in an iambic foot:

(63) Attested syncope patterns in iambic languages

a.	/takapa/	→	(ták)pa	not *(ta.ká)pa
	LLL		HL	(LL)L
b.	/takapana/	→	(ta.káp)na	not *(ta.ká)pa.na
	LLLL		(LH)L	(LL)LL

The outputs of syncope in (63) perform better than the faithful alternatives on the STRESS-TO-WEIGHT PRINCIPLE because their foot heads are heavy, not light. Syncope patterns just like this are found in Hopi (Jeanne 1978, 1982) Southeastern Tepehuan (Kager 1997, Willett 1982), Aguaruna (Alderete 1998, Payne 1990), and Central Alaskan Yupik (Gordon 2001, Hayes 1995, Jacobson 1985, Miyaoka 1985, Woodbury 1987).

Southeastern Tepehuan is unusual among these languages because it also deletes long vowels in some circumstances. Kager 1997 argues that such deletion minimizes the number of unfooted syllables. Long vowels syncopate only when the result is footed more exhaustively, so syncope applies only in the second example in (64) (the pattern is only shown schematically; for a more detailed discussion see chapter 4).

(64) Syncope of long vowels

a.	/takaapa/	(ta.káa)pa	*(ták)pa
	LHL	(LH)L	HL
b.	/taakaapan/	(táak)pan	not *(táa)kaa.pan or *(taa)ka.pan
	HHH	(H)H	(H)HH (H)LH

The output of syncope performs better on PARSE-σ than the faithful alternative—syncope allows the winner to pack more syllables into the foot. The important point here is that *syllables* are not counted—*unfooted syllables* are.

What we do not find, however, is an iambic language with a pattern just like Southeastern Tepehuan except that long vowels are deleted wherever it is possible to reduce the number of syllables:

(65) Non-occurring syncope pattern in iambic languages

a.	/takaapa/	tak.pa	σσ	not *ta.kaa.pa	σσσ
b.	/taakapa/	taa.kap	σσ	not *taa.ka.pa	σσσ

Yet with *STRUC(σ) in the grammar, this sort of pattern is predicted. Consider the tableau in (66), which includes metrical constraints, MAX-V, and *STRUC(σ). The constraints that are instrumental here are SWP ("if stressed, then heavy"), PARSE-σ ("no unfooted syllables"), NONFINALITY ("no final stress"), MAXV ("no V deletion"), and *STRUC(σ) ("no syllables"). As long as NONFINALITY dominates PARSE-σ and *STRUC(σ) dominates MAXV, /takaapa/ will map to *(ták)pa*:

(66) Iambic syllable reduction syncope with *STRUC(σ)

		SWP	NONFIN	*(σ)	PARSE-σ	MAXV
/takaapa/	a. (ta.káa)pa			***!	*	
	b. (ta.káap)		*!	**		*
	c. ☞(ták)pa			**	*	*
/taakapa/	d. (táa)ka.pa			***!	**!	
	e. ☞(táak)pa			**	*	*
/takapa/	f. ☞(ták)pa			**	*	*
	g. (ta.káp)		*!	**		*
	h. (ta.ká)pa	*!		***	*	

There is no metrical preference for (H) iambs over (LH) iambs—none of the metrical constraints in (66) favors *(tak)pa* over *(ta.kaa)pa*. These two types of feet are distinguished only by the number of syllables they have, i.e., by their performance on *STRUC(σ). If *STRUC(σ) is excluded from (66), *tak.pa* does not have a chance of emerging as the winner in any grammar—from the point of view of markedness (as opposed to economy principles), the deletion of the second vowel in /takaapa/ is gratuitous. The unattested pattern /takaapa/ → *tak.pa* is economy for economy's sake.

2.5.2.2 Syllable Alignment as an economy device

Some gradient Alignment constraints (McCarthy and Prince 1993a) can have a very similar effect. Consider the syllable alignment constraints of Mester and Padgett 1994, which assign a violation mark for every mora that stands between, a given edge of a syllable and the corresponding edge of a prosodic word: ALIGN-L(σ, PrWd). In fact, although Mester and Padgett proposed these constraints to analyze the so-called directional syllabification pattern in dialects of Arabic (see Broselow 1992a, Farwaneh 1995, Ito 1986), their economy potential was quickly realized. Spaelti 1997 and Walker

1998 use syllable alignment to limit the size of the reduplicant[33] (see §2.3.3.2 for a non-economy alternative), Davis and Zawaydeh 1996 rank syllable alignment constraints above MAXV to analyze Cairene Arabic syncope, Kager 1995 uses syllable alignment to derive stem disyllabicity in Guugu Yimidhirr, and Ussishkin 2000 proposes a different twist on syllable alignment, σ-ALIGN, to derive the disyllabic maximum size of stems in Hebrew, which is also enforced through syncope.

Under Ussishkin's (2000) theory of Hierarchical Alignment, binarity is optimal at all prosodic levels because it ensures that every constituent shares at least one edge with the prosodic word, thereby achieving prominence: if the prosodic word consists of one or two syllables, each syllable stands at an edge, but if the prosodic word consists of three syllables, the middle syllable is in a non-prominent position. The difference between this version of syllable alignment and that of Mester and Padgett 1994 is in the nature of the quantification over edges: in Mester and Padgett's version, the edge of every syllable must coincide with the *same* edge of a prosodic word, while Ussishkin's σ-ALIGN requires that the edge of every syllable coincide with *some* edge of a prosodic word.

Syllable alignment constraints are not fully equivalent to *STRUC(σ)—they differ in their assessment of monosyllabic words. *STRUC(σ) starts counting at *one* syllable, but syllable alignment is a bit more lenient—it starts counting at *two* syllables (except for Ussishkin's (2000) version, which starts counting at *three*):

[33] Walker 2000 actually departs from the syllable alignment analysis of Mbe in favor of *STRUC(σ), noting that the two strategies achieve nearly identical results.

(67) Hypothetical syllable alignment constraints and their economy effects

	*STRUC(σ)	ALIGN-L(σ, PrWd)	σ-ALIGN
a. [σ]	*		
b. [σσ]	**	*	
c. [σσσ]	***	**	*
d. [σσσσ]	****	***	**

Formally, syllable alignment constraints are not *STRUC constraints—they do not penalize the least non-null member on a harmonic scale. When it comes to scales, though, gradient syllable alignment is fairly suspect—it necessitates either scales of infinite length or $n \succ n+1$ scales. Infinitely long scales are an impossibility since CON must be finite, while $n \succ n+1$ scales add a powerful device to the theory that is otherwise unnecessary (this point was first raised in §2.2.6).

As for σ-ALIGN, it is neither a *STRUC constraint nor a gradient alignment constraint—it does not assess the distance between a medial syllable and a word edge, distinguishing only between medial syllables (bad) and edge syllables (good). Nevertheless, it may be necessary to give σ-ALIGN the slip as well, since it has the same effect as *STRUC(σ) in the matter of /takaapa/ → *(ták)pa*. The problem is that neither gradient syllable alignment nor σ-ALIGN pay any regard the prosodic status of the syllables in question—the thing that matters to these constraints is the number of syllables in the output, not metrical well-formedness.

Consider the tableaux below, which are versions of (66) with *STRUC(σ) replaced by gradient syllable alignment and σ-ALIGN, respectively. The third output, *(ták)pa*, is harmonically bounded by *(takáa)pa* if ALIGN-L(σ, PrWd) and σ-ALIGN are excluded from CON, but if they are present, *(ták)pa* has a serious shot at being the winner—all that's required is that the relevant syllable-counting constraint dominate MAXV and that NONFINALITY dominate PARSE-σ.

(68) Economy for economy's sake, with gradient syllable alignment of Mester and Padgett (1994)

/takaapa/	NONFIN	ALIGN-L(σ, PrWd)	MAXV	PARSE-σ
a. (takáa)pa		**!		*
b. (takáap)	*!	*	*	
c. ☞(ták)pa		*	*	*

(69) Economy for economy's sake, with σ-ALIGN of Ussishkin (2000)

/takaapa/	NONFIN	σ-ALIGN	MAXV	PARSE-σ
a. (takáa)pa		*!		*
b. (takáap)	*!		*	
c. ☞(ták)pa			*	*

Exclusion of these constraints from CON still leaves the analyst some devices for analyzing maximum size restrictions—see §2.3.3.2.

To summarize, I argue that the ability to penalize syllables without reference of their metrical status is harmful whether it is an attribute of a true economy constraint like *STRUC(σ) or of syllable alignment constraints. All three kinds of constraints discussed here can favor an unattested pattern where H is chosen over the otherwise well-formed LH iambic foot. The only way to avoid this situation is to not let constraints penalize syllables except qua their metrical affiliation.

PH-referring economy constraints are not the only constraints with harmful effects—in the next section, I explore some predictions of having nihilistic constraints "against everything."

2.5.2.3 Emergence of the marked in reduplication and positional faithfulness

Even when nihilistic *STRUC constraints are dominated, they can have effects in situations that McCarthy and Prince (1994a) dub 'the emergence of the unmarked.' The effect of nihilistic *STRUC constraints, however, is more appropriately described as emergence of the marked—by penalizing unmarked segments, they can favor outputs that are marked. Two environments where the effects of nihilistic *STRUC constraints can be felt are reduplicants and non-privileged positions.

Reduplicants often contain a subset of the language's sound inventory, and it has been claimed that it is always the unmarked subset (Alderete et al. 1999, McCarthy and Prince 1994a, 1995). For example, in Tübatulabal, the first onset of the base is copied into the reduplicant as a stop with the least marked place of articulation, glottal:

(70) ʔ-reduplication in Tübatulabal (Alderete et al. 1999, Voegelin 1958)

a.	pɨtɨta	→	ʔɨ-pɨtɨta	'to turn over'
b.	to:yan	→	ʔo:-doyan	'he is copulating'
c.	ʃiʔɨwɨ	→	ʔɨ:-ʃiʔɨwɨ	'it looks different'
d.	ʔa:baʔiw	→	ʔa:-ʔabaʔiw	'it is showing'

Alderete et al. 1999 argue that *ʔ* is the default segment in Tübatulabal because it violates the lowest-ranked place markedness constraint, *PL/PHAR:

(71) *PL/LAB, *PL/DORS>>*PL/COR>>*PL/PHAR (Lombardi 2001, 2002, Prince and Smolensky 1993)

This hierarchy is ranked between MAX-C_{IO} and MAX-C_{BR}, as shown in (72): in normal input-output mappings, consonants with any place are mapped faithfully (cf. (d) and (d)), but in reduplication copying, only glottal stops[34] are permitted to surface (cf. (a) and (b)). Non-glottal consonants are deleted and replaced by epenthetic *ʔ*. Alderete et al. argue that the reason any consonants surface at all in the reduplicant is that ONSET is high-ranked (cf. (a) and (c)):

(72) The Tübatulabal onset (from Alderete et al. 1999:345)

/RED-toyan	MAX-C_{IO}	ONS	*PL/COR	*PL/PHAR	MAX-C_{BR}	DEP-C_{BR}
a. ☞ ʔo:-doyan			d, y, n	ʔ	d, y, n	ʔ
b. to:-doyan			t!, d, y, n		y, n	
c. o:-doyan		*!	d, y, n		d, y, n	
d. ʔo:-ʔoʔaʔ	d, y, n!			ʔ, ʔ, ʔ, ʔ	ʔ, ʔ	

If the ranking of ONSET and *PL/PHAR were reversed, however, the result is a pattern where *no* consonants are permitted in the reduplicant. This would look like this:

(73) Onsetless reduplicants (an unattested pattern)

a. /RED+napa/ a.a-na.pa
b. /RED+ʔita/ i.a-ʔi.ta
c. /RED+weta/ e.a-we.ta

The same result can be obtained by ranking the non-lenient version of the onset sonority hierarchy (see (10)) below ONSET and MAX-C_{IO}. Since the onset sonority hierarchy and the place markedness hierarchy penalize the entire range of possible consonants, their ranking between Max-C_{IO} and Max-C_{BR} obliterates consonants from reduplicants. This is while the normal onset inventory of the language is harmonic:

[34] The place hierarchy analysis alone cannot explain why *h* is copied as *ʔ*: /RED-hu:ʔ/ → *ʔu:-hu:ʔ* 'it leaked (Crowhurst 1991:52). Presumably, either the constraint against fricatives or *ONS/FRIC rules out the faithful copying of *h*.

(74) Onsetless reduplicants

/RED+ʔita/	MAX-IO	*ONS/t	MAX-BR
a. ʔiʔa-ʔita		*!***	
b. ☞i.a-ʔita		**	**
c. i.a-i.a	*!*		

The culprits here are *PL/PHAR and *ONS/t: because they penalize the least marked elements on their respective scales, they act as *STRUC constraints. If these constraints were eliminated, *not* copying ʔ would not be an option because it gratuitously violates MAX-BR.

These sorts of constraints can have a similar effect when they interact with positional faithfulness. Beckman 1998 reports numerous patterns where marked structure is allowed to surface only in special positions, e.g., the initial segment of the word but not elsewhere. The prediction is, then, that given the ranking F_{pos}>>*STRUC>>F, structure marked with respect to nihilistic *STRUC constraints should only be present in designated positions. For example, consider the following hypothetical language, which has consonants only in the initial syllable but hiatus elsewhere:

(75) Consonants in initial syllable only

a. /nalikepati/ → ná.i.e.a.i
b. /wata/ → wá.a
c. /aina/ → á.i.a

All onset constraints including *ONS/t are dominated by MAX-INITIAL but not by the non-positional MAX. Thus, word-initial consonants are preserved but word-internal ones must delete:

(76) Consonants deleted except in first syllable

/wata/	MAX-INITIAL	*ONS/w	*ONS/l	*ONS/n	*ONS/t	MAX
a. ☞wa.a		*				*
b. wa.ta		*			*!	
c. a.a	*!					**

Nihilistic constraints against vowels also have the potential for favoring unattested syncope patterns. In chapter 4, I discuss the effects of context-free markedness constraints *LOW and *NONLOW (Lombardi 2003) in more detail. In brief, the issue is that there is an asymmetry in differential syncope patterns: there are languages where low sonority vowels (e.g., *ə* or *i*) delete wherever possible, and there are languages where high sonority vowels (e.g., *a*) delete in unstressed positions, but there are no languages where high sonority vowels delete wherever possible but other vowels do not. This asymmetry can be explained only if there are no context-free markedness constraints against high sonority vowels. If *LOW is allowed into the grammar, the pattern is wrongly predicted to exist. In chapter 4, I discuss this prediction in more detail and provide an alternative to the context-free markedness theory of epenthetic vowel quality that allows us to expunge *LOW and *NONLOW from CON.

The patterns discussed here are inevitable under the view that "everything is marked." The only way to get around such predictions is to exclude certain constraints from CON. A straightforward way to do that is to formulate constraints leniently based on harmonic scales.

2.5.3 Section summary

This section has defined *STRUC constraints and discussed some of their harmful effects: unattested inventory gaps (e.g., languages without obstruents or vowels) and bizarre structure-reducing patterns such as syncope to reduce the number of syllables, "emergence of the marked" in reduplication, and absence of elements like consonants (not traditionally seen as marked) outside privileged positions. These patterns are nothing more than slight improvisations on the originally intended function of Economy constraints: favoring smaller structures. The problems that Economy constraints cause cannot generally be solved by restricting their ranking—they suggest that constraints of this sort must be excluded from the theory altogether.

2.6 Chapter summary

This chapter presented a theory of economy effects without economy principles. Economy effects, it was argued, are nothing but a consequence of a language-specific ranking of constraints. Moreover, economy effects never target unmarked structure—if something is deleted, the goal is a less marked output rather than a shorter output.

The theory relies on a different conception of markedness: markedness is always a relative property. A structure can only be marked if there is another non-null structure that is not marked. This is formally encoded in the Lenient Theory of CON, whereby constraints penalize every element on their respective harmonic scale except for the most harmonic one. The scales themselves cannot stipulate nihilistic comparisons, e.g., "$x \succ \varnothing$." Language-specific grammars can prefer ∅ to every other candidate in a comparison, but individual constraints do not.

A consequence of this approach is that economy constraints are banned from CON, which I argue is necessary in any case because they have harmful typological effects. The argument takes a different turn in the next chapter, where I show that a particular economy effect, metrical syncope, can be analyzed to great effect in terms of independently motivated constraints, which account not only for the details of the syncope processes in the languages examined but also for other aspects of their phonologies. Conversely, economy constraints contribute nothing to the understanding of these processes.